



**HAL**  
open science

# On the Limits of Machine Perception and Interpretation

Bart Lamiroy

► **To cite this version:**

Bart Lamiroy. On the Limits of Machine Perception and Interpretation. Image Processing [eess.IV]. Université de Lorraine, 2013. tel-00940209

**HAL Id: tel-00940209**

**<https://theses.hal.science/tel-00940209>**

Submitted on 31 Jan 2014

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

# On the Limits of Machine Perception and Interpretation

## MÉMOIRE

présenté et soutenu publiquement le 3 décembre 2013

pour l'obtention de l'

**Habilitation à Diriger des Recherches de l'Université de Lorraine**  
(Spécialité Informatique)

par

Bart LAMIROY

### Composition du jury

<i>Rapporteurs :</i>	Jean-Philippe DOMENGER	Université de Bordeaux 1 – France
	Rolf INGOLD	Université de Fribourg – Suisse
	Simone MARINAI	Université de Florence – Italie
<i>Examineurs :</i>	Karl TOMBRE	Université de Lorraine – France
	Josep LLADÒS	Université Autonome de Barcelone – Espagne
	Jean-Marc OGIER	Université de La Rochelle – France

Mis en page avec la classe thloria.

*La recherche est un métier d'incompréhension... vanitas vanitatis*

*"Il n'y a que les imbéciles qui ne changent jamais d'avis", dit le sage ; et comme il n'était pas imbécile, il changea immédiatement d'avis.*



## Remerciements

*"Beauty is in the Eye of the Beholder" ...*

Par ordre décroissant d'importance, il me semble essentiel de remercier d'abord mon épouse et mes enfants, qui ont eu (et ont toujours) à subir mes déplacements à l'autre bout du monde, sans compensation quelconque et toujours à des moments où cela interfère avec la vie de famille, les laissant se débrouiller tous seuls, ma tendance à oublier l'heure quand je suis "sur un truc", les nuits blanches et l'énervement à l'approche des *deadlines*, et les vacances pas assez nombreuses ou en demi-teinte pour cause de tête dans les nuages.

Ensuite le système de fonctionnariat à la française et la liberté incroyable qu'il octroie à ses serviteurs chercheurs et enseignant-chercheurs.

En enfin, l'ensemble des collègues, étudiants, administratifs et autres personnes qui ont fait que mon parcours professionnel est celui qui est repris dans ce document. D'aucuns pour m'avoir fait confiance, avoir toléré et parfois même apprécié mes regards et opinions décalés – voir, pour quelques uns, les avoir encouragés –, beaucoup pour m'avoir fait bénéficier de leurs soutiens financiers et qui m'ont permis, à tout moment, de faire ce que me plaisait de faire ; certains, qui par leur opposition, ou leur regard critique m'ont amené à réfléchir et à garder une humilité productive et qui m'ont aussi parfois permis de rebondir dans des directions inattendues ; tous, sans exception, pour m'avoir fait avancer dans la vie ... j'espère dans la bonne direction.



# Table des matières

<b>1</b>	<b>Synthèse en Français</b>	<b>1</b>
1.1	Curriculum vitae . . . . .	1
1.1.1	Indicateurs et statistiques de recherche . . . . .	1
1.1.2	Synthèse des activités de recherche . . . . .	3
1.1.3	Séjours internationaux et collaborations . . . . .	6
1.2	Projet scientifique . . . . .	7
<b>2</b>	<b><i>Curriculum Vitae</i> and Achievements</b>	<b>11</b>
2.1	Synopsis . . . . .	12
2.2	Research . . . . .	13
2.3	Teaching . . . . .	23
2.4	Publications . . . . .	28
<b>3</b>	<b>Scientific Positioning</b>	<b>35</b>
3.1	Image Recognition and Indexing . . . . .	35
3.1.1	Scientific Context and State-of-the-Art . . . . .	36
3.1.2	Highlights and Contributions . . . . .	36
3.1.3	Publications on this Topic . . . . .	38
3.2	Visual Servoing . . . . .	40
3.2.1	A Genuine Industrial Implementation Context . . . . .	40
3.2.2	Scientific Context and State-of-the-Art . . . . .	41
3.2.3	Highlights and Contributions . . . . .	42
3.2.4	Publications on this Topic . . . . .	43
3.3	Graphical and General Document Analysis . . . . .	45
3.3.1	Scientific Context and State-of-the-Art . . . . .	45

3.3.2	Highlights and Contributions . . . . .	52
3.3.3	Publications on this Topic . . . . .	52
<b>4</b>	<b>Selected Papers</b>	<b>57</b>
	Scan-to-xml: Using Software Component Algebra for Intelligent Document Ge- neration . . . . .	59
	Robust and Precise Circular Arc Detection . . . . .	70
	Symbol Recognition using Spatial Relations . . . . .	82
	How Carefully Designed Open Resource Sharing Can Help and Expand Docu- ment Analysis Research . . . . .	110
	Computing Precision and Recall with Missing or Uncertain Ground Truth . . . .	124
<b>5</b>	<b>the Limits of Image Interpretation</b>	<b>139</b>
5.1	Introduction . . . . .	140
5.2	Comparing Machine Perception Algorithms . . . . .	145
5.2.1	Notations and Definitions Relating to Reference Data . . . . .	145
5.2.2	On the (Correct) Use of Ground Truth . . . . .	147
5.2.3	On the Subjectivity of Ground Truth . . . . .	150
5.2.4	Interpreting Ground Truth . . . . .	155
5.3	Analysis and Interpretation . . . . .	158
5.3.1	Interpretation . . . . .	158
5.3.2	Context . . . . .	159
5.3.3	Analysis . . . . .	163
5.4	Modeling Contexts and Interpretations . . . . .	164
5.4.1	Interpretation is Undecidable . . . . .	165
5.4.2	Interpretation as a Computational Problem . . . . .	166
5.4.3	Interpretations as Data Mining . . . . .	168
5.5	Related Initiatives . . . . .	173
5.6	Conclusions . . . . .	173
	<b>Bibliography</b>	<b>175</b>
	<b>Appendices</b>	<b>187</b>
<b>A</b>	<b>Funding Proposals</b>	<b>189</b>

PEPS CNRS INS2I Rupture : Contresens . . . . .	190
PEPS CNRS HuMaIn : Inter-Est . . . . .	193
PICS CNRS : Speedit . . . . .	197



# Chapitre 1

## Synthèse en Français

### Préambule

L'essentiel de ce document est écrit en anglais, pour diverses raisons. Nous fournissons néanmoins une synthèse du document en français. C'est donc à quoi répond ce chapitre. Il reprend et traduit certaines parties du texte complet en anglais, qui, lui, constitue le document de référence. Il n'est pas conseillé au lecteur averti de s'attarder sur la version française, dont le but n'est pas de fournir un regard différent ou des informations complémentaires au texte de référence. Il n'a pas été sujet à une relecture aussi approfondie que le reste du texte, et il est possible que certaines incohérences subsistent.

### 1.1 Curriculum vitae

La totalité du chapitre 2, p. 11 comprend mon curriculum vitae et couvre mes activités de recherche et d'enseignement durant mes périodes de thèse de doctorat, de Post-Doc et an tant que maître de conférences. La version anglaise a une portée plus large qu'un *curriculum vitae* classique et fournit une synthèse de mes contributions scientifiques dans les domaines où j'ai été impliqué, fournissant ainsi les informations de base nécessaires pour comprendre les raisons et les origines des projets de recherche développés dans les autres parties de ce document. La version ci-dessous reprend quelques éléments significatifs et référence les développements plus détaillés dans les autres parties de ce document.

#### 1.1.1 Indicateurs et statistiques de recherche

Cette section reprend, de façon purement comptable, l'ensemble des indicateurs "significatifs" de mon activité de recherche. Ces indicateurs sont mis en contexte et commentés dans les sections *ad hoc* du manuscrit général.

### Activités d’encadrement

L’ensemble des activités d’encadrement est décrit en détail dans la section 2.2, p. 16.

Co-encadrement de 3 thèses : Jan RENDEK (1 publi, abandon de la thèse par le thésard), Jean-Pierre SALMON (0 publi, retrait de l’encadrant), Santosh K.C. (4 revues, 2 LNCS-GREC, 5 conférences internationales, 3 autres).

Accompagnement de travaux de thèses internationales de 3 autres candidats (J. MAS, M. ILIE, T. SUN – 4 publications)

33 projets/stages de niveau Master, principalement des projets d’un semestre en École d’Ingénieur, 3 stages de M2 Recherche – 12 publications.

### Production scientifique

La liste exhaustive des publications et productions scientifiques est disponible pp.28–34 : 1 livre, 5 revues internationales, 8 chapitres, ouvrages collectifs ou LNCS, 4 vulgarisations ou conférences invitées, 40 conférences internationales, 7 ateliers internationaux (avec comité de relecture et publication d’actes), 6 ateliers internationaux sans actes publiés (mais avec publication post-atelier), 4 colloques nationaux.

### Valorisation et transfert

- Transfert technologique de logiciels de suivi de cibles vers OSS Steel Shipyard (Odense, Danemark) dans le projet Européen Vigor
- Bourse CIFRE Jan Rendek (France Télécom)
- Transfert technologique avec la société française Algo’Tech and le cadre du projet STREP "FRESH"
- Projet de transfert en cours sur la détection de cercles avec la société française Exameca

### Implications et rayonnement scientifiques

- “Data Curator” et membre du bureau du TC-10 de l’IAPR (*International Association for Pattern Recognition*),
- membre du comité “*Publications & Publicity*” de l’IAPR,
- General Chair de GREC 2013 (*Tenth International Workshop on Graphics Recognition*)

- membre du comité éditorial d'IJDAR (*International Journal of Document Analysis and Recognition*, édité par Springer)
- membre du comité de pilotage d'RFIA 2014 (dix-neuvième congrès francophone sur la Reconnaissance des Formes et l'Intelligence Artificielle)
- trésorier de l'AFRIF (Association française de reconnaissance et interprétation de formes)
- membre du comité de programme de 4 conférences internationales (ICDAR 2013, ICDAR 2014, GREC 2011, ICCVG 2012)
- membre du comité de programme de 4 conférences nationales (CIFED 2010, 2012, 2014, CIDE 2008)
- 1 mention spéciale aux attributions des meilleurs papiers à ICDAR (*International Conference on Document Analysis and Recognition*) 2011, pour [21]
- vainqueur de la sixième édition du GREC *Arc Segmentation Contest* en 2011 avec [12].

### 1.1.2 Synthèse des activités de recherche

Cette partie est la traduction de la section 2.2.

Ma recherche a toujours été liée au traitement d'images : pendant mon doctorat (à Grenoble dans l'équipe MOVI, dirigé par R. MOHR) j'ai couvert des aspects d'indexation et de reconnaissance, au cours de mon post-doc (à Grenoble, dans l'équipe BIP, dirigé par B. ESPIAU) j'ai évolué vers la modélisation 3D pour l'asservissement visuel de robots, et depuis mon recrutement en tant que maître de conférences (à Nancy, dans l'équipe QGAR, dirigée par K. TOMBRE, puis par S. TABBONE), j'ai contribué au domaine de l'analyse de documents graphiques. Pendant mon séjour en tant que chercheur invité à l'Université de Lehigh (D. LOPRESTI et H. BAIRD) j'ai élargi mon champ d'application à l'analyse de documents en général.

#### 1994 - 1998 La reconnaissance d'image et indexation

Mes travaux de thèse concernaient la reconnaissance d'objets dans les images à partir des données structurelles. Ces données ont servies comme base à des techniques d'indexation en utilisant quasi-invariants affines et projectives, afin d'absorber les déformations de l'image dues à des changements de point de vue. Dans un premier temps, les travaux ne considéraient uniquement les contours de l'image [38,57]. L'approche a ensuite progressivement été étendue à des combinaisons géométriques cohérentes de contours et de zones d'intérêt [19,33,15,46]. Notre approche a été testée avec succès dans un environnement industriel pour l'identification de pièces mécaniques de moteurs d'automobiles

[56]. Une des parties les plus fondamentales de ce travail a été l'étude de la complexité algorithmique inhérente à l'indexation d'images basée sur leur contenu [37,55]. [37] a été l'un des premières publications à remarquer que l'indexation de données multimédia doit impérativement prendre en compte la complexité induite par la recherche de voisinages dans des espaces de descripteurs de dimension élevée, tout en offrant quelques pistes de solutions.

### 1998 - 2000 asservissement visuel

Pendant la période 1998 - 2000, j'étais en charge de la coordination du projet européen VIGOR dont notre équipe de recherche était coordinateur principal. Les partenaires du consortium étaient : l'équipe de R. CIPPOLA à l'Université de Cambridge, celle de H-H. NAGEL au IITB Fraunhofer de Karlsruhe, A. SASHUA et M. WERMAN de l'Université Hébraïque de Jérusalem, Inria Rhône-Alpes – équipes BIP (B. ESPIAU) et MOVI (R. HORAUD) – puis les entreprises Odense Steel Shipyard, Ltd (Danemark) et Sintors SA (France).

Outre les tâches d'organisation et de coordination, j'étais également chargé de superviser le processus de transfert de technologie avec les partenaires industriels. Mes principales contributions scientifiques concernaient le contrôle stéréoscopique visuel métrique et non métrique. Pendant le projet, nous avons étendu l'état de l'art dans l'asservissement visuel du contrôle monoculaire étalonnée au contrôle stéréo non calibré, en intégrant les contraintes épipolaires [35] dans les résultats publiés antérieurement [7,39]. L'approche a ensuite été validée sur des données réelles *in situ* à Odense Steel Shipyard, dans le cas de soudage asservi visuellement [32].

Parallèlement à ce travail, j'ai contribué avec un doctorant, A. RUF, à la caractérisation des mouvements rigides dans l'espace projectif  $\mathbb{P}^3$ . En parallèle, le Master de F. MARTIN a aidé à appliquer la théorie développée à la calibration de robots à la volée. Ce travail a ensuite conduit à une méthode de contrôle stéréo non calibré [14,36].

### Depuis 2000 : analyse de documents graphiques

En 2000 j'ai Intégré l'équipe QGAR à l'INRIA Nancy Grand-Est/LORIA, ce qui m'a requis de me concentrer sur l'interprétation des documents graphiques. Ces types de documents ont la propriété d'être exprimés dans un langage visuel qui porte habituellement une sémantique plus claire ou plus explicite que les images photographiques d'environnements réels. Mes principales contributions présentées ici concernent toutes l'extraction des structures résultant d'un pipeline de traitement de documents [13]. Elles peuvent être classées selon trois thèmes principaux :

#### Structuration sémantique et Scan-to-XML

Afin de modéliser la sémantique des documents, nous avons travaillé sur la navigation automatisé dans les documents graphiques. Cela nécessite notamment d'exprimer les relations sémantiques permettant de relier les zones portant un sens proche. Les résultats obtenus [53,31,45] (dont [31] a été réalisée en collaboration avec le laboratoire CVC à l'Universitat Autònoma de Barcelona, Espagne), ont donné lieu à une représentation plus générique de la façon d'exprimer la sémantique des documents à l'aide d'une algèbre de composants [13].

Ces travaux ont mené à la conclusion que la sémantique d'un document peut être exprimée comme une structure en treillis, qui à son tour permet de séparer correctement le contexte (ou ontologie liée au contexte d'application) de l'analyse bas niveau. Cela nécessite, toutefois d'identifier correctement tous les composants et leurs interactions [18]. Ce travail a conduit au financement d'une thèse par France Télécom R&D (J. RENDEK) en Novembre 2004. Il a également conduit à une tentative de collaboration avec l'équipe TexMex d'INRIA à Rennes, en ce qui concerne l'interaction entre les techniques de fouille de textes et graphiques en analyse du document. Étant donné que le lien entre le texte et l'image passe nécessairement par un niveau conceptuel plus élevé, et que ces concepts sont inévitablement extrait de l'graphiques à travers un pipeline de détecteurs et analyseurs, nous avons comblé l'écart avec l'algèbre de composants développée dans [13]. De cette réflexion commune ont ensuite découle tous nos autres travaux sur de la programmation logique inductive [27,52,54] que nous développons dans la section suivante.

## Indexation et Reconnaissance

Le contexte décrit précédemment explique bien l'origine de mon sujet de recherche suivant, qui est basé sur deux thèmes principaux, et visant à une intégration de la boucle globale d'extraction d'information visuelle et l'utilisation de la sémantique :

- Les méthodes d'apprentissage pour identifier l'information graphique pertinente. Beaucoup de ce travail a été fait sous contrat avec France Télécom R&D, et en étroite collaboration avec le CVC à Barcelone, principalement à travers les thèses de J. RENDEK et J. MAS-ROMEU. La première concerne l'interprétation et la caractérisation off-line de symboles dessinés à main levée dans une collection de documents non structurés. La méthode est basée sur l'apprentissage par retour de pertinence combinée à de classification statistique [44,60,61]. La seconds a donné lieu à une approche robuste et assez novatrice de génération dynamique de grammaires d'adjacence pour la reconnaissance *on-line* de symboles graphiques [29,30,43].
- L'utilisation d'un vocabulaire visuel pour la reconnaissance de symboles en utilisant des composants de détection robustes. Ce travail a commencé lors d'un projet européen FP-6 STREP appelé FRESH et dans lequel la partie de reconnaissance de symboles est basée sur l'algèbre de composants. De janvier 2007 à novembre 2008, j'ai été en charge de la partie consacrée à la reconnaissance graphique. Outre la gestion de projet et les rapports, j'étais également

co-encadrant de la thèse de JP SALMON [28,62] 2004-2008, et responsable d'un ingénieur embauché pour le transfert de technologie avec les partenaires industriels (L. FRITZ) [28].

Les conclusions de ces travaux ont donné lieu à la thèse CORDI – INRIA de S. K.C. de 2008-2011. Son objectif principal était d'étudier les moyens d'utiliser les relations spatiales et le positionnement relatif des éléments visuels afin d'obtenir une description d'images efficace et utile à la reconnaissance. La méthode s'appuie sur le vocabulaire décrit précédemment [28,12,62] combiné à la programmation logique inductive (PLI). La PLI a déjà montré son utilité pour l'extraction relations non triviales entre les concepts dans des documents textuels. Ce travail a permis d'évaluer l'approche dans le contexte spécifique de documents graphiques où l'incertitude, le bruit et complexité de calcul sont un problème. Ce projet comporte trois phases principales : tout d'abord, l'évaluation la qualité des descriptions des symboles en utilisant les relations spatiales [9,50] ; ensuite, la formalisation du vocabulaire visuel afin de pouvoir l'utiliser pour la description de symboles complexes et d'être intégré avec la PLI [54,27,52], et enfin, dans la phase finale, l'apprentissage automatique de descriptions de symboles complexes [3].

**Analyse de performance et interprétation** La structuration sémantique était ensuite le déclencheur de mes activités en tant que chercheur invité à Lehigh University, où j'ai été responsable de la supervision d'un projet cherchant à proposer une approche globale et collaborative à l'accès à de ressources (données, annotations et algorithmes de référence) pour l'analyse de documents [21,23,24,40,41,42] appelée DAE (*Document Analysis and Exploitation*<sup>1</sup>). La direction de recherche engagée comprend l'exploration combinée des approches d'analyse de documents et la représentation des connaissances, plus précisément liées aux ontologies et aux logiques de description et tente de fournir une base formelle pour l'évaluation de performances et le vérification (ou certification) des résultats expérimentaux publiés. Cette collaboration a également conduit à l'organisation d'un concours international d'évaluation des performances d'analyse de documents de bout en bout [16], la première édition duquel a eu lieu à ICDAR 2011.

### 1.1.3 Séjours internationaux et collaborations

**Août 2000, février 2001** Odense, au Danemark, dans les installations du chantier naval Odense Steel afin d'intégrer et de transférer des logiciels pour l'asservissement visuel de robots de soudage et le démonstrateur final du projet VIGOR.

**Juillet - Décembre 2001** collaboration avec E. VALVENY du Centre de Visió per Computador (CVC) de l'Universitat Autònoma de Barcelona, pendant son post-doc. à Nancy sur la détection de relations sémantiques dans les diagrammes de type "éclaté" [31].

---

<sup>1</sup><http://dae.cse.lehigh.edu>

**Novembre 2005** Centre de Visió per Computador, Barcelone, avec prof. J. LLADÓS et encadrement conjoint des thèses de J. RENDEK et J. MAS ROMEU (CVC) [30,43]

**Novembre 2007** City University Hong Kong (prof. LIU Wenyin) sur la détection de cercles [28].

**Mai 2009** Centre de Visió per Computador (CVC) at the Universitat Autònoma de Barcelona (prof. J. LLADÓS) de travail conjoint avec J. MAS ROMEU sur les grammaires d'adjacence.

**Juin 2009** Collaboration avec prof. E. BARNEY SMITH de Boise State University, en visite à Nancy en tant que professeur invitée. Les travaux communs sont toujours en cours sur l'évaluation de la performance de la détection de cercles et la binarization et un séjour est prévu, en tant que chercheur invité prévue à BSU courant 2014.

**2010-2012** Chercheur invité au Computer Science and Engineering Department de Lehigh University (pendant 18 mois en 2010-2011 avec D. LOPRESTI et H. BAIRD), suivie d'une poursuite de la collaboration avec D. LOPRESTI : je bénéficie actuellement de une position de courtoisie à Lehigh, me donnant accès à leurs ressources informatiques ; D. LOPRESTI a fait des séjours cours au LORIA en décembre 2011, juillet 2012 et en juin 2013 [42,24,16,21,23,40,41].

**Juillet 2011** Chercheur invité à la National Library of Medicine, Bethesda, MD, USA. Collaborations avec G. THOMA et S. ANTANI sur adaptation de la plate-forme DAE pour leurs données expérimentales.

**Septembre 2012** M. ILIE, doctorant à l'Université du Danube inférieur (*Dunărea de Jos*) de Galați, Roumanie, a séjourné pendant 2 semaines à Nancy pour le démarrage d'une collaboration et de consultation sur son Ph.D. sujet lié à la recherche d'images par le contenu et l'analyse de documents. D'autres séjours sont prévus en 2014.

## 1.2 **Projet scientifique**

Le projet scientifique est développé en détail dans le chapitre 5. Les paragraphes qui suivent en donnent un rapide tour d'horizon et sont, pour l'essentiel, la traduction de l'introduction trouvée à la page 140.

Le projet s'articule principalement autour de l'évaluation de performances à travers les notions d'ambiguïté d'interprétation et tente d'aboutir une modélisation de la notion de contexte dans le domaine de l'interprétation d'images. Il vise également le développement d'une plateforme de référence pour l'hébergement de données et outils d'évaluation de performances dans le domaine de l'analyse de documents en proposant d'utiliser des outils du domaine de la fouille de données et de la modélisation de connaissances.

L'idée générale du projet de recherche développé dans ce manuscrit est de constater qu'il est très difficile, au vue de la quantité de résultats publiés faisant état de résultats de recherche expérimentale dans l'analyse de documents, de clairement établir une cartographie de l'état de l'art. Plus particulièrement, comment pouvons-nous décider, dans le cadre de tous les données, algorithmes et approches publiés, lesquels conviennent au mieux à un problème particulier ? En d'autres termes, comment l'état de l'art est-il capable de résoudre un problème spécifique, ou lesquels des sous-problèmes peuvent être considérés comme résolus [Lopresti and Nagy, 2011, Lopresti and Nagy, 2012] ? Ce sont des questions essentielles, auxquels il est important de répondre si l'on veut faire de la recherche de façon efficace, quel que soit le domaine d'étude. Nous démontrerons que les réponses sont loin d'être triviales.

Répondre à ces questions conduit nécessairement à relier le concept général d'interprétation de données (principalement appliquée aux images de documents, pour des raisons pratiques, mais qui peut facilement être étendu à la perception artificielle en général) à celui d'analyse des performances et à la façon de rendre compte des résultats de recherche expérimentale. Nous avons déjà abordé ces points dans [23] et [41].

Les points développés dans [41], notamment, tout en considérant la question sous l'angle de l'analyse d'images de documents, s'appliquent facilement à la recherche plus large en perception artificielle et soulèvent les questions suivantes :

1. Comment peut-on évaluer objectivement des contributions individuelles à un problème de perception artificielle ? Peuvent-elles être comparées à des travaux antérieurs ? Peut-il y avoir un ensemble de critères mesurables pour établir si elles contribuent effectivement à l'amélioration de l'état de l'art ?
2. Dans quelle mesure ces contributions sont contraintes par un contexte spécifique d'utilisation ? Qu'est un contexte d'utilisation ? Peut-il être décrit, formalisé ou mesuré ?
3. Est-il réellement possible d'évaluer la performance d'une contribution à l'égard de la perception humaine ? Cela a-t-il un sens de le faire ? Que faudrait-il pour être en mesure de le faire ?

Ces questions sont faussement naïves de de bon sens, et, il semble qu'il n'existe pas de cadre établi pour les aborder. Elles semblent, en fait, être pris pour acquises et «évidentes». Nous allons montrer dans ce qui suit qu'elles sont loin de l'être, et que de les considérer à la légère peut conduire à introduire des biais dans les perceptions de la qualité de la recherche évaluée de plusieurs façons. C'est loin d'être un nouveau problème, et il a été considéré auparavant. Par exemple, une partie de ces questions peut être retracée aux origines de la recherche expérimentale et le besoin d'une traçabilité et une nécessité de pouvoir reproduire les résultats [21].

Pour d'atteindre ces considérations plus fondamentales, nous nous penchons sur l'analyse de la performance, et comment la recherche expérimentale doit être menée afin d'être valide, comme nous l'évoquons dans la section 3.3.

Le document est organisé comme suit : la section 5.2 commence par proposer quelques approches plus formelles pour aborder l'analyse de performance d'algorithmes de perception artificielle, et comment elles se rapportent aux limites des spécifications de la vérité terrain (et, comme nous le verrons, sa subjectivité inévitable). On démontrera, par conséquent, l'ambiguïté intrinsèque de l'interprétation et des méthodes développées dans l'état de l'art lorsqu'elles sont utilisées en conjonction avec la perception artificielle ; section 5.3 jette alors un regard plus approfondi sur les concepts d'interprétation et d'analyse, principalement en relation avec l'analyse d'images de documents, de sorte à ce que dans la section 5.4, nous pouvons introduire des approches permettant de modéliser les notions d'*interprétation* et de *contexte* et de les rapporter à l'analyse de performances. Étant donné que les arguments de cette thèse sont construits dans un ordre croissant de l'abstraction, nous commençons par un ensemble de considérations de base, en utilisant des définitions communément admises. Au fur et à mesure que les limites et les contradictions vont commencer à apparaître, la nécessité d'aborder plus formellement diverses questions va commencer à apparaître.

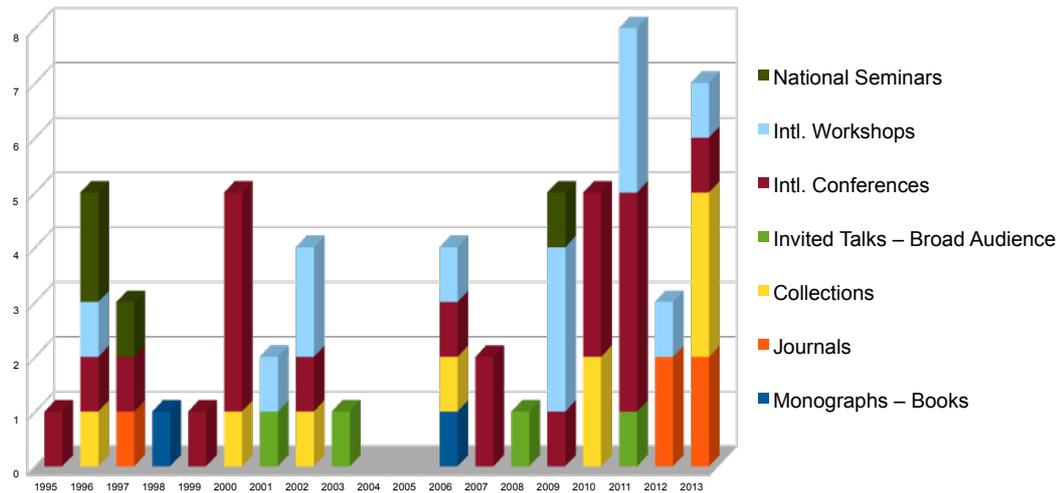
Le reste du chapitre 5 tentera de déterminer si on peut :

- établir une forme de description du contexte qui est approprié pour la perception artificielle (et l'analyse d'images de documents en particulier) et s'il peut être obtenu automatiquement par des techniques d'apprentissage statistique ou formel ?
- utiliser cette description du contexte pour évaluer les performances des algorithmes ?
- utiliser la description de contexte pour décrire les données, de sorte à ce que celui-ci puisse être utilisé pour pour orienter des tâches de recherche d'informations ?
- établir des limites ou des restrictions formelles pour les descriptions proposées précédemment et déterminer s'il existe des interprétations dont on peut prouver qu'il est impossible de les obtenir par un algorithme. S'il existe une classe de problèmes d'interprétation qui ne peut être résolu par un algorithme, la seconde question serait de savoir si cette classe peut être caractérisée en d'une façon ou d'une autre.



## 2.4 Publications

Overview of Publications: 1995 - 2013



### Referencing, Statistics and Rewards

53 of the following publications are referenced through *Harzing's Publish or Perish* using Google Scholar, and are cited 603 times. On CiteseerX, 24 are referenced with 79 citations (excluding self-citations). ISI Web of Knowledge references 17 (*Cited Reference Search*), with 90 citations (excluding self-citations).

- [21] received a special mention at the *Best Paper Awards* of the International Conference on Document Analysis and Recognition 2011, in Beijing, China, for potentially ground breaking impact on document image analysis research.
- [12] won the 2011 IAPR International Contest on Circle Detection, organized at the Ninth International Graphics Recognition Workshop in Seoul, Korea.

### Monographs and Books

- [1] "*Systèmes d'exploitation*", LAMIROY B., NAJMAN L., TALBOT H. Pearson Education France (Ed.), November 2006
- [2] "*Reconnaissance et modélisation d'objets 3D à l'aide d'invariants projectifs et affines*", B. LAMIROY (1998), thèse de doctorat de l'Institut National Polytechnique de Grenoble.

### International Journals

- [3] “*Integrating Vocabulary Clustering with Spatial Relations for Symbol Recognition*”, S. K.C., B. LAMIROY, Laurent WENDLING in International Journal on Document Analysis and Recognition, Springer Verlag, 2013.
- [4] “*DTW-Radon-based Shape Descriptor for Pattern Recognition*”, S. K.C., B. LAMIROY, Laurent WENDLING in International Journal of Pattern Recognition and Artificial Intelligence, World Scientific Publishing, 2013.
- [5] “*Relative Positioning of Stroke Based Clustering: A New Approach to On-line Handwritten Devanagari Character Recognition*”, S. K.C., C. NATTEE, B. LAMIROY International Journal of Image and Graphics, World Scientific Publishing, 2012
- [6] *Symbol Recognition using Spatial Relations* S. K.C., B. LAMIROY, L. WENDLING Pattern Recognition Letters, Elsevier, 2012, 33 (3), pp. 331-341
- [7] “*Object Pose: The Link between Weak Perspective, Paraperspective and Full Perspective*”, R. HORAUD, F. DORNAIKA, B. LAMIROY and S. CHRISTY (1997) in “International Journal of Computer Vision”, No. 2, pp. 173–189

### Selected Article Collections with Blind Review

- [8] “*Computing Precision and Recall with Missing or Uncertain Ground Truth*”, B. LAMIROY, T. SUN extended version of [49] in “Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16”, Revised Selected Papers, Lecture Notes in Computer Science 7423, Springer, pp. 149-162, Feb. 2013, Ogier, Jean-Marc; Kwon, Young-bin (Eds.)
- [9] “*Spatio-structural Symbol Description with Statistical Feature Add-on*”, S. K.C., B. LAMIROY, L. WENDLING, extended version of [48] in “Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16”, Revised Selected Papers, Lecture Notes in Computer Science 7423, Springer, pp. 228-237, Feb. 2013, Ogier, Jean-Marc; Kwon, Young-bin (Eds.)
- [10] “*Report on the Symbol Recognition and Spotting Contest*”, E. VALVENY, M. DELALANDRE, R. RAVEAUX, B. LAMIROY in “Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16”, Revised Selected Papers, Lecture Notes in Computer Science 7423, Springer, pp. 198-207, Feb. 2013, Ogier, Jean-Marc; Kwon, Young-bin (Eds.)
- [11] “*Dynamic Angle Based Theory in Learning Relative Directional Spatial Relationships on Components of Raster Symbols*”, S. K.C., L. WENDLING, B. LAMIROY extended version of [51] in “Graphics recognition: achievements, challenges, and evolution, Eighth IAPR International Workshop on Graphics Recognition, Selected Papers”, Lecture Notes in Computer Science, Ogier, Jean-Marc; Liu, Wenyin; Llados, Josep (Eds.) 2010, pp. 163–174, Springer.
- [12] “*Robust Circular Arc Detection*”, B. LAMIROY, Y. GUEBBAS, extended version of [50] in “Graphics recognition: achievements, challenges, and evolution, Eighth IAPR International

Workshop on Graphics Recognition, Selected Papers”, Lecture Notes in Computer Science, Ogier, Jean-Marc; Liu, Wenyin; Llados, Josep (Eds.) 2010, Springer<sup>12</sup>.

- [13] “*Scan-to-XML: Using Software Component Algebra for Intelligent Document Generation*”, B. LAMIROY and L. NAJMAN, in *Proceedings of the Fourth IAPR International Workshop on Graphics Recognition*, Springer-Verlag, Lecture Notes in Computer Science, 2002.
- [14] “*Visual Control Using Projective Kinematics*”, A. RUF, F. MARTIN, B. LAMIROY and R. HORAUD in John Hollerbach and Dan Koditschek (Eds.) *Robotics Research: the Ninth International Symposium*, Springer-Verlag, London, 2000
- [15] “*An Experimental Comparison of Appearance and Geometric Model Based Recognition*”, C. SCHMID, P. BOBET, B. LAMIROY and R. MOHR in Jean Ponce, Andrew Zisserman, Martial Hebert (Eds.) *Object Representation in Computer Vision II*, Springer-Verlag, Lecture Notes in Computer Science 1144, 1996

### Professional or Broad Audience Publications – Invited Communications

- [16] “*Document Analysis Algorithm Contributions in End-to-End Applications: Report on the ICDAR 2011 Contest*”, B. LAMIROY, D. LOPRESTI, T. SUN in 11th International Conference on Document Analysis and Recognition - ICDAR 2011, Sep 2011, Beijing, China. IEEE Computer Society
- [17] “*Pattern recognition methods for querying and browsing technical documentation*, TOMBRE K., LAMIROY B., in 13th Iberoamerican Congress on Pattern Recognition Progress in Pattern Recognition, Image Analysis and Applications Lecture Notes in Computer Science , Springer-Verlag, Lecture Notes in Computer Science, vol. 5197, pages 504-518 , 2008.
- [18] “*Graphics Recognition – form Re-engineering to Retrieval*”, K. TOMBRE and B. LAMIROY, in *Seventh International Conference on Document Analysis and Recognition*, Edinburgh, UK, 3-6 August 2003.
- [19] “*Combining Local Recognition Methods for Better Image Recognition*”, B. LAMIROY, P. GROS et S. PICARD in “*Vision*”, vol. 17, no. 2, The Society of Manufacturing Engineers (SME), Dearborn, Michigan, USA, 2001. (*reprint of [33] with minor language corrections*)

### International Conferences with Blind Review and Edited Proceedings

- [20] “*Relation Bag-of-Features for Symbol Retrieval*”, S. K.C., Laurent WENDLING, B. LAMIROY in Twelfth International Conference on Document Analysis and Recognition (ICDAR 2013), Aug 2013, Washington DC, United States.
- [21] “*An Open Architecture for End-to-End Document Analysis Benchmarking*”, B. LAMIROY, D. LOPRESTI in Eleventh International Conference on Document Analysis and Recognition (ICDAR 2011), oral, September 18 – 21, Beijing, China. (**Special Mention at Best Paper Awards**)

---

<sup>12</sup>**Winner of the Sixth International Arc Segmentation Contest**, held at the Ninth IAPR International Workshop on Graphics Recognition (GREC2011) Chung-Ang University, Seoul, South Korea.

- [22] “*DTW for Matching Radon Features: A Pattern Recognition and Retrieval Method*”, S. K.C., B. LAMIROY, Laurent WENDLING in Jacques Blanc-Talon and Richard P. Kleihorst and Wilfried Philips and Dan C. Popescu and Paul Scheunders. 13th International Conference on Advanced Concepts for Intelligent Vision Systems - ACIVS 2011, 6915, pp. 249-260, Ghent, Belgium. Springer
- [23] “*Document Analysis Research in the Year 2021*”, D. LOPRESTI, B. LAMIROY in Twenty-fourth International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE 2011), June 28 – July 1, Syracuse, NY.
- [24] “*How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research*”, B. LAMIROY, D. LOPRESTI, H. KORTH, J. HEFLIN in Document Recognition and Retrieval XVIII, IS&T/SPIE 23rd Annual Symposium on Electronic Imaging, 23-27 January 2011, San Francisco, CA USA.
- [25] “*Spatial Similarity based Stroke Number and Order Free Clustering*”, S. K.C., C. NATTEE, B. LAMIROY in 12th International Conference on Frontiers in Handwriting Recognition (ICFHR), Kolkata, India, November 2010.
- [26] “*Learning Spatial Relations for Graphical Symbol Description*”, K.C., L. WENDLING, B. LAMIROY in International Conference on Pattern Recognition, August 2010, Istanbul, Turkey.
- [27] “*Inductive Logic Programming for Symbol Recognition*”, S. K.C., B. LAMIROY, J-P. ROPERS in “International Conference on Document Analysis and Recognition”, poster, July 2009, Barcelona, Spain.
- [28] “*Robust Circle Detection*”, LAMIROY B., GAUCHER O., FRITZ, L., in 9th International Conference on Document Analysis and Recognition - ICDAR’07 , pages 526-530, oral, volume 1, 2007,
- [29] “*An Incremental On-line Parsing Algorithm for Recognizing Sketching Diagrams*”, MAS ROMEU J., SANCHEZ, G., LLADOS, J. LAMIROY, B., in 9th International Conference on Document Analysis and Recognition - ICDAR’07 , pages 452-456 , volume 1, 2007.
- [30] “*Automatic Adjacency Grammar Generator from User Drawn Sketches*”, J. MAS ROMEU, B. LAMIROY, G. SANCHEZ and J. LLADOS, in International Conference on Pattern Recognition, poster, pages 1026–1029, Hong Kong, 20–24 August 2006.
- [31] “*Scan-to-XML: Automatic Generation of Browsable Technical Documents*”, E. VALVENY and B. LAMIROY, in *Proceedings of the Sixteenth International Conference on Pattern Recognition*, poster, Québec City, Canada, 12-15 August 2002.
- [32] “*What Metric Stereo Can Do for Visual Servoing*”, B. LAMIROY, C. PUGET and R. HORAUD, “IEEE/RSJ International Conference on Intelligent Robots and Systems”, 30 October – 5 November 2000, Kagawa University, Takamatsu, Japan.
- [33] “*Combining Local Recognition Methods for Better Image Recognition*”, B. LAMIROY, P. GROS et S. PICARD, “The Eleventh British Machine Vision Conference”, 11–14 September 2000, Bristol, UK.
- [34] “*Visually Guided Robots for Ship Building*”, B. LAMIROY, T. DRUMMOND, R. HORAUD and O. KNUDSEN, “1st International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT’2000)”, March 30 – April 2, 2000, Potsdam/Berlin, Germany, pp. 262–275

- [35] “Controlling Robots with Two Cameras: How to Do it Properly”, B. LAMIROY, B. ESPIAU, N. ANDREFF and R. HORAUD, “IEEE International Conference on Robotics and Automation”, oral, April 24-28 2000, San Francisco, United States of America
- [36] “Visual Control Using Projective Kinematics”, A. RUF, F. MARTIN, B. LAMIROY and R. HORAUD, “9<sup>th</sup> International Symposium of Robotics Research”, October 1999, Snowbird, Utah, United States of America
- [37] “Object Indexing is a Complex Matter”, B. LAMIROY and P. GROS in « Proceedings of the 10th Scandinavian Conference on Image Analysis », oral, June 1997, Lappeenranta, Finland, Vol. I, pp. 277-283
- [38] “Rapid Object Indexing and Recognition Using Enhanced Geometric Hashing”, B. LAMIROY and P. GROS in « Proceedings of the 4th European Conference on Computer Vision », oral, April 1996, Cambridge, UK, Vol. 1, pp. 59-70
- [39] “Object Pose: Links Between Paraperspective and Perspective”, R. HORAUD, S. CHRISTY, F. DORNAIKA and B. LAMIROY in “Proceedings of the 5th International Conference on Computer Vision”, June 1995, Cambridge, Massachusetts, United States of America, pp. 426-433

### International Workshops with Edited Proceedings

- [40] “The Non-Geek’s Guide to the DAE Platform” Bart LAMIROY, Daniel LOPRESTI in DAS - 10th IAPR International Workshop on Document Analysis Systems, Mar 2012, Gold Coast, Queensland, Australia. IEEE, pp. 27-32.
- [41] “A Real-World Noisy Unstructured Handwritten Notebook Corpus for Document Image Analysis Research”, J. CHEN, Daniel LOPRESTI, Bart LAMIROY in Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data - (J-MOCR-AND 2011), Sep 2011, Beijing, China.
- [42] “A Platform for Storing, Visualizing, and Interpreting Collections of Noisy Documents”, B. LAMIROY, D. LOPRESTI in “Fourth Workshop on Analytics for Noisy Unstructured Text Data - AND’10”, Oct. 26, 2010, Toronto, Canada, col. ACM International Conference Proceeding Series.
- [43] “Automatic Learning of Symbol Descriptions Avoiding Topological Ambiguities”, MAS ROMEU J., LAMIROY B., SÁNCHEZ G., LLADÓS J, in 3rd Eurographics Workshop on Sketch-Based Interfaces and Modeling, pages 27–34, September 2006.
- [44] “A Few Steps Towards On-the-Fly Symbol Recognition with Relevance Feedback”, Jan Rendek, Bart Lamiroy and Karl Tombre, in *7th International Workshop, Document Analysis and Systems*, Nelson, New Zealand, Springer-Verlag, Lecture Notes in Computer Science, Volume 3872, January 2006, pp. 604–615.
- [45] “Text/Graphics Separation Revisited”, K. TOMBRE, S. TABBONE, L. PÉLISSIER, B. LAMIROY and Ph. DOSCH in Proceedings of the Fifth IAPR International Workshop on Document Analysis Systems, Princeton, NJ, USA, August 19-21 2002.
- [46] “An Image Oriented CAD Approach”, C. SCHMID, Ph. BOBET, B. LAMIROY and R. MOHR in Proceedings of the ECCV’96 International Workshop *Object Representation in Computer Vision*, Avril 1996, Cambridge, UK, pp. 221–245

**International Workshops without Review or Edited Proceedings**

- [47] “*Evaluation and the Semantic Gap ... What if we Were on a Side-Track?*”, B. LAMIROY Tenth IAPR International Workshop on Graphics RECOgnition - GREC 2013, Aug 2013, Lehigh University, Bethlehem, PA, United States of America.
- [48] “*Spatio-structural Symbol Description with Statistical Feature Add-on*”, S. K.C., B. LAMIROY, Laurent WENDLING Ninth IAPR International Workshop on Graphics RECOgnition - GREC 2011, Sep 2011, Seoul, Korea, Republic Of.
- [49] “*Precision and Recall Without Ground Truth*”, B. LAMIROY, T. SUN, Ninth IAPR International Workshop on Graphics RECOgnition - GREC 2011, Sep 2011, Seoul, Korea, Republic Of.
- [50] “*Dynamic Angle Based Theory in Learning Relative Directional Spatial Relationships on Components of Raster Symbols*”, S. K.C., L. WENDLING, B. LAMIROY in “Eighth IAPR International Workshop on Graphics Recognition”, July 2009, La Rochelle, France.
- [51] “*Robust Circular Arc Detection*”, B. LAMIROY, Y. GUEBBAS, in “Eighth IAPR International Workshop on Graphics Recognition”, July 2009, La Rochelle, France.
- [52] “*Assessing Classification Quality by Image Synthesis*”, B. LAMIROY, in “Eighth IAPR International Workshop on Graphics Recognition”, July 2009, La Rochelle, France.
- [53] “*Scan-to-XML for Vector Graphics: an experimental setup for intelligent browsable document generation*”, B. LAMIROY, L. NAJMAN, R. EHRHARD, C. LOUIS, F. QUÉLAIN, N. ROUYER and N. ZEGHACHE in Proceedings of the Fourth IAPR International Workshop on Graphics Recognition, Kingston, Ontario, Canada, September 7-8 2001.

**National Conferences or Colloquia**

- [54] “*Utilisation de Programmation Logique Inductive pour la reconnaissance de symboles*”, S. K.C., B. LAMIROY, J-P. ROPERS in “5ème Atelier ECOI : Extraction de COnnnaissance et Images”, GRCE, January 2009, Strasbourg, France
- [55] “*Indexation et recherche d’images*”, R. MOHR, P. GROS, B. LAMIROY, S. PICARD and C. SCHMID in “Actes du 16<sup>e</sup> colloque GRETSI sur le traitement du signal et des images”, September 1997, Grenoble, France
- [56] “*Computer Aided (dis)Assembly Using Visual Cues*”, B. LAMIROY, C. SCHMID, R. MOHR, M. TONKO, K. SCHÖFER and H.-H. NAGEL in “Proceedings of the IAR Annual Meeting”, November 1996, Karlsruhe, Germany
- [57] “*Reconnaissance d’objets par indexation géométrique étendue*”, B. LAMIROY and P. GROS in Journées ORASIS, May 1996, Clermont-Ferrand, France, pp. 19-24

**Films/Videos**

- [58] *Stereo Vision for Robot Control*, directed by Christian BLONZ, SICS – INRIA, scientific validation: Bart LAMIROY, Radu HORAUD et Bernard ESPIAU.

### Associated and Supervised Publications

Short list of selected works by Ph.D. students, directly related to previously cited work but of which I am not a co-author.

- [59] “*Character Recognition based on DTW-Radon*”, S. K.C., 11th International Conference on Document Analysis and Recognition - ICDAR 2011, Beijing, China, pp. 264-268, DOI: 10.1109/ICDAR.2011.61
- [60] “*Browsing Graphics Without Prior Knowledge*”, D. ZUWALA, J. RENDEK, International Conference on Pattern Recognition, Hong Kong, 20-24 August 2006, vol. 1, p. 735-738.
- [61] “*Extraction of Consistent Subsets of Descriptors Using Choquet Integral*”, J. RENDEK, L. WENDLING, International Conference on Pattern Recognition, Hong Kong, 20-24 August 2006, vol. 3, p. 208-211.
- [62] “*A New Method to Detect Arcs and Segments from Curvature Profiles*”, J.-P. SALMON, I. DEBLED-RENNESON, L. WENDLING, International Conference on Pattern Recognition, Hong Kong, 20-24 August 2006, vol. 3, p. 387-390.
- [63] “*Metadata for structured document datasets*”, H. F. Korth, D. Song and J. Hefflin, DAS '10: Proceedings of the 8th IAPR International Workshop on Document Analysis Systems, 547–550, ACM, New York, NY, USA, June 2010.

# Chapter 3

## Scientific Positioning

### Introduction

This chapter is traditionally devoted to positioning a career into its scientific environment and highlight the impact of its contributions. It also introduces the incubation context and the maturation of the scientific project developed in Chapter 5, with respect to the state-of-the-art and the evolution of knowledge in a particular domain.

This chapter is divided in three main sections, each of them corresponding to the domains in which I have been active during my career, essentially corresponding to my PhD., Post-Doc and faculty positions. They cover neither comparable periods of time nor comparable scientific maturity or even connex scientific domains. Each of them has, in its own ways, contributed to the corresponding state-of-the-art, but above all, has forged the conception of my current views on where some the current hard limits in Machine Perception lie, and what medium and long term investigations are needed to address them. The chapter will be limited to report the research done in these three domains and how it relates (or related) to the rest of the research community. Discussions on how to handle shortcomings and hard problems will be held in further chapters.

### 3.1 Image Recognition and Indexing

My PhD. was devoted to object recognition in real-world images from structural data. These data were indexed using affine and projective quasi-invariants, in order to absorb image deformations resulting from viewpoint changes. In a first time, the work only considered image contours [38,57], but was then progressively extended to geometrically coherent combinations of contours and interest patches [19,33,15,46]. Our approach was successfully tested in an industrial environment for the identification of mechanical parts of car engines [56]. One of the more fundamental parts of this work has been the

study of the inherent computational complexity related to content-based image indexing [37,55]. [37] has been one of the first published works noticing that indexing multimedia data imperatively needs to take into account computational complexity related to high dimensional metric spaces, and offering some workarounds.

### 3.1.1 Scientific Context and State-of-the-Art

Placing this work in its original context is very insightful, since it was done slightly before, and at the verge of a profound paradigm shift in Image Recognition and Indexing, and the explosion of Content Based Image Retrieval (CBIR), and more recently, Big Data in multimedia. The insightful part of the story is that much of this work was on the “wrong” side of the evolutionary scale.

The scientific context of this work is detailed in the first and second chapters of [2], and given the relative obsolescence of the results presented, we refer the interested reader to it, rather than to reformulate it here. We shall just give a synoptic overview of it.

The main focus on object recognition and image analysis was related to the still generally present works of D. MARR and I. BIEDERMAN [Biederman, 1981, Marr, 1982a, Biederman, 1985], in that there was a tight correlation between the recognition of objects in images and their 3D nature. This is what fueled much of the 3D vision activity at that time, and the fact that recognition was very much considered under a structural angle.

New work was starting to emerge, however, trying to look at Machine Perception under a completely 2D angle, by using appearance based models, and by considering visual structure from a purely numerical and statistical point of view [Sirovitch and Kirby, 1987, Turk and Pentland, 1991, Murase and Nayar, 1995, Schiele and Crowley, 1996] (opening the way to learning techniques that have flourished since) and not from a hierarchical 3D geometrical viewpoint.

The work conducted in this thesis can best be seen as a hybrid, transitional approach, in the sense that it is resolutely appearance based, but that it still relies on 2D edge-based descriptions of shapes and objects, and their overall geometric coherence, which links them to the previously mentioned paradigm of 3D geometrical descriptions of the world.

### 3.1.2 Highlights and Contributions

The highlights and contributions of this period are to have introduced one of the many competing *appearance based* recognition methods at that time (*e.g.* [Schmid and Mohr, 1996, Schiele and Crowley, 1996]) modeling segmented images by geometric configurations. By introducing a characterization of these configurations using *em* quasi-invariant we have implemented an indexing scheme that can effectively establish a mapping between local

configurations of an image and those in an indexed corpus. It raises some fundamental issues related to the organization and structure of efficient indexing spaces, and the notion of verification of overall consistency between plausible candidates among proposed models and the unknown image.

## Contributions

We distinguish three main contributions in this thesis.

1. Using *quasi-invariants* on extracted geometric configurations and for recognition had already been used by various authors [Lamdan and Wolfson, 1988, Gros, 1993, Rothwell, 1995, Nelson, 1996], but were considered too limited, because of their low descriptive power. We introduced a new indexing scheme combined with a global geometric consistency check, thus overcoming the limitations of the *quasi-invariants* and obtain a recognition method based on appearance that is exploitable in real-world situations [38,46,56].
2. Guided by the need to make our method more robust and extend its usability in broader contexts and in uncontrolled environments, we were interested in other approaches that also built upon local appearance and indexing (*e.g.* [Schmid and Mohr, 1996]). We have identified local geometric configurations as a common factor, and we have extended our method to other descriptors than only quasi-invariants. This integration has allowed us to propose a recognition method and which can be used in a broad range of situations, and beyond largely polyhedral models we dealt with initially [19,33].
3. By studying the limitations of our method, including resource requirements, we have shown formally that indexing, as used in many approaches at that time, is not always the answer. Our study shows that in cases where it is not necessary to take into account uncertainty around characteristics, indexing is a very powerful tool, significantly reducing the computational complexity of recognition when the size of descriptors is large. However, when it is necessary to introduce the notion of uncertainty in the indexing process, the complexity of problem is changed, and an indiscriminate increase in the size descriptors does not reduce the execution time in all cases [37, 55].

## Analysis and Impact

The impact of this work has been limited. On the one hand, the parts related to image recognition [38,19,33] have been proven much less efficient for recognition and retrieval than the subsequent interest key point approaches that were developed almost at the same time [Schmid et al., 2000, Lazebnik et al., 2003, Lowe, 2004], and that have ultimately given rise to the widespread visual bag-of-word approaches ([Sivic and Zisserman, 2003]

and all subsequent work stemming from it), currently almost universally adopted. On the other hand, the analysis in [37] was relatively ahead of its time, and the complexity issues related to it have been starting to emerge only a few years later [Amsaleg et al., 2000, Amsaleg et al., 2004, Sigurdardottir et al., 2005, Lejsek et al., 2006].

One of the reasons (and the lessons to be learnt) of this lack of success (especially concerning recognition) is the significant change of viewpoint between both approaches. Although I didn't quite realize this at the time (as the introductory part of [2] clearly shows), notwithstanding the fact that it was clearly an appearance based take to recognition, the fact of using contour structures and quasi-invariants, still anchored it as a natural evolution of the structural and hierarchical vision of image interpretation [Biederman, 1981, Marr, 1982a, Biederman, 1985], with the belief that recognition invariably results from some 3D perception of reality. The problem with those is that they have a scalability problem, if they want to capture all possible appearances and shape variations that can occur. Our approach managed to partially avoid this problem, but due to its lack of discriminant nature, reintroduced computational complexity with its global geometric coherence check. The competing approaches having taken a completely non-structural approach have been able to circumvent these issues for a while, and thus push the frontiers of the state-of-the-art. However, recent developments show there is an increasing interest in revisiting local geometric coherence [Jegou et al., 2008, Perd'och et al., 2009, Heath et al., 2010], a path we already started exploring more than a decade ago in [19,33].

### 3.1.3 Publications on this Topic

This is an excerpt of the full publication list, pp. 28–34.

#### Monographs and Books

- [2] *“Reconnaissance et modélisation d’objets 3D à l’aide d’invariants projectifs et affines”*, B. LAMIROY (1998), thèse de doctorat de l’Institut National Polytechnique de Grenoble.

#### Professional or Broad Audience Publications

- [19] *“Combining Local Recognition Methods for Better Image Recognition”*, B. LAMIROY, P. GROS et S. PICARD in “Vision”, vol. 17, no. 2, The Society of Manufacturing Engineers (SME), Dearborn, Michigan, USA, 2001. (*reprint of [33] with minor language corrections*)

#### International Conferences with Blind Review and Edited Proceedings

- [33] *“Combining Local Recognition Methods for Better Image Recognition”*, B. LAMIROY, P. GROS et S. PICARD, “The Eleventh British Machine Vision Conference”, 11–14 September 2000, Bristol, UK.
- [37] *“Object Indexing is a Complex Matter”*, B. LAMIROY and P. GROS in « Proceedings of the 10th Scandinavian Conference on Image Analysis », oral, June 1997, Lappeenranta, Finland, Vol. I, pp. 277-283

- [38] “*Rapid Object Indexing and Recognition Using Enhanced Geometric Hashing*”, B. LAMIROY and P. GROS in « Proceedings of the 4th European Conference on Computer Vision », oral, April 1996, Cambridge, UK, Vol. 1, pp. 59-70
- [39] “*Object Pose: Links Between Paraperspective and Perspective*”, R. HORAUD, S. CHRISTY, F. DORNAIKA and B. LAMIROY in “Proceedings of the 5th International Conference on Computer Vision”, June 1995, Cambridge, Massachusetts, United States of America, pp. 426-433

#### International Workshops with Edited Proceedings

- [46] “*An Image Oriented CAD Approach*”, C. SCHMID, Ph. BOBET, B. LAMIROY and R. MOHR in Proceedings of the ECCV’96 International Workshop *Object Representation in Computer Vision*, Avril 1996, Cambridge, UK, pp. 221–245

#### National Conferences or Colloquia

- [55] “*Indexation et recherche d’images*”, R. MOHR, P. GROS, B. LAMIROY, S. PICARD and C. SCHMID in “Actes du 16<sup>e</sup> colloque GRETSI sur le traitement du signal et des images”, September 1997, Grenoble, France
- [56] “*Computer Aided (dis)Assembly Using Visual Cues*”, B. LAMIROY, C. SCHMID, R. MOHR, M. TONKO, K. SCHÖFER and H.-H. NAGEL in “Proceedings of the IAR Annual Meeting”, November 1996, Karlsruhe, Germany
- [57] “*Reconnaissance d’objets par indexation géométrique étendue*”, B. LAMIROY and P. GROS in Journées ORASIS, May 1996, Clermont-Ferrand, France, pp. 19-24

#### Reports

- [64] “*Reconnaissance d’objets polyédriques à l’aide d’invariants projectifs*”, B. LAMIROY, Rapport de DEA, 1994.
- [65] “*Mise en correspondance dense de deux images par corrélation*”, B. LAMIROY, Rapport de Magistère, 1993.

## 3.2 Visual Servoing

As already mentioned p. 13, this part only covers two years in my research career and is a brief incursion into the visual servoing domain, during which I was in charge of the coordination of the European Commission funded VIGOR project of which our research group was lead contractor.

My main scientific contributions concerned research on metric and non metric visual stereoscopic control and have have extended the state-of-the-art in visual servoing from calibrated monocular control to uncalibrated stereo control by integrating epipolar constraints [35] based on previously published results [7,39]. The approach has then been validated on real-world data on-site at the Odense Steel Shipyard, Denmark, in the case of visually servoed welding [32].

In parallel, work with A. RUF and F. MARTIN led to a fully un-calibrated stereo control method [14,36].

### 3.2.1 A Genuine Industrial Implementation Context

One of the challenging parts of this work is that much of it has been applied and transferred to an industrial setup at Odense Steel Shipyard, Ltd. (OSS) ship welding facilities where 11 degree of freedom robots undertake tasks on manufactured workpieces. On the one hand, this kind of tasks are usually thoroughly modeled and tested within the virtual CAD environments of the workpiece and the robot. On the other hand, directly using the CAD model for programming the real task is far from desirable or sometimes not even directly transposable. The reasons for this are that :

1. The workpieces may considerably differ from the initial CAD model, due to accumulated errors during the assembly process, combined with the large scale of the objects with respect to the robot, or even due to simple phenomena like thermal expansion related to varying climatic conditions.
2. One is not guaranteed that the workpiece arrives in the CAD model predicted position with respect to the robot. This is mainly due to the size and inertia of the workpieces that need to be handled.
3. In the Odense Steel Shipyard, Ltd. (OSS) context, every operation is one-of-a-kind (unlike what happens in the car industry, or any other mass production scheme, for that matter). This means that off-line operations are not profitable since they cannot be factored out to speed up numerous repetitive on-line operations.

Therefore, one easily sees the need for a mechanism that needs to offer the following services :

- rapidly detect, locate and track the workpiece within the work area of the robot [Drummond and Cipolla, 1999a, Drummond and Cipolla, 1999b];

- be extremely robust to discrepancies between the reality and the CAD model (possibly highlighting the regions where the differences are quite noticeable), as well as occlusions, movement, lighting conditions and physical setup parameters [32];
- securely drive the robot end-effector to its programmed position with the required precision for executing its task and execute the latter under constant guidance [35].

### 3.2.2 Scientific Context and State-of-the-Art<sup>13</sup>

Our work addresses the problem of visually controlling a robot with two cameras. Unlike previous approaches, it explicitly handles stereo vision by embedding the epipolar constraint [Longuet-Higgins, 1981, Luong and Faugeras, 1996] directly into the control law and therefore has theoretically provable behavior. We established, for instance, that, unlike the single-camera case, there is NO control singularity [35].

The epipolar constrained visual servoing method has direct practical implications, in particular in situations where one of the two cameras is partially or totally occluded. We suggest an approach, namely *virtual stereo servoing* that allows a smooth trajectory even when the left or the right cameras have image processing problems.

The overall advantage of these studies was to establish, for the first time, the theoretical and operational bases for robust and safe visual control of robots, with guaranteed behavior in 3D. The 3D trajectory executed by the robot under monocular or stereo servoing is fundamentally. In both cases, the expected image trajectory (2D) is usually a straight line. However, in the monocular case, there is no guarantee that this straight 2D line translates into a straight 3D trajectory, since any planar curve can project into a line. On the other hand, the 3D trajectory in the stereo case projects into a line on either image, imposing *de facto* a straight line movement in space (except for extreme cases, where the trajectory would lie in the focal plane).

We address visual servoing from the viewpoint of stereo vision. Existing methods [Maru et al., 1993, Hager et al., 1995, Hager, 1997, Chang et al., 1997], were based on the stacking of monocular servo image Jacobian matrices, resulting in control commands with no specific geometric constraints in 3D. Our contributions formally establish that the epipolar constraint between two images can be taken into account explicitly [35] and that the use of stereo vision significantly increases the quality of the task execution, especially where precision, robustness and smoothness of movement is concerned [32].

Existing methods of visual servoing using one or multiple cameras were based on the stacking of monocular Jacobian matrices. In this context, visual servoing is considered as a task function [Samson et al., 1990]. Several approaches of robot control using a single camera [Espiau et al., 1992, Horaud et al., 1998] or stereo rigs [Maru et al., 1993,

---

<sup>13</sup>This state-of-the-art overview is limited to the period in time (1998-2000) when this research was conducted. Since I did not pursue work in this domain beyond that period, it does not make much sense to pretend having a keen sense of its evolutions *post partem*.

[Hager et al., 1995, Hager, 1997, Chang et al., 1997] pre-existed ours. In the case where only a single camera is concerned, a certain number of operational conditions, such as camera calibration, were required. In that case, however, the problem is sufficiently constrained and a minimal set of control variables can be used. However, one has to bear in mind that a number of singularities exist, making visual control difficult near those configurations (*e.g.* 180 degree rotations [Chaumette, 1997]).

The use of a stereo rig avoids control singularities associated with a single camera, and requires less domain knowledge. The cited approaches are restricted to particular cases of stereo vision, however, where cameras are non verged. Although verged or non verged cameras are theoretically equivalent [Brown, 1992], in practice, and more particularly where visual servoing is concerned, there are important drawbacks. The formal framework we have proposed for using stereo servoing in a general case has shown that the epipolar constraint can be taken into account for all camera configurations and that stereo control is inherently more robust than monocular servoing.

### 3.2.3 Highlights and Contributions

We have addressed the problem of visually controlling a robot with two cameras. The vast majority of visual servoing methods used one camera or combined several cameras by stacking together the single camera case. We formally introduced the geometric (epipolar) constraint into the study and the design of the control law. We showed that there are no intrinsic control singularity problems, thus allowing stereo-based control to be a suitable setup for many practical configurations. We introduced the concept of virtual stereo servoing, allowing one of the two cameras to temporarily lose signal, due to occlusions during robot task execution.

Besides establishing a theoretical framework that has been extended to take into account more complex correlated multi-sensor configurations [Cervera et al., 2003, Hynes et al., 2006, Pari et al., 2008, Sebastián et al., 2009], our experiments have shown that the approach offers the following operational advantages:

- Image trajectories are more stable under stereo servoing.
- Although image convergence is better under monocular servoing, the 3D positioning is twice more precise when using stereo servoing than when using monocular servoing (average error for stereo is half the one for mono with a standard deviation an order of magnitude less).

In practice we observed that when the cameras are at 3 to 5 meters away from the robot the gain in precision (between monocular and stereo servoing) is of a factor of 10, thus dropping the positioning of the robot tool relative to a ship part from 1cm (one camera) to 1mm (stereo).

- The kinematic screw is less erratic in the case of stereo servoing.

- *Virtual stereo* servoing behaves extremely smoothly in presence of signal loss during servoing. The image trajectories are smoother than in the monocular case, and the overall system performance at convergence is preserved even in presence of rough calibration.

### 3.2.4 Publications on this Topic

This is an excerpt of the full publication list, pp. 28–34.

#### International Journals

- [7] “*Object Pose: The Link between Weak Perspective, Paraperspective and Full Perspective*”, R. HORAUD, F. DORNAIKA, B. LAMIROY and S. CHRISTY (1997) in “International Journal of Computer Vision”, No. 2, pp. 173–189

#### Selected Article Collections with Blind Review

- [14] “*Visual Control Using Projective Kinematics*”, A. RUF, F. MARTIN, B. LAMIROY and R. HORAUD in John Hollerbach and Dan Koditschek (Eds.) *Robotics Research: the Ninth International Symposium*, Springer-Verlag, London, 2000

#### International Conferences with Blind Review and Edited Proceedings

- [32] “*What Metric Stereo Can Do for Visual Servoing*”, B. LAMIROY, C. PUGET and R. HORAUD, “IEEE/RSJ International Conference on Intelligent Robots and Systems”, 30 October – 5 November 2000, Kagawa University, Takamatsu, Japan.
- [34] “*Visually Guided Robots for Ship Building*”, B. LAMIROY, T. DRUMMOND, R. HORAUD and O. KNUDSEN, “1st International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT’2000)”, March 30 – April 2, 2000, Potsdam/Berlin, Germany, pp. 262–275
- [35] “*Controlling Robots with Two Cameras: How to Do it Properly*”, B. LAMIROY, B. ESPIAU, N. ANDREFF and R. HORAUD, “IEEE International Conference on Robotics and Automation”, oral, April 24-28 2000, San Francisco, United States of America
- [36] “*Visual Control Using Projective Kinematics*”, A. RUF, F. MARTIN, B. LAMIROY and R. HORAUD, “9<sup>th</sup> International Symposium of Robotics Research”, October 1999, Snowbird, Utah, United States of America

#### Films/Videos

- [58] *Stereo Vision for Robot Control*, directed by Christian BLONZ, SICS – INRIA, scientific validation: Bart LAMIROY, Radu HORAUD et Bernard ESPIAU.

#### Registered Software<sup>14</sup>

- [66] `libTracking` software library for target tracking implementation. APP registration reference: IDDN.FR.001.450016.00.R.O.1999.000.00000.
- [67] `libServo` stereoscopic visual servoing implementation, based on [66]. APP registration reference: IDDN.FR.001.080011.00.R.P.2000.000.00000.

#### Reports

---

<sup>14</sup>APP : Agence pour la Protection des Programmes

- [68] “*Requirements for an open TCP Speed and Position Control interface*”, B. LAMIROY and T. DRUMMOND, Technical Specifications, VIGOR, Esprit-IV reactive LTR project, number 26247, 19 octobre 1999, INRIA Rhône-Alpes and University of Cambridge.
- [69] “*Description of Demonstrators, INRIA Uncalibrated Visual Servoing Prototype*”, B. LAMIROY, Public Deliverable D4.1.c, VIGOR, Esprit-IV reactive LTR project, number 26247, 1 mars 1999, INRIA Rhône-Alpes.
- [70] “*Year 1 Demo, Preliminary Specs*”, B. LAMIROY and C. GRAMKOW, Restricted Deliverable D1.3.a, VIGOR, Esprit-IV reactive LTR project, number 26247, 14 septembre 1998, INRIA Rhône-Alpes, Odense Steel Shipyard, Ltd. (OSS).

### 3.3 Graphical and General Document Analysis

As already detailed previously (*cf.* pp. 14-16), my research activity switched to graphical document image analysis in 2000, when integrating the QGar team at the INRIA Nancy Grand-Est/LORIA.

The three major research themes developed during this period, are “*Scan-to-XML and Semantic Structuring*”, “*Indexing and Recognition*” and “*Performance and Interpretation Analysis*”. While the first two themes could be rated as traditional and mainstream, the last one is more difficult to qualify, and will receive most of the focus on this section.

The research conducted during this period has been almost exclusively within an overlapping network of international collaborations: with the QGar team in Nancy, obviously, but also with the CVC lab at the Universitat Autònoma de Barcelona, the IAPR TC-10 and TC-11 communities, the European Strep FRESH project members, the City University of Hong-Kong, and, finally, but not the least, Lehigh University; be it through the exchange of students, short or extended stays, or visiting positions.

This period also corresponds to a progressively increasing implication and recognition in the international document image community, through the participation in PhD. defense committees abroad (J. MAS at the CVC, Barcelona), program committee duties for international conferences (GREC, ICDAR), general chair functions (GREC 2013) and active implication as board member in one of the technical committees and one of the standing committees of the International Association for Pattern Recognition (IAPR).

#### 3.3.1 Scientific Context and State-of-the-Art

Graphic Documents are basically documents containing a significant (if not exclusive) part of line drawings. They are usually considered a separate class between full text documents and photo-realistic documents. Many applications or research areas related to the interpretation of documents in general, usually consider it good practice to segment composite documents into at least text, line drawings and photo-realistic sub-parts, to which more specialized treatments are then applied. Many applications within graphic document image analysis revolve around schematic representations (maps, blueprints, wiring diagrams ...) that bear an intrinsic meaning, expressed in a visual language. Hence, these representations are often referred to as “*semantics*”.

The larger part of the work conducted in this domain, consists, in various ways, to extract or capture these semantics. Our own contributions have explicitly stated this goal on multiple occasions [13,53,31,45]. However, since graphical document image analysis falls within the broader domain of Machine Perception, it suffers from what is called the Semantic Gap [Smeulders et al., 2000]. One of its effects on the document image analysis domain, is a constant back-and-forward movement between the extraction of lower

level (considered “*syntax*”) image analysis, and higher level interpretations (considered “*semantics*”)

Our own work did not escape from that movement: our work Inductive Logic Programming [27,52,54] was an attempt to model a certain class of interpretation contexts, as were some of the structural approaches for hand-drawn symbol recognition [29,30,43]. But since they needed robust data to work with (especially the ILP framework) we also investigated the extraction of “syntactical” graphical entities [28,62,51] or relations [50], for instance.

Our initial idea was to develop a complete framework, based on an extracted vocabulary elements [28,51,62] that would assess the quality of symbol descriptions using spatial relationships [50], express them in a formal description language [54,27,52] and allow to deploy machine learning for complex symbol descriptions.

Establishing a complete comparative state-of-the-art of our work, would probably not contribute very much to the debate, and can, in part, be found in the referenced papers. Furthermore, establishing this study would also very likely neither establish the pertinence of our work. The reason for this is the core of our current focus of investigation: the “*Performance and Interpretation Analysis*” research theme.

To explain and establish the rationale behind this apparent counter-intuitive assertion that establishing a state-of-the-art review has limited use, we shall use the examples of some of our previously cited work. In the context of our work done on symbol recognition (essentially due to S. K.C. during his Ph.D. [6,9,22]) for instance, we have established, using extensive experimental verification, that it outperforms the most reputed state-of-the-art symbol recognition approaches, like global signal region based descriptors: *Zernike* [Kim and Kim, 2000], *Generic Fourier Descriptors* [Zhang and Lu, 2002], *Shape Context* [Belongie et al., 2002] and *R-signature* [Tabbone et al., 2006], or pixel-based approaches: Statistical Integration of Histogram Array (SIHA) [Yang, 2005] and 2D kernel density [Zhang et al., 2006]. The main question, however, remains to define what the term “*outperform*” actually means. Concerning symbol recognition, the generally admitted consensus is related to expressing recognition rates or precision/recall [van Rijsbergen, 1979] measures over shared datasets. This is what was used in our published work, and this is the main reason why the research community is actively involved in the organization of international contests [Valveny and Dosch, 2004, Tombre et al., 2005, Dosch and Valveny, 2005, Dosch et al., 2008] and/or distribution of reference datasets. As we shall develop in Chapter 5, this dependency on the specific context within which the evaluation takes place (*e.g.* the contest), is never completely explicit and always contains a non-negligible part of arbitrary choice. As a consequence, although these widely accepted protocols give an indication of quality of the approaches under comparison for the datasets used, it is not sure whether the increase in efficiency or in “quality” reported in the various publications, based on the observation of performance on fixed data, actually allows to draw any conclusions on intrinsic quality of each of them.

An excellent illustration of the limitations of these limitations can be found in our work

on circle detection [28,12], which won the 2011 IAPR International Contest on Circle Detection [Al-Khaffaf et al., 2012], organized at the Ninth International Graphics Recognition Workshop in Seoul, South Korea, but for which many competing approaches exist [Hilaire and Tombre, 2006, Ayala-Ramirez et al., 2006, Chen and Chung, 2001, Cheng and Liu, 2003, Hanahara and Hiyane, 1991, Chiu and Liaw, 2005, Wenyin, 2006].

The Nancy QGar team, in which this research was done, has a record of successes in vectorization and segmentation, and [28] was directly inspired by previous work done there [Hilaire and Tombre, 2006]. Since [Hilaire and Tombre, 2006] won the 2005 edition of the IAPR International Contest on Circle Detection, [28], building on a comparable approach, had been noticed by the 2007 edition organizers. On their invitation, it competed unofficially (after the contest was closed [Shafait et al., 2008]) and obtained more than decent results (unpublished, since after the deadline, but giving rise to a collaboration with Prof. Liu and stay at Hong Kong City University in 2007). The final version [12] competed in full in both the 2009 and 2011 editions [Al-Khaffaf et al., 2010, Al-Khaffaf et al., 2012].

The most important impact of this work (especially with respect to future research directions) stems from this double participation of the exact same algorithm in two successive editions of the same contest. In 2009 [12] was ranked third or fourth [Al-Khaffaf et al., 2010], in 2011 it came in first [Al-Khaffaf et al., 2012], which seems quite contradictory<sup>15</sup>. On the one hand this raises some fundamental questions on the scientific foundations of "state-of-the-art" published results and how to interpret them when based on reporting with respect to standard data sets, and how a research community as a whole can benefit from analyzing in greater detail the implications of using contests and reference data sets for state-of-the-art evaluation<sup>16</sup>. On the other hand, it calls for tools to better take into account legacy results and confront them to both evolutions of the state-of-the-art knowledge and data or interpretation contexts. Especially considering the inflation of initiatives trying to evaluate results based on organized contests: at ICDAR 2011 [DBL, 2011] alone, eighteen different, and sometimes overlapping contests were proposed: Arabic Handwriting Recognition, CROHME: Competition on Recognition of Online Handwritten Mathematical Expressions, Arabic Writer Identification Contest, Book Structure Extraction, Page Dewarping, Arabic Recognition Competition: Multi-font Multi-size Digitally Represented Text, SigComp: Signature Verification Competition for On- and Offline Skilled Forgeries, DIBCO: Document Image Binarization Contest, Writer Identification Contest, Table competition, AHTR: Arabic Handwritten Textline Recognition, Music Scores: Writer Identification and Staff Removal, Online Arabic Handwriting Recognition, Historical Document Layout Analysis, Specific Document Analysis Algorithm Contributions in End-to-End Applications, French Handwriting Recognition, Robust Reading, Chinese

---

<sup>15</sup>When investigating further, it still is awkward, but can be explained both by the fact that all other competitors were either new, and had never participated before, or may have been "overtrained" on specific data, combined with the fact that the evaluation data had changed between the two editions, including higher resolution scans than before.

<sup>16</sup>Some of the arguments developed here have already been given (separately or partially) in our more recently published works [21,23,24,40,41,42].

Handwriting, Interactive Handwriting.

A thorough review of most published work in general, shows that there is a fundamental lack of reproducibility and traceability of the reported experiments for various reasons: paper size restrictions limit the level of thorough description of the algorithms, of used parameters and of the complete experimental protocol; data sets may be unavailable due to copyright restrictions; downloadable code often is very platform dependent, and rapidly becomes obsolete ... organization of evaluation campaigns in the Document Analysis community [Grosicki et al., 2009, Geoffrois, 2009], or in the broader Machine Perception domain [Garris and Klein, 1998, Müller et al., 2010, Smeaton et al., 2006], reference datasets and contest only partially address the global problem performance evaluation and do not offer any answer to how to reproduce or trace the results of the reported experiments. Furthermore, they require a significant amount of effort and resources to maintain. This problem is not contained to the sole domain of Document Image Analysis, and various other domains where computational experimental research is conducted are confronted to the same issues [Gent and Kotthoff, 2011]. Examples of tentatives of addressing the problem are numerous. For instance,

- AMIC@ (<http://bioalgo.iit.cnr.it/amica>) is an initiative from the BioAlgo group at the Institute of Informatics and Telematics (IIT) in Pisa, in collaboration with the Department of Computer Engineering of the University of Pisa, and the Department of Computer Engineering of the University of Modena and Reggio Emilia. It stands for “All Microarray Clusterings @ once”. The description on its website defines it as “a web application aiming to provide users with a common user interface to a wide range of microarray gene expression data clustering algorithms.” that “allows [...] to run several algorithms (and different configurations of them) on the same data set, view all the resulting clusterings on-line [...], view the clustering homogeneity, visualize de expression level of each heat map cell, download the outcome of the clustering activities as a standard clustered data files [...], or the hierarchical dendrogram [... and ] allows you to execute simultaneously all the algorithms on a given data set”.
- EvaluatIR.org [Zobel et al., 2011, Armstrong et al., 2009] (University of Melbourne – <http://wice.csse.unimelb.edu.au:15000/evalweb/ireval/>) is targeted toward Information Retrieval and provides an online reference and analytical tool for retrieval effectiveness results. They start from the observation that up-to-date and comparable benchmarks for collections like TREC are often difficult to find. They claim to allow people to:
  - see what the state of the art in retrieval access is, using our database of evaluation results. It includes past retrieval runs submitted to TREC, benchmarks of out-of-the box IR systems and runs published by other researchers. Information about how the results were obtained will be collected and documented wherever possible;
  - privately upload their own runs in TREC runfile format;

- compare any results in our database with each other or your private uploaded runs – in a range of different ways. A range of different evaluation metrics are available, and results can be compared graphically query-by-query, using statistical significance tests, ranked against other runs, using score standardisation, and more;
  - share their uploaded runs with other researchers, and link to back to relevant publications.
- Galaxy [Goecks et al., 2010] (<http://galaxy.psu.edu>) at the Center for Comparative Genomics and Bioinformatics at Penn State, and the Biology and Mathematics and Computer Science departments at Emory University defines itself as “an open, web-based platform for accessible, reproducible, and transparent computational biomedical research. **Accessible:** Users without programming experience can easily specify parameters and run tools and workflows. **Reproducible:** Galaxy captures information so that any user can repeat and understand a complete computational analysis. **Transparent:** Users share and publish analyses via the web and create Pages, interactive, web-based documents that describe a complete analysis.”
  - IPOL [IPOL, ] (<http://www.ipol.im>) is a journal of image processing and image analysis. Each article contains a text describing an algorithm, a source code, an online execution facility, and an archive of all online experiments. The text, source code and demonstration are peer-reviewed.
  - Kaggle (<http://www.kaggle.com>), as the company defines itself, “is an arena where you can match your data science skills against a global cadre of experts in statistics, mathematics, and machine learning [...] [it] is a platform for data prediction competitions that allows organizations to post their data and have it scrutinized [...] In exchange for a prize, winning competitors provide the algorithms that beat all other methods of solving a data crunching problem. Most data problems can be framed as a competition.”
  - RabbitCT [Rohkohl et al., 2009] (<http://www.rabbitct.com>) is a collaboration of the Department of Neuroradiology and the Pattern Recognition Lab at the Friedrich-Alexander-University Erlangen-Nuremberg and ensues from the observation that many publications in the domain of 3D reconstruction of medical images “are not comparable, mainly due to variations in data acquisition, preprocessing, chosen geometries, and the lack of a common publicly available test dataset.” The project consists in “[providing] an open platform for worldwide comparison in back projection performance and ranking on different architectures using one specific, clinical, high resolution C-arm CT dataset of a rabbit [...] [including a] benchmark interface, a prototype implementation in C++, and image quality measures.”
  - Re3data (<http://www.re3data.org>) is a German DFG funded academic initiative that, as stated on their website “is to create a global registry of research data

repositories. The registry will cover research data repositories from different academic disciplines. re3data.org will present repositories for the permanent storage and access of data sets to researchers, funding bodies, publishers and scholarly institutions. In the course of this mission re3data.org aims to promote a culture of sharing, increased access and better visibility of research data. In the first phase of the project the following tasks are prioritized: the conception and construction of a web-based registry of research data repositories, the definition of selection criteria of research data repositories, the formulation of a metadata schema to describe research data repositories.”

- TunedIT [Wojnarski et al., 2010] (<http://tunedit.org>) is a start-up company from the Academic Technology Incubator of the University of Warsaw, focused on providing web-based tools to data mining scientists for conducting repeatable experiments and easily evaluating data-driven algorithms. They have developed a platform for hosting data competitions and try to offer services connecting academia with industry interested by data mining. The concept consists un uploading data sets, or algorithms to a centralized platform, thus allowing to replay existing experiments, on the same or on different data, or compare results of algorithms with previous executions. Il also allows to review and comment *etc.*. It also offers services for hosting, running and analyzing complete contest setups.
- Zenodo (<http://zenodo.org>) is hosted at the CERN facilities, and is one of the outcomes of the FP7 OpenAire (<http://www.openaire.eu/>) infrastructure. As their website mentions, “ZENODO builds and operate a simple and innovative service that enables researchers, scientists, EU projects and institutions to share and showcase multidisciplinary research results (data and publications) that are not part of the existing institutional or subject-based repositories of the research communities. ZENODO enables researchers, scientists, EU projects and institutions to: easily share the long tail of small research results in a wide variety of formats including text, spreadsheets, audio, video, and images across all fields of science; display their research results and get credited by making the research results citable and integrate them into existing reporting lines to funding agencies like the European Commission; easily access and reuse shared research results.”

It is within this global mindset that our latest work is to be considered.

**The DAE platform,** developed during my position as visiting scientist at Lehigh University, and its related publications [23,24,40,41,42] investigate the interactions between reproducible and accountable experimental research, and the tools that would be needed to accomplish these goals.

It starts from the observation of how an ideal research community should be structured. On the one hand: the community, made of collective knowledge, state-of-the-art, claims, methods and algorithms. On the other hand: individuals proposing new ideas, new claims and methods not yet adopted or acknowledged by the community. In order to stake their

claims, the individuals need to know they perform and compare to the collective state-of-the-art knowledge. In reaction, the community needs to be able to assess the legitimacy of the proposed claims, and be assured that they are verifiable and reproducible.

In [23] we introduce a fictional character, Jane, a young starting researcher looking to solve a specific knowledge extraction problem. She typically needs to find appropriate experimental data, prove the generality of her proposed solution or establish boundaries, compare her results to the state-of-the-art and, finally, report her results.

The current consensus on how this should be done is through peer reviewed publications, use of code repositories and reference data collections. The first, however, poses a problem related to re-implementation issues for various, and often legit and unavoidable reasons related to actual claim verification and code re-implementation. Code repositories partially address these issues, but are technology dependent and therefore subject to obsolescence and may infringe on specific intellectual property and Copyright issues. Data collections have a very high maintenance cost, are usually conceived for a very specific interpretation context and tend not to evolve over time, therefore progressively diverging from the constantly evolving community focus.

The approach we develop [24,42] consists in offering a community-based platform that allows to formalizing experiments and their environment. It contains all required experimental data on the one hand, and ways to describe it, distribute it and question, dispute, challenge and extend it (*e.g.* [41]). On the other hand, it also offers the same services for formalizing experiments and to reproduce them both in fully controlled and reproducible environments *and* in new, yet unexplored contexts or with new data, allowing them to be questioned, challenged, adapted and extended [21].

The main differences between the proposed solution and other existing approaches, like the ones described previously, or the EU FP-7 IMPACT (Ref: 215064) project, which has adopted a extremely similar technological approach to a comparable, yet fundamentally different problem, are that: the environment, specifications and formats are completely open; datasets are fully expandable and reusable in other contexts then the ones they were initially conceived for; moreover, they can be combined with other data without additional cost or burden; data allows multiple interpretations; all data is formalized and typed (with the possibility to add new types) , associated with its full provenance and can be queried using standard SQL or SPARQL.

Furthermore, not only is the raw experimental data stored, but all interactions and results, interpretations and modifications are registered in such a way that all experiments become traceable, reproducible and can be analyzed; even in new contexts. This is being made possible by the use and referencing of all algorithms as being integral part of the model, and being made accessible as web-services, operating in a controlled and reproducible environment. To the best of our knowledge, there is no comparable, existing work in this domain that groups all these features.

**The organization of a new kind of contest** [16] constitutes a scale test of the paradigm the DAE platform is supposed to support. Organized as one of the multiple contests of ICDAR 2011 (*cf. supra*) it distinguishes itself by the fact that

- it measures the impact of individual, partial contributions, on a complete document analysis process: contestants can compete by providing only one single sub-process of the whole pipeline. This guarantees that focused research can be measured in actual complete application contexts, without the need for the contestants to be concerned by integration issues;
- the whole contest environment is open and fully reproducible: at any point in time, the contest can be re-run on the same data, with the same algorithms or by adding other algorithms or by replacing the experimental data.
- all results are logged and archived, and are available for further analysis and scrutiny so that the conclusions of the contest can be challenged or re-used in other contexts.

Here again, we are not aware of similar performance evaluation setups.

**The development of new evaluation metrics** [8] provides new tools for exploiting the data produced in the previously described setups, in such a way that it can take into account absence, inaccuracy or even competing ground-truth information such that the same experimental data can be studied in different interpretation contexts.

### 3.3.2 Highlights and Contributions

While our work has received decent credit from the research community, and we obtained some nice achievements with our work on circle detection [28,51], or symbol recognition [6,9], the most influential work (at least, potentially influential) is our most recent research on performance and interpretation analysis.

We have established and validated a completely new approach to experimental research validation and reporting for Machine Perception (although it is currently only applied to Document Image Analysis). This work is not only of theoretical or hypothetical nature, as described in the referenced publications [23,24,42], it has actually been deployed and used in large scale evaluation contexts [21,16,40,41] and was recognized as groundbreaking at the ICDAR 2011 best paper award ceremony.

### 3.3.3 Publications on this Topic

This is an excerpt of the full publication list, pp. 28–34.

#### International Journals

- [3] “*Integrating Vocabulary Clustering with Spatial Relations for Symbol Recognition*”, S. K.C., B. LAMIROY, Laurent WENDLING in *International Journal on Document Analysis and Recognition*, Springer Verlag, 2013.
- [4] “*DTW-Radon-based Shape Descriptor for Pattern Recognition*”, S. K.C., B. LAMIROY, Laurent WENDLING in *International Journal of Pattern Recognition and Artificial Intelligence*, World Scientific Publishing, 2013.
- [5] “*Relative Positioning of Stroke Based Clustering: A New Approach to On-line Handwritten Devanagari Character Recognition*”, S. K.C., C. NATTEE, B. LAMIROY *International Journal of Image and Graphics*, World Scientific Publishing, 2012
- [6] *Symbol Recognition using Spatial Relations* S. K.C., B. LAMIROY, L. WENDLING *Pattern Recognition Letters*, Elsevier, 2012, 33 (3), pp. 331-341

#### Selected Article Collections with Blind Review

- [8] “*Computing Precision and Recall with Missing or Uncertain Ground Truth*”, B. LAMIROY, T. SUN extended version of [49] in “*Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16*”, Revised Selected Papers, *Lecture Notes in Computer Science 7423*, Springer, pp. 149-162, Feb. 2013, Ogier, Jean-Marc; Kwon, Young-bin (Eds.)
- [9] “*Spatio-structural Symbol Description with Statistical Feature Add-on*”, S. K.C., B. LAMIROY, L. WENDLING, extended version of [48] in “*Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16*”, Revised Selected Papers, *Lecture Notes in Computer Science 7423*, Springer, pp. 228-237, Feb. 2013, Ogier, Jean-Marc; Kwon, Young-bin (Eds.)
- [10] “*Report on the Symbol Recognition and Spotting Contest*”, E. VALVENY, M. DELALANDRE, R. RAVEAUX, B. LAMIROY in “*Graphics Recognition. New Trends and Challenges. 9th International Workshop, GREC 2011, Seoul, Korea, September 15-16*”, Revised Selected Papers, *Lecture Notes in Computer Science 7423*, Springer, pp. 198-207, Feb. 2013, Ogier, Jean-Marc; Kwon, Young-bin (Eds.)
- [11] “*Dynamic Angle Based Theory in Learning Relative Directional Spatial Relationships on Components of Raster Symbols*”, S. K.C., L. WENDLING, B. LAMIROY extended version of [51] in “*Graphics recognition: achievements, challenges, and evolution, Eighth IAPR International Workshop on Graphics Recognition, Selected Papers*”, *Lecture Notes in Computer Science*, Ogier, Jean-Marc; Liu, Wenyin; Lladós, Josep (Eds.) 2010, pp. 163–174, Springer.
- [12] “*Robust Circular Arc Detection*”, B. LAMIROY, Y. GUEBBAS, extended version of [50] in “*Graphics recognition: achievements, challenges, and evolution, Eighth IAPR International Workshop on Graphics Recognition, Selected Papers*”, *Lecture Notes in Computer Science*, Ogier, Jean-Marc; Liu, Wenyin; Lladós, Josep (Eds.) 2010, Springer.
- [13] “*Scan-to-XML: Using Software Component Algebra for Intelligent Document Generation*”, B. LAMIROY and L. NAJMAN, in *Proceedings of the Fourth IAPR International Workshop on Graphics Recognition*, Springer-Verlag, *Lecture Notes in Computer Science*, 2002.

#### Professional or Broad Audience Publications – Invited Communications

- [16] “*Document Analysis Algorithm Contributions in End-to-End Applications: Report on the ICDAR 2011 Contest*”, B. LAMIROY, D. LOPRESTI, T. SUN in *11th International Conference on Document Analysis and Recognition - ICDAR 2011*, Sep 2011, Beijing, China. IEEE Computer Society

- [17] “*Pattern recognition methods for querying and browsing technical documentation*”, TOMBRE K., LAMIROY B., in 13th Iberoamerican Congress on Pattern Recognition Progress in Pattern Recognition, Image Analysis and Applications Lecture Notes in Computer Science, Springer-Verlag, Lecture Notes in Computer Science, vol. 5197, pages 504-518, 2008.
- [18] “*Graphics Recognition – from Re-engineering to Retrieval*”, K. TOMBRE and B. LAMIROY, in *Seventh International Conference on Document Analysis and Recognition*, Edinburgh, UK, 3-6 August 2003.

#### International Conferences with Blind Review and Edited Proceedings

- [20] “*Relation Bag-of-Features for Symbol Retrieval*”, S. K.C., Laurent WENDLING, B. LAMIROY in Twelfth International Conference on Document Analysis and Recognition (ICDAR 2013), Aug 2013, Washington DC, United States.
- [21] “*An Open Architecture for End-to-End Document Analysis Benchmarking*”, B. LAMIROY, D. LOPRESTI in Eleventh International Conference on Document Analysis and Recognition (ICDAR 2011), oral, September 18 – 21, Beijing, China. (**Special Mention at Best Paper Awards**)
- [22] “*DTW for Matching Radon Features: A Pattern Recognition and Retrieval Method*”, S. K.C., B. LAMIROY, Laurent WENDLING in Jacques Blanc-Talon and Richard P. Kleihorst and Wilfried Philips and Dan C. Popescu and Paul Scheunders. 13th International Conference on Advanced Concepts for Intelligent Vision Systems - ACIVS 2011, 6915, pp. 249-260, Ghent, Belgium. Springer
- [23] “*Document Analysis Research in the Year 2021*”, D. LOPRESTI, B. LAMIROY in Twenty-fourth International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems (IEA/AIE 2011), June 28 – July 1, Syracuse, NY.
- [24] “*How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research*”, B. LAMIROY, D. LOPRESTI, H. KORTH, J. HEFLIN in Document Recognition and Retrieval XVIII, IS&T/SPIE 23rd Annual Symposium on Electronic Imaging, 23-27 January 2011, San Francisco, CA USA.
- [25] “*Spatial Similarity based Stroke Number and Order Free Clustering*”, S. K.C., C. NATTEE, B. LAMIROY in 12th International Conference on Frontiers in Handwriting Recognition (ICFHR), Kolkata, India, November 2010.
- [26] “*Learning Spatial Relations for Graphical Symbol Description*”, K.C., L. WENDLING, B. LAMIROY in International Conference on Pattern Recognition, August 2010, Istanbul, Turkey.
- [27] “*Inductive Logic Programming for Symbol Recognition*”, S. K.C., B. LAMIROY, J-P. ROPERS in “International Conference on Document Analysis and Recognition”, poster, July 2009, Barcelona, Spain.
- [28] “*Robust Circle Detection*”, LAMIROY B., GAUCHER O., FRITZ, L., in 9th International Conference on Document Analysis and Recognition - ICDAR'07, pages 526-530, oral, volume 1, 2007,
- [29] “*An Incremental On-line Parsing Algorithm for Recognizing Sketching Diagrams*”, MAS ROMEU J., SANCHEZ, G., LLADOS, J. LAMIROY, B., in 9th International Conference on Document Analysis and Recognition - ICDAR'07, pages 452-456, volume 1, 2007.

- [30] “Automatic Adjacency Grammar Generator from User Drawn Sketches”, J. MAS ROMEU, B. LAMIROY, G. SANCHEZ and J. LLADOS, in International Conference on Pattern Recognition, poster, pages 1026–1029, Hong Kong, 20–24 August 2006.
- [31] “Scan-to-XML: Automatic Generation of Browsable Technical Documents”, E. VALVENY and B. LAMIROY, in *Proceedings of the Sixteenth International Conference on Pattern Recognition*, poster, Québec City, Canada, 12–15 August 2002.

#### International Workshops with Edited Proceedings

- [40] “The Non-Geek’s Guide to the DAE Platform” Bart LAMIROY, Daniel LOPRESTI in DAS - 10th IAPR International Workshop on Document Analysis Systems, Mar 2012, Gold Coast, Queensland, Australia. IEEE, pp. 27–32.
- [41] “A Real-World Noisy Unstructured Handwritten Notebook Corpus for Document Image Analysis Research”, J. CHEN, Daniel LOPRESTI, Bart LAMIROY in Joint Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data - (JMOCR-AND 2011), Sep 2011, Beijing, China.
- [42] “A Platform for Storing, Visualizing, and Interpreting Collections of Noisy Documents”, B. LAMIROY, D. LOPRESTI in “Fourth Workshop on Analytics for Noisy Unstructured Text Data - AND’10”, Oct. 26, 2010, Toronto, Canada, col. ACM International Conference Proceeding Series.
- [43] “Automatic Learning of Symbol Descriptions Avoiding Topological Ambiguities”, MAS ROMEU J., LAMIROY B., SÁNCHEZ G., LLADÓS J, in 3rd Eurographics Workshop on Sketch-Based Interfaces and Modeling, pages 27–34, September 2006.
- [44] “A Few Steps Towards On-the-Fly Symbol Recognition with Relevance Feedback”, Jan Rendek, Bart Lamiroy and Karl Tombre, in *7th International Workshop, Document Analysis and Systems*, Nelson, New Zealand, Springer-Verlag, Lecture Notes in Computer Science, Volume 3872, January 2006, pp. 604–615.
- [45] “Text/Graphics Separation Revisited”, K. TOMBRE, S. TABBONE, L. PÉLISSIER, B. LAMIROY and Ph. DOSCH in Proceedings of the Fifth IAPR International Workshop on Document Analysis Systems, Princeton, NJ, USA, August 19–21 2002.

#### International Workshops without Review or Edited Proceedings

- [48] “Spatio-structural Symbol Description with Statistical Feature Add-on”, S. K.C., B. LAMIROY, Laurent WENDLING Ninth IAPR International Workshop on Graphics RECOgnition - GREC 2011, Sep 2011, Seoul, Korea, Republic Of.
- [49] “Precision and Recall Without Ground Truth”, B. LAMIROY, T. SUN, Ninth IAPR International Workshop on Graphics RECOgnition - GREC 2011, Sep 2011, Seoul, Korea, Republic Of.
- [50] “Dynamic Angle Based Theory in Learning Relative Directional Spatial Relationships on Components of Raster Symbols”, S. K.C., L. WENDLING, B. LAMIROY in “Eighth IAPR International Workshop on Graphics Recognition”, July 2009, La Rochelle, France.
- [51] “Robust Circular Arc Detection”, B. LAMIROY, Y. GUEBBAS, in “Eighth IAPR International Workshop on Graphics Recognition”, July 2009, La Rochelle, France.
- [52] “Assessing Classification Quality by Image Synthesis”, B. LAMIROY, in “Eighth IAPR International Workshop on Graphics Recognition”, July 2009, La Rochelle, France.

- [53] “*Scan-to-XML for Vector Graphics: an experimental setup for intelligent browsable document generation*”, B. LAMIROY, L. NAJMAN, R. EHRHARD, C. LOUIS, F. QUÉLAIN, N. ROUYER and N. ZEGHACHE in Proceedings of the Fourth IAPR International Workshop on Graphics Recognition, Kingston, Ontario, Canada, September 7-8 2001.

**National Conferences or Colloquia**

- [54] “*Utilisation de Programmation Logique Inductive pour la reconnaissance de symboles*”, S. K.C., B. LAMIROY, J-P. ROPERS in “5ème Atelier ECOI : Extraction de COonnaissance et Images”, GRCE, January 2009, Strasbourg, France

## Chapter 4

# Selected Papers

This chapter reproduces a selection of published papers, in chronological order. They all concern work done relative document image analysis. They are not necessarily the highest cited papers, but the papers I, subjectively, believe best introduce the next chapters.

1. **“Scan-to-xml: Using Software Component Algebra for Intelligent Document Generation”** [13], p. 59

This paper is very interesting because of its premonitory nature. It was one of my very first steps into the domain of document analysis, and, at the time, was a “one-shot” set of general thoughts and ideas, I didn’t really consciously return to the topic until much later [21], with a slightly different point of view (although we had reiterated the ideas in [18]). The research project developed in Chapter 5 is very much influenced by this initial paper, although the concepts of it, at the time of its writing were still fairly unstructured and naïve.

2. **“Robust and Precise Circular Arc Detection”** [12], p. 70

This paper made it to my shortlist because its the final outcome of a process that started 2 years earlier [28] as a quick hack with a student, but which very quickly provided incredibly impressive experimental results. It shows how, as a researcher, we need to remain humble with respect to achievements, that sometimes (often?) are due to chance.

It also has greatly contributed to my understanding of the limits of interpretation and some of the unavoidable shortcomings of contests as evaluation tools and metrics of advances in the state-of-the-art. Furthermore it is currently being negotiated with a private company for being licensed in their production lines for quality purposes.

3. **“Symbol Recognition using Spatial Relations”** [6], p. 82

This is a strong reminder of what a tight-rope walking exercise Ph.D. advisory can be. The results in this paper are mainly due to S. K.C. and his investigations during his Ph.D. Besides the fact that it illustrates how easily initial research goals [27,54] can be thwarted and “abandoned” when reality kicks in and requires investigating unexpected side problems, it also offers another viewpoint to interpretation contexts, and how those influence the perceived quality or pertinence of observed results.

4. **“How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research”** [24], p. 110

This is the first published account of the DAE server<sup>17</sup> and the founding concepts of the performance evaluation paradigm it supports. This theme is of very high importance and will drive my personal short-term research for number of years to come. It is also emblematical for the extremely fruitful collaboration with Lehigh University and Prof. Lopresti it has sparked [16,21,23,24,40,41,42]. It will also provide the experimental testbed on which it shall be possible to implement and validate most of the experiences that will eventually support the theses developed in the following chapters.

Incidentally, it also is a good example to ponder for the chicken-and-egg problem whether the tools shape the research topics, or whether it is the other way round.

5. **“Computing Precision and Recall with Missing or Uncertain Ground Truth”** [8], p. 124

This paper was selected because it is the first step in the direction of actually trying to formalize and measure some of the ideas, conjectures and projects developed in the following chapters.

---

<sup>17</sup><http://dae.cse.lehigh.edu>

## Scan-to-XML: Using Software Component Algebra for Intelligent Document Generation

B. Lamiroy<sup>1</sup> and L. Najman<sup>2,3</sup>

<sup>1</sup> LORIA – INPL / Projet QGAR  
Campus scientifique – B.P. 239  
54506 Vandœuvre-lès-Nancy CEDEX – FRANCE

<sup>2</sup> Océ Print Logic Technologies SA / CCR Department  
1, rue Jean Lemoine – BP 113  
94015 Créteil CEDEX – FRANCE

<sup>3</sup> Laboratoire A2SI, Groupe ESIEE  
Cité Descartes, BP99  
93162 Noisy-le-Grand CEDEX – FRANCE

**Abstract.** The main objective of this paper is to experiment a new approach to develop a high level document analysis platform by composing existing components from a comprehensive library of state-of-the art algorithms. Starting from the observation that document analysis is conducted as a layered pipeline taking syntax as an input, and producing semantics as an output on each layer, we introduce the concept of a *Component Algebra* as an approach to integrate different existing document analysis algorithms in a coherent and self-containing manner. Based on XML for data representation and exchange on the one side, and on combined scripting and compiled libraries on the other side, our claim is that this approach can eventually lead to a universal representation for real world document analysis algorithms.

The test-case of this methodology consists in the realization of a fully automated method for generating a browsable, hyper-linked document from a simple scanned image. Our example is based on cutaway diagrams. Cutaway diagrams present the advantage of containing simple “browsing semantics”, in the sense that they consist of a clearly identifiable legend containing index references, plus a drawing containing one or more occurrences of the same indices.

### 1 Introduction and Objectives

In this paper, we aim to validate a new approach to develop high level document analysis tools by composing existing components from a comprehensive library of state-of-the art algorithms, by developing a fully automated method for generating a browsable, hyper-linked document from a simple scanned image. The work presented here was conducted in collaboration with the QGAR research group of the LORIA laboratory, Océ Print Logic Technologies, and students from the École des Mines de Nancy [8].

The expressed need for composing existing algorithms naturally emerges from the observation that there exist a large number of papers describing ideas aimed at solving particular points of specific problems. Reusing this work for solving other applications seems to be quite a difficult task. In order to attain a functional level that can be used to tackle real-world problems, there is a pressing need to evaluate, experiment and combine existing approaches into an operational whole.

Moreover, the reuse issue is crucial, since the generic reasoning behind document analysis is to adopt a multi-level **syntax+context=semantics** paradigm. In other terms, most of the time, there is a loop where the semantics of a lower level algorithm, become the syntax of a higher level approach, and so on and so forth. The main problem is that the notion of the three terms: *syntax*, *context* and *semantics* tend to vary rapidly in function of a great number of parameters, one of the most important being the improvements of algorithms and advances in the state-of-the-art.

**The syntax** of a document tends to come from individual algorithms (vectorization, segmentation, skeletization, ...). These algorithms alone cannot decide on the "*truth*" (veracity, accuracy) of its output. It is only the whole treatment chain that can give information on that. Therefore, the question arises of how to use, to the best, imperfect results.

**The context** is given by the type of document and the problem that needs to be solved. It is generally expressed in terms of combination and conditions of syntactic expressions and expert knowledge. It is the correct formulation of this contextual information that gives the added value to a new approach that uses basic components to construct a higher level solution. Finding the correct contextual formulation is generally a process of trial and error, and requires a flexible framework for testing and reuse.

We insist on the fact that the paradigm we are describing is a general conception of how document analysis is conducted: a layered pipeline taking syntax as an input, and producing semantics as an output on each level. We are not referring to syntactic methods within the document analysis domain ([1,3,7,11] among others), but we are thinking of a more general framework.

Another important aspect that needs to be considered in the light of reuse and combination of existing components is the need to describe the results of one component in a way that can be easily translated (used) for further (more advanced) purposes. In other terms, *finding a flexible way to express the results*. There is some kind of intelligence already present in the way we express the result (there are two kinds of problems: the solved ones, and the ill-expressed ones). In today's world, the expression of choice is the XML format and its various derivatives (SVG, or in a way HTML). Most of the work in the XML context consists in using XSSL, XLT to transform the data, in a form easier to use for other purposes. We address this point in this paper.

As a summary we would like to emphasize the facts that

1. More than 20 years of advances in document analysis has given rise to an enormous pool of high performance algorithms. However, they tend to address individual and specific problems, and (independently of any benchmarking or whatsoever) are more or less suited than competing approaches for particular tasks.
2. In order to solve higher level, real-world problems, there is a need for building upon those existing components in an open manner that integrates the demand for flexibility and interchangeability.

In that sense, we need:

- (a) a way to communicate results of one given treatment to another without being restricted to particular formats of their representation.
- (b) an experimentation and development framework that allows a flexible combination of different components, and facilitates their interchange. The principle objective is to form meta-components expressing an enhanced context and opening the way to a straightforward assembly of intelligent components, and finally achieve a full treatment pipeline architecture.

The outline of the paper is as follows. First, we more thoroughly analyze the needs in terms of operational requirements for a platform that would suit the needs of flexible inter-component communication, and allow for efficient experimental research. Secondly, we present a test-case for this environment, that consists in converting a cutaway diagram into a browsable document, and present the result of an automated treatment pipeline solving this problem.

## 2 The model

In this section, we describe our approach to achieve the objectives expressed in section 1. At first, we refine the requirements induced by the introduction, then we propose our solution.

### 2.1 Requirements

The model for testing ideas in graphic analysis is the pipeline: an image is run through a series of operations, resulting in other images and other kinds of information extracted from those images and in various formats.

Our model should comply to the following requirements:

1. We want this pipeline model evolutionary, in the sense that it should possible to be modified in a straightforward manner. This is necessary for several reasons. First, there is a need for flexible experimentation of new algorithms, combining existing components that are still in a non-finalized stage of development. Second, advances in the state-of-the-art may require fine-tuning of details or modifications and additions of new concepts.

2. We want to be able to add our home-made components as easily as possible. If some (part) of those components play the same role, they need to be interchangeable for comparison purposes. Thus, our platform has to be modular.
3. Ideally, our platform should give us access to the greatest number of existing components, such components offering high-level services, like the ability to build a user-interface for demonstration purposes. Furthermore, the use of a console mode, i.e. a mode in which we can test step-by-step some idea, deciding only when seeing the result of the current step what will be the next step, is a must for debugging or fine-tuning of parameters.

## 2.2 Formal Analysis

If we formally represent the whole pipeline process as the application of a certain number of well chosen algorithms on a set of initial data, the usual way of seeing things is to represent algorithms as operators, and data as operands. There is, however, a more subtle way to represent the same situation, it has the particularity to be less stringent on the separation between what is algorithm and what is data.

Let us assume data is *truth*. Algorithms, therefore, must be seen as reasonings producing new truths from the initial data. In other terms, the initial data form *axioms*, subsequent data produced by the application of algorithms form corollaries or other theorems. In that sense, they are proofs. Re-expressing the previously cited requirements leads us to separate two classes of proofs: the *well established* ones, which are the experimented class of algorithms, found in the state of the art ; and the more experimental ones, built upon the previous ones by empirical trial and error until proven worthy enough to pass over to the first class.

This allows us to shed a new light on the **syntax+context=semantics** paradigm referred to in the introduction. The same equation can be rewritten into **data + algorithm + meta-knowledge = data**, where the meta-knowledge groups the following (non-exhaustive) list of items: choice of the algorithm (or combination of algorithms), parameterization of the algorithm, thresholding and selection of pertinent data, ... The individual algorithms are usually available, tend to come in multiple flavors and sorts, and are particularly specialized for a given context. The meta-knowledge itself is difficult to come by in an automated way<sup>1</sup>, and is more generally the result of a fuzzy empirical process (which is a more snobbish way for saying “guesswork”). This meta-knowledge has the further tweak of being data-like without actually being part of the data, nor really part of the algorithm.

Solving complex problems thus consists of starting with the lowest possible syntax or semantics: a raw image, and progressively applying well chosen algorithms and adding context information (or meta-knowledge) as the process

---

<sup>1</sup> Which is rather obvious, once one thinks of it ... if it were, it would be part of the algorithm itself.

advances and intermediate data is produced, eventually converging to the final semantics.

This is not without raising a number of practical concerns:

- Any intermediate algorithm solving a particular problem within the pipeline is probably the result of a choice between many existing alternatives. In the experimental context of finding the solution, it is absolutely necessary that this algorithm can be replaced by any other equivalent.
- Slightly changing the goal semantics should result in only a minimal change in the intermediate set of combined algorithms

### 2.3 A Component Algebra for Document Analysis

All the points discussed in the two previous sections naturally lead to what we'd like to call a Component Algebra:

- Valid data belongs to a well defined set.
- Components transform input data to output data belonging to the same, well defined set.
- Any intermediate result of a component is valid data.
- Data is self-explaining. Any component should take any data as input, and be able to determine if it is part of its application domain.
- Components are interchangeable.
- Context is easily represented and passed from one component to another as an attribute of the data itself.

In the following sections we shall develop the more practical choices that lead to a fully operational framework. XML is the most appropriate data representation format for the needs we have expressed. Determining the most efficient DTD to cover our ambitions is currently ongoing work. The glue for component construction and interaction is given by scripting. Combined with a compiled library it gives all latitude to realize our goal, as we are describing in the next section.

### 2.4 Implementation Proposal: Scripting Languages

It is now well known that object-oriented technology alone is not enough to ensure that the system we want to develop is flexible and adaptable. Scripting languages [9] go a step further than object oriented framework, as they integrate concepts for component-based application development. As it is stated in [9], *Scripting languages assume that there already exists a collection of useful components written in other languages. Scripting languages aren't intended for writing applications from scratch; they are intended primarily for plugging together components.* They thus tend to make the architecture of applications much more explicit than within an object-oriented framework [10].

- Scripting languages are extendible [2]: this means that new abstractions can easily be added to the language, encouraging the integration of legacy code.

- They are embeddable, which means that they offer a flexible way to adapt and extend applications. As it is stated in [10], *the composition of components using a script leads to a reusable component. Therefore, components and scripts can be considered as a kind of component algebra.*
- Scripting languages support a specific architectural style of composing components, namely the *pipe and filter* style supported by UNIX shells. As we have stated before, this is a good prototyping architecture, well adapted to image processing.

To obtain a fully operational experimentation platform, we build up a new component from our own graphic analysis library QGAR [4,5]. QGAR is a library written in C++, which implements basic methods for graphic recognition<sup>2</sup>: among the methods implemented, we found some low-level algorithms (binarization, segmentation, gradient methods, convolution masks, ...) and some vectorization methods. The use of SWIG [2] allows to automate the creation of the component: SWIG is a compiler that takes C/C++ declarations, turns them into the "glue" to access them from common scripting languages, including Python and Tcl.

### 3 An Example of an Application: a Browsable Cutaway

The most difficult part of succeeding this kind of generic architectures is the realization of a component algebra itself. The basic idea being the flexible combination of existing components to form new components, there is a crucial need of representing data for interchange between components. We have good conscience that the complete conception of such an algebra, and the representation of the data that is exchanged between components goes far beyond the scope of this paper. A thorough study of the formal requirements of such a construct will probably end at the fundamental difficulties of graphics analysis: what are the image semantics and how to represent them? In our paper, we restrict ourselves to *a priori* known and very simple semantics. It is our ambition to continue this study in depth and beyond.

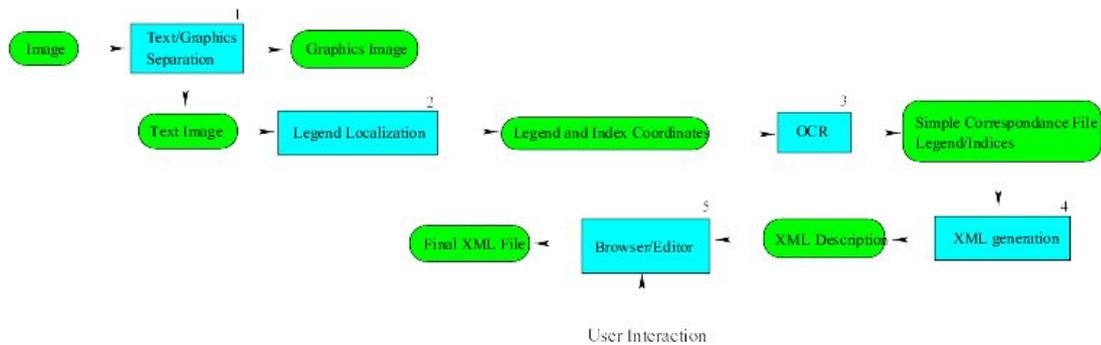
The application we have chosen for illustrating our ideas is to render a cutaway diagram browsable [8]. Cutaway diagrams present nicely identifiable image semantics, but extracting them requires the collaboration between a number of image treatments. They are composed of a graphical image, containing a number of *tags* (which we shall refer to as *indexes*) denoting zones of interest in their vicinity. A copy of all tags, together with an explaining text, is also present along the graphical image, and is called "the legend".

In what follows we take the following assumptions:

1. Tags are formed by alphanumeric characters we shall refer to as *strings*. There is no supplementary assumption concerning the tags, more particularly, we do not assume there are particular signs (such as surrounding

---

<sup>2</sup> The complete source code can be downloaded or tested on-line at <http://www.qgar.org>.



**Fig. 1.** General Software Architecture: A modular pipeline with easily interchangeable components.

circles, arrows, *etc.* – *cf.* Figure 2) that would allow easy detection of the tags.

2. The legend has an array structure, consisting of an undefined number of macro-columns, each of which contains two micro-columns: the leftmost containing the tag, the rightmost containing the explaining text.

Our goal is to detect the tags in the image, to detect the legend, and match the tags with the corresponding explaining text in the image by producing a mapping in the form of an XML document. A home-brew browser, will allow navigation through the document: clicking on a tag in the image or in the legend highlights the corresponding tag(s).

### 3.1 General Algorithm Outline

In order to achieve this goal, we use the pipeline architected software model that takes a raw image as input, and sequentially executes the following treatment (Figure 1):

1. Separate text from graphics, and produce two new, raw images, one only containing text, the other only containing graphics.

The text–graphics separation is an enhanced implementation of the FLETCHER–KASTURI algorithm [6]. It consists in an analysis of the connected components of the initial raw binary image. The algorithm is based on a statistical analysis of the distribution of the bounding boxes of all connected components. Given the fact that alphanumeric characters roughly have small squarish bounding boxes, a simple thresholding with respect to the bounding boxes’ surface and ratio allows it to quite nicely separate text from graphics. It is noteworthy to mention that the quality of the Text/Graphic segmentation does not need to be absolutely perfect for our algorithm to work

- properly. As will be shown further on, we can cope with a certain level of badly classified glyphs.
2. Taking the text image as input, analyze it, and locate the position of the legend containing the needed references. Separate the text image into another two raw images: one containing the legend area, another containing the rest. Since we know that the legend is generally a concentrated “blob” of texts (independent of its content), and that the referenced indexes in the drawing are most often sparsely scattered over the image, we use a simple algorithm that is general and robust enough to extract the location of the legend (although it might need some tuning in the case of awkwardly formed, *i.e.* non-rectangular, legends): we apply a Run-Length Smoothing Algorithm closure [12] on the image, with its parameter roughly the size of a text element. This will result in the legend form a homogeneous zone of filled black pixels. A simple search of very high density pixel clusters rapidly locates the legend with respect to the rest of the image or other textual elements.
  3. Taking the legend area as input, analyze its structure and output a list of image zones, containing the references to be searched for in the rest of the image. These zones are fed to an OCR, and the final result of the treatment consists of a list of image zones and the corresponding references, as recognized by the OCR.
  4. Taking the rest of the image as input, we also produce a list of the remaining text zones as well as the recognition results using the same OCR. This approach has the advantage to be rather robust. On the one hand, clutter in the Text layer will give at best a “not recognized” result from the OCR, or a random text label, at worst. The probability to encounter an identical text label (provided it contains a sufficient number of individual characters) is very low. Clutter therefore has very few chances to be matched with anything. A similar reasoning holds for the micro-analysis of the legend we mentioned before. Since the legend normally is composed of a tag, followed by a longer text string, there are few chances that the text string will be mismatched by a tag.
  5. Both lists are analyzed in order to produce a coherent browsable structure, expressed in XML.
  6. The XML file is then given as input for a customized browser-editor, written in Tcl/Tk, that allows navigation and editing of the final document.

### 3.2 XML for Data Exchange

In order to achieve our ambition of having complete interoperability, XML is missing link for collaborating modules. Indeed, in the context of a component algebra, any image treatment module can be either a final implementation, a component within a more complex treatment pipeline, or both. A sound, and universal knowledge representation scheme is absolutely necessary in this context, and XML [13] and its derivatives seem to be the perfect key to this framework. XML was designed to be an expression language in which it is possible to represent the underlying semantics of a document, and that that offers the possibility to easily extend and enrich its description.

**XML Applied to Browsing Semantics** Within the experimental context of this paper, we were principally concerned with expressing the content of a browsable document. We therefore decided on a very simple and straightforward DTD that would allow the interpretation of a cutaway tags, as detected by our algorithm. A browsable document (`NAVIGABLE_IMAGE`) contains three sub-parts:

- The path to a source image (`IMG`), mainly for displaying reasons.
- A list of rectangles (`LEGEND`), that contains a unique identifier as an attribute, and is supposed to be composed of two corners (`UPPER_LEFT` and `BOTTOM_RIGHT`) and an OCR recognition tag. This list represents the recognized items of the legend zone of the document.
- A similar list of rectangles (`DRAWING`, that also contains a unique identifier as an attribute, two corners (`UPPER_LEFT` and `BOTTOM_RIGHT`) and an OCR recognition tag. This lists enumerates the tags scattered over the drawing part of the document.

Our browser (*cf.* Figure 2) simply parses both lists and activates links between items of the `DRAWING`-list with those of the `LEGEND`-list, when they present the same OCR recognition tag. Unlinked zones can be highlighted as warnings for possible further fine-tuning and error correction methods. Moreover, the formal description of the document and the availability of a browser/editor, allows for a straightforward inclusion of human intervention: the browser/editor can allow for modification of the tag dimensions (merge/split/rezise) and tag labels (and indirectly the links) or even allow for creation and deletion of tags. Having a formal description of the document as well as a browser/editor nicely turns our experiment into a complete, exploitable platform for semi-automated hyper-linked document generation.

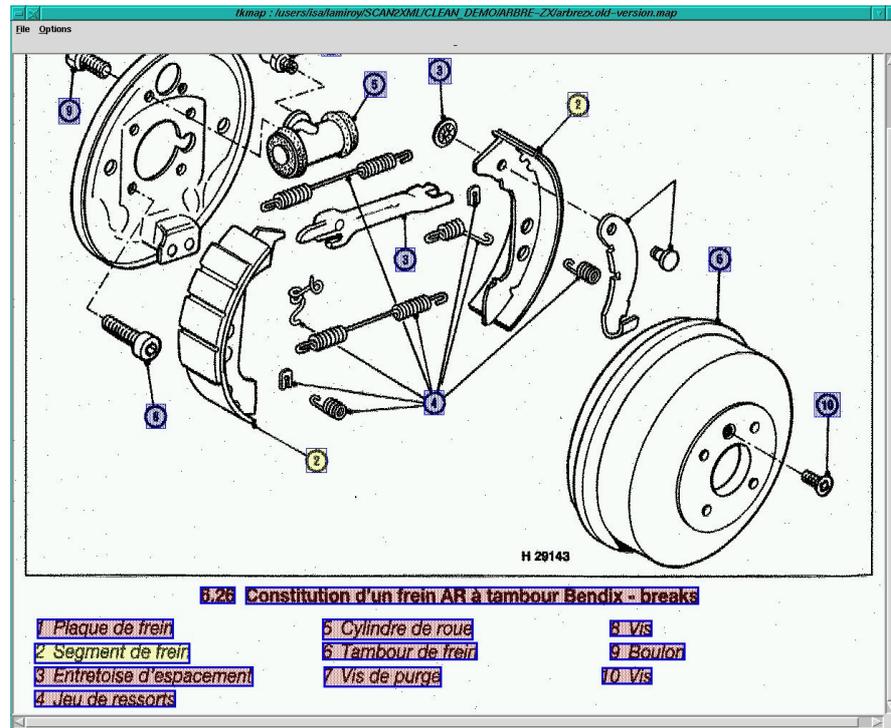
## 4 Conclusion and Perspectives

In this paper we have presented a first approach toward the creation of intelligent browsable documents, applied to cutaway diagrams. We identify the tags in the drawing and correlate them with the occurrences of the same tags in the legend adjoining the graphics.

Moreover, we identified the need for a real component algebra for document analysis applications and research and defined the major requirements for such an environment:

- flexibility and interchangeability, which we obtain through scripting basic image treatment components
- and interoperability, which we want to obtain by defining a sound XML-based description format for image and document analysis results.

The very first implementation we made of this component algebra, based on the Tcl/Tk scripting language and our own graphic analysis library `QGAR`,



**Fig. 2.** Screenshot of Browser/Editor, showing a fully browsable cutaway diagram. Red and blue zones are “clickable”. Highlighted zones are the results of a selection.

has demonstrated the power of such a design for rapidly obtaining a complete, exploitable platform for semi-automated hyper-linked document generation<sup>3</sup>.

One of the major issues remains the data representation. In our test-case, we controlled the compiled software components, and their APIs were coherent with each other. Furthermore, we only used a limited part of our library, reducing the data types to be exchanged (finally, everything is reduced to their respective bounding boxes). Before testing the scalability of our approach (especially by integrating other sources for document analysis components), we need to formalize the data framework.

We are currently investigating the most appropriate way for expressing the data flow between different components. In our example, we developed an *ad*

<sup>3</sup> All programming was done by 5 students, totaling roughly 300 man-hours, considering that they had no knowledge in image analysis, had to implement the pipeline, develop the XML browser/editor, compare different public domain OCR engines, etc.

*hoc* DTD that is anything but generic and extendible. It is clear that, in order to achieve a realistic framework, the final representation should encompass most graphical entities and attributes. SVG is a very important initiative in that direction, and we are currently investigating its possibilities. However, at the printing of this document, it is still unclear whether this W3C recommendation remains suited for free, unlimited use, and other challengers, like MPEG7–VRML also need to be considered.

## References

1. R. H. Anderson. Syntax directed recognition of hand-printed two-dimensional mathematics. In M. Klerer and J. Reinfelds, editors, *Interactive Systems for Experimental Applied Mathematics*. Academic Press, New York, 1968.
2. D. M. Beazley. SWIG and automated C/C++ scripting extensions. *Dr. Dobbs Journal*, (282):30–36, February 1998.
3. D. Dori. A Syntactic/Geometric Approach to Recognition of Dimensions in Engineering Drawings. *Computer Vision, Graphics and Image Processing*, 47:271–291, 1989.
4. Ph. Dosch, C. Ah-Soon, G. Masini, G. Sánchez, and K. Tombre. Design of an Integrated Environment for the Automated Analysis of Architectural Drawings. In S.-W. Lee and Y. Nakano, editors, *Document Analysis Systems: Theory and Practice. Selected papers from Third IAPR Workshop, DAS'98, Nagano, Japan, November 4–6, 1998, in revised version*, Lecture Notes in Computer Science 1655, pages 295–309. Springer-Verlag, Berlin, 1999.
5. Ph. Dosch, K. Tombre, C. Ah-Soon, and G. Masini. A complete system for analysis of architectural drawings. *International Journal on Document Analysis and Recognition*, 3(2):102–116, December 2000.
6. L. A. Fletcher and R. Kasturi. A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images. *IEEE Transactions on PAMI*, 10(6):910–918, 1988.
7. S. H. Joseph and T. P. Pridmore. Knowledge-Directed Interpretation of Mechanical Engineering Drawings. *IEEE Transactions on PAMI*, 14(9):928–940, September 1992.
8. B. Lamiroy, L. Najman, R. Ehrhard, C. Louis, F. Quélain, N. Rouyer, and N. Zeghache. Scan-to-XML for vector graphics: an experimental setup for intelligent browsable document generation. In *Proceedings of Fourth IAPR International Workshop on Graphics Recognition*, Kingston, Ontario, Canada, September 2001.
9. John K. Ousterhout. Scripting: Higher-level programming for the 21st century. *Computer*, 31(3):23–30, March 1998.
10. J.-G. Schneider and O. Nierstrasz. Components, scripts and glue. In J. Hall L. Barroca and P. Hall, editors, *Software Architectures - Advances and Applications*, pages 13–25. Springer, 1999.
11. M. Viswanathan. Analysis of Scanned Documents — a Syntactic Approach. In H. S. Baird, H. Bunke, and K. Yamamoto, editors, *Structured Document Image Analysis*, pages 115–136. Springer-Verlag, Heidelberg, 1992.
12. K.Y. Wong, R.G. Casey, and F.M. Wahl. Document analysis system. *IBM J. Res. Develop.*, 26(2):647–656, 1982.
13. Extensible markup language (XML) 1.0 (second edition). Technical report, w3c, 2000. <http://www.w3.org/TR/2000/REC-xml-20001006>.

# Robust and Precise Circular Arc Detection

Bart Lamiroy and Yassine Guebbas

Nancy Université – INPL – LORIA  
Équipe Qgar – Bât. B  
Campus Scientifique – BP 239  
54506 Vandoeuvre-lès-Nancy Cedex – France  
Bart.Lamiroy@loria.fr

**Abstract.** In this paper we present a method to robustly detect circular arcs in a line drawing image. The method is fast, robust and very reliable, and is capable of assessing the quality of its detection. It is based on Random Sample Consensus minimization, and uses techniques that are inspired from object tracking in image sequences. It is based on simple initial guesses, either based on connected line segments, or on elementary mainstream arc detection algorithms. Our method consists of gradually deforming these circular arc candidates as to precisely fit onto the image strokes, or to reject them if the fitting is not possible, this virtually eliminates spurious detections on the one hand, and avoiding non-detections on the other hand.

## 1 Introduction

Finding circular arcs is one the recurring problems in graphical document interpretation or symbol recognition. The main difficulty with the existing approaches is that they often are of considerable complexity (e.g. Hough-like [1] or feature grouping approaches [2]) sensitive to image quality, line thickness, or rely on a number of user defined parameters or thresholds that make them extremely difficult to apply to generic problems or on heterogeneous document sets.

The approach developed in this paper reduces the set of needed parameters to a minimal set of very elementary and visually significant values and can be applied without prior knowledge of the document set, regardless of line widths, connectedness or complexity. It relies on elementary (3,4)-distance transform skeletonization [3] and segment detection [4]. Unlike extremely efficient methods like [5], ours does not require reasonable segmentation of arcs. This work is tightly related to [6].

The following section establishes how to determine if a single circular arc is present, provided we have a rough initial guess of its position, and how to robustly detect and locate it using RANSAC (Random Sample Consensus [7]). Section 3 then explains how to generalize to detecting and localizing any number of circles, without *a priori* knowledge of their position. The last two sections conclude by eliminating spurious detections and by establishing the limits of the approach.

## 2 Determining the Presence of a Circular Arc

In this section we address the problem of detecting a circular arc, given an initial estimate of its center  $(x_c, y_c)$ , its radius  $\sigma$ , and its two endpoints  $p_l$  and  $p_r$ <sup>1</sup>. This estimate, as we shall see further, can be very approximate. The main goal, in this first stage, is to detect whether or not, an arc is present in the image, near the vicinity of the given parameters.

### 2.1 General Algorithm

We are mainly exploiting the algorithm described in [6], with one major adjunction. The cited method has been developed to identify and locate full circles, and therefore only needs to consider adapting to two variables: the center  $(x_c, y_c)$ , and the radius  $\sigma$ . This is not the case anymore for detecting arcs, since two parameters are added:  $p_l$  and  $p_r$ , the left and right endpoints.

The general approach we develop consists of taking the set  $\mathcal{P} = \{p_i\}$  of all pixels  $p_i$  lying on the discrete circular arc  $\mathcal{A}^0$  defined by  $(x_c, y_c)$ ,  $\sigma$ ,  $p_l$  and  $p_r$ . As in, [6], we define, for each of these pixels  $p_i$ , the discrete line  $\Delta_i$ , starting at  $(x_c, y_c)$ , and passing through  $p_i$ . Let  $q_i$  be the pixel on  $\Delta_i$  that is the closest black pixel to  $p_i$ . Let  $\mathcal{Q}_a^0 = \{q_i\}$ .  $\mathcal{Q}_a^0$  therefore is the set of all black pixels closest to the initial estimate  $\mathcal{A}^0$  in the direction of the circle radius.

Figure 1 gives an illustration of this estimation. Initial guesses are drawn in blue. For each conjectured circle, green pixels are those found at the correct distance from the center, while red ones lie on the radius and are closest to the circle.

Now, let  $\mathcal{C}^1$  be the best fitting circle over  $\mathcal{Q}_a^0$  (any criterion can be used, but we are using the Least Median of Squares – cf. section 2.2), and let us generalize the previous step, such that  $\mathcal{Q}_c^t$  contains the set of all black pixels closest to the theoretical circle  $\mathcal{C}^t$  in the direction of the circle radius (and similarly for  $\mathcal{Q}_a^t$ ).

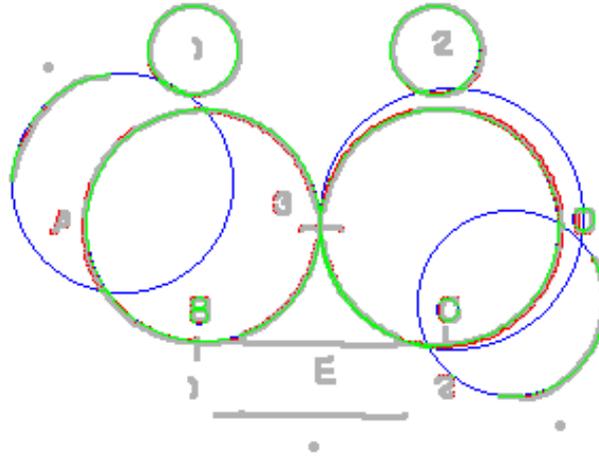
By construction,  $\mathcal{Q}_a^t \subset \mathcal{Q}_c^t$ , and while this new set of points allows for a re-estimation of  $(x_c, y_c)$ ,  $\sigma$ , the other parameters  $p_l$  and  $p_r$  need to be re-evaluated as well. The approach is the following:

Let  $\tau^t$  be the error measure between  $\mathcal{A}^t$  and  $\mathcal{Q}_a^t$ . *i.e.*  $\tau^t$  represents the fitness between the model  $\mathcal{A}^t$  and its corresponding data  $\mathcal{Q}_a^t$ . Let  $\mathcal{A}_<^t \subset \mathcal{A}^t$  a smaller circular arc<sup>2</sup> than  $\mathcal{A}^t$  such that  $\tau_{<}^t > \tau^t$  and that

$$\forall \mathcal{A}^{t*} | \mathcal{A}_<^t \subset \mathcal{A}^{t*} \subset \mathcal{A}^t : \tau^{t*} < \tau^t. \quad (1)$$

<sup>1</sup> In this document we shall conveniently ignore the fact that there is a small ambiguity with defining an arc by the center of its corresponding circle, the radius and the endpoints: one also has at least the orientation of the arc to consider as to know what part of the circle between the two endpoints is belonging to the arc, and which part isn't.

<sup>2</sup> “smaller” meaning having the same center and radius, but having a smaller aperture while being fully included in the “larger” one.



**Fig. 1.** Example of circle hypotheses: in blue, the initial guess; in green, points correctly lying on the conjectured circle; in red, point closest to the circle.

In other terms,  $\mathcal{A}_{<}^t$  is a smaller arc that fits the dataset better than  $\mathcal{A}^t$  and all intermediate arcs fit less. This means that  $\mathcal{A}_{<}^t$  is the largest sub-arc fitting the data better than  $\mathcal{A}^t$ .

We do a similar search by increasing the arc size thus obtaining  $\mathcal{A}^t \subset \mathcal{A}_{>}^t$  a larger circular arc than  $\mathcal{A}^t$  such that  $\tau_{>}^t > \tau^t$  and that

$$\forall \mathcal{A}^{t*} | \mathcal{A}^t \subset \mathcal{A}^{t*} \subset \mathcal{A}_{<}^t : \tau^{t*} < \tau^t, \quad (2)$$

$\mathcal{A}_{>}^t$  thus being the smallest super-arc fitting the data better than  $\mathcal{A}^t$ .

We can then define  $\mathcal{A}^{t+1}$  as being the  $\operatorname{argmax}_{\tau} \{\mathcal{A}_{<}^t, \mathcal{A}^t, \mathcal{A}_{>}^t\}$ . Continuing this iteration until  $\mathcal{A}^t = \mathcal{A}^{t+1}$  will yield the best estimate of the arc (if any) closest to the initial  $\mathcal{A}^0$ .

In the following sections we detail the different steps of this general approach.

## 2.2 Using RANSAC and LMedS

Since there is no guarantee that any  $\mathcal{A}^t$  or  $\mathcal{Q}^t$  may effectively contain points that form a circle, it may be extremely hazardous to use global minimization approaches (like Least Squares, for instance) [8]. It is known that these estimators are very sensitive to outliers or spurious data that does not conform to the required model [9]. Using these functions would invariably lead to degenerate convergence.

RANSAC [7] is much better suited for fitting very noisy data – especially data containing measures that do not belong to the model that is to be estimated –

The approach consists of selecting the strict minimum of data points required for estimating an instance of the model (*e.g.* three points for estimating a circle) and then computing the residual error of the other data points to this model. This is done a number of times, and the final model is the one with the lowest residual error.

More formally: let  $\mathcal{Q}^t$  be the set of model points.  $\mathcal{Q}^t$  supposedly, and in the worst case, contains a ratio of  $\tau$  outliers. Let  $q_n, q'_n$  and  $q''_n$  be three random points belonging to  $\mathcal{Q}^t$ , and let  $\mathcal{C}_n$  be the circle defined by and passing through  $q_n, q'_n$  and  $q''_n$ . Let  $\delta(\mathcal{C}, p)$  be the distance of a point  $p$  to a circle  $\mathcal{C}$ , and let  $\text{Med}_\tau(\mathcal{S})$  be the  $\tau$ -quantile median value of the set  $\mathcal{S}$ . We then define the residual error of a set of model points  $\mathcal{Q}^t$  to a circle  $\mathcal{C}_n$  as

$$\text{RsdErr}(\mathcal{Q}^t, \mathcal{C}_n) = \text{Med}_\tau(\{\delta(\mathcal{C}_n, p) \mid p \in \mathcal{Q}^t\}). \quad (3)$$

$\text{RsdErr}$  gives the maximum distance of a set of points to a circle, discarding a proportion of  $\tau$  outliers.

With RANSAC we choose  $R$  random subsets of 3 points within  $\mathcal{Q}^t$ , each giving rise to the computation of a circle  $\mathcal{C}_n$ . For each subset, we compute the corresponding  $\text{RsdErr}(\mathcal{Q}^t, \mathcal{C}_n)$ , thus obtaining

$$\mathcal{C}^{t+1} = \underset{\mathcal{C}_n, n \in [1 \dots R]}{\text{argmin}} (\text{RsdErr}(\mathcal{Q}^t, \mathcal{C}_n)). \quad (4)$$

The number of required subsets can be formally deduced from both the quality of the data (expected rate of outliers  $\tau$ ), the dimensionality of the problem (here 6, since we need three points for estimating a circle, each point having two dimensions) and the required confidence in the result [7].

### 3 Robust Arc Detection

The previously presented method does a very good job of robustly determining whether there is a circular arc close to a given center and radius  $(x_c, y_c)$  and  $\sigma$ . However, it needs some initial guess on where to search. The method we are developing here proceeds in three main phases:

1. Generate a high number of possible arc candidates, without consideration of uniqueness, overlapping or exact localization.
2. Verify the quality of each candidate using the approach described in section 2. The output of this verification is a list  $\mathcal{A}$  of genuine arcs, correctly fitted on the image data.
3. Detect and merge multiple and/or partial detections of the same curves as to obtain a set of unique, disjoint arcs.

#### 3.1 Arc Candidate Generation

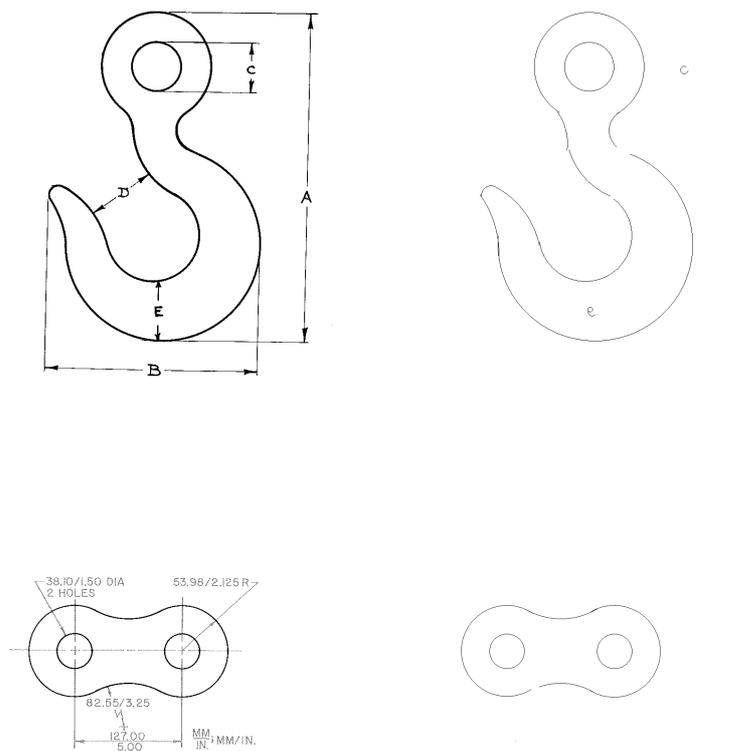
In order to obtain the largest possible set of arc candidates, we automatically segment the image using a basic Rosin & West line segment vectorization [4]. We

then simply enumerate all connected pairs of segments. Each pair gives us three points, which is exactly the amount of data that allows for getting an initial guess for a circular arc:  $(p_1, p_2, p_3)$ . These points define a unique circle on the one hand, and furthermore, since they are ordered –  $p_2$  being in the middle – they define the left and right extrema for the definition of an arc.

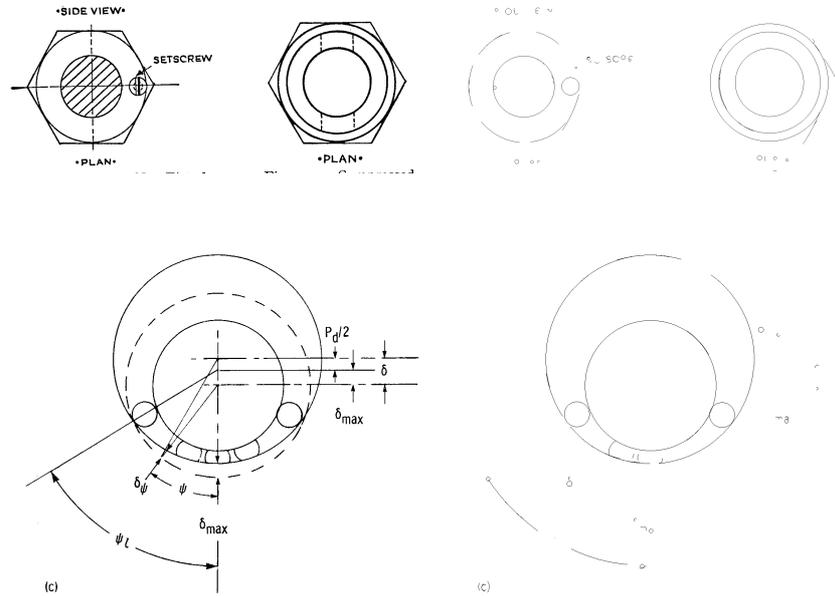
This approach is combined with direct arc detection from [4] as to produce the largest possible set of arc candidates to bootstrap our localization method (*cf.* section 2).

### 3.2 Merging of multiple detections

Since the method is based on unfiltered hypothesis generation, it has a clear tendency toward over-segmentation, as shown in Figure 2. The main idea behind



**Fig. 2.** GREC 2007 contest images: original image (left) – final segmentation (right)

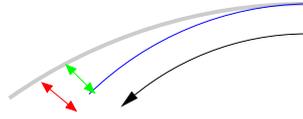


**Fig. 3.** GREC 2009 contest images: original image (left) – final segmentation (right)

being tolerant towards this over-segmentation is to be confident that (almost) all image pixels belonging to an arc are covered by at least one initial arc candidate. Merging arcs should therefore result in a full coverage of each arc of the image by one unique, genuine arc. Merging arc candidates representing the same circular arc in the image requires two distinct operations: merging estimates covering the same pixels and merging arc candidates not sharing the same pixels but being partial estimates of a same wider arc. These two operations can be performed by first increasing the aperture of the arcs (*cf.* section 3.2.1), thus making hypothetical arcs share pixels, and, secondly, merging the arc candidates sharing pixels (*cf.* section 3.2.3). For merging arcs, we do not use the full circle image, but the image skeleton [6].

**3.2.1 Increasing Aperture** To increase the aperture of an arc, we first set a threshold to the maximum distance between a point from the discrete hypothetical arc and the closest pixel, as shown in Figure 4, where the distance is measured on the line going through the center of the hypothetical arc and a point on the hypothetical arc.

The increase of the aperture of an arc is done by starting from the endpoints of the candidate arc  $p_l$  and  $p_r$  and then increasing the aperture pixel-wise, as long as the distance to the closest pixel remains below the threshold.



**Fig. 4.** Distance between a point of the hypothetical arc and the closest pixel: in blue the hypothetical arc; in Grey, image pixels; in green, distance between a point on the hypothetical arc and the closest pixel; in red, threshold

**3.2.2 Finding Which Arcs to Merge** Once the set of maximal arc candidates obtained, overlapping ones or those lying on the same image curve need to respond to the following criteria in order to be merged:

1. A non-empty intersection between two arcs means that these two arcs are likely part of a same covering arc. However two arcs having common closest pixels are not necessarily sub-arcs of a same arc as shown in Figure 5.
2. Arcs having comparable radii are merge candidates. This is checked through a radius ratio with the formula:

$$\frac{|r_1 - r_2|}{\max(r_1, r_2)} < \text{RatioRadiusError}. \quad (5)$$

Checking the center of the circle is less robust since small changes in curvature may be visually insignificant, but generate large differences in the center position.

3. Arcs having opposed normal vectors (*cf.* Figure 5) are not eligible for merging, even though they may overlap. This criterion is verified by choosing a point  $I$  from the overlapping part of two arcs, and constructing a vector  $\overrightarrow{IO_1}$  that originates from  $I$  and finishes at  $O_1$  the center of the first arc, and defining similarly a vector  $\overrightarrow{IO_2}$ . We then compute the scalar product of these vectors:

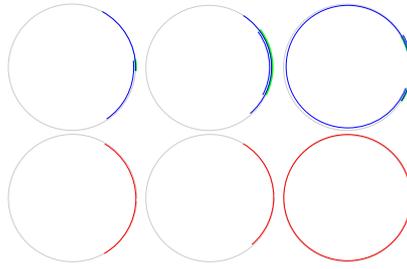
$$\overrightarrow{IO_1} \cdot \overrightarrow{IO_2} = |\overrightarrow{IO_1}| |\overrightarrow{IO_2}| \cos \theta \quad (6)$$

If the sign of the product is positive, we consider that the two arcs stem from the same covering one.



**Fig. 5.** Intersection of arcs with opposed curvature signs or with significantly different radii

Once these criteria are verified, three different configurations may occur for merging. They are depicted in Figure 6. In configuration A the two arcs are “adjacent” sharing some pixels, in configuration B one arc includes another and in configuration C the two arcs are “explementary”. The covering arc is formed by considering that the two arcs belong to the same circle. Therefore the resulting arc is the union of the two arcs as if they had the same center and the same radius, in other words, the computation of the arc’s angle and aperture is based on the angles and apertures of the two arcs.



**Fig. 6.** Intersection configurations (top) and corresponding covering arcs (bottom). Configurations are labeled from left to right: A,B and C.

**3.2.3 Merging Arcs** The last phase consists of creating the final, genuine arcs by merging the selected candidates corresponding to the previously described criteria. The method developed here tries to find the three points of the equilateral triangle which is circumscribed by the merged arc circle. This increases the odds of having the best fitting circle as with this method we avoid choosing either noisy or numerically instable points. This procedure begins by choosing either the arc 1 or 2, and then computes the edge length of the circumscribed equilateral triangle

$$\text{edgeLength} = 2r_i \sin \frac{\pi}{3}, \quad (7)$$

where  $r_i$  is the radius of the circle.

Now, let  $\mathcal{R}_i$  be the subset of image curve points ( $\mathcal{Q}$ ) which are close to the arc candidate  $\mathcal{C}_i$

$$\mathcal{R} = \{p \in \mathcal{Q} | \delta(\mathcal{C}, p) < \text{RsdErr}(\mathcal{Q}, \mathcal{C})\}. \quad (8)$$

We can then define a partitioning of the points ( $\mathcal{D}_1$  and  $\mathcal{D}_2$ ) belonging to the two arc candidates, as well as their intersection  $\mathcal{I}$

$$\mathcal{I} = \mathcal{R}_1 \cap \mathcal{R}_2, \mathcal{D}_1 = \mathcal{R}_1 \setminus \mathcal{R}_2, \mathcal{D}_2 = \mathcal{R}_2 \setminus \mathcal{R}_1. \quad (9)$$

We then define  $p_3$  and  $p_1$  such that

$$p_3 p_1 = \min_{p_i \in \mathcal{L}, p_j \in \mathcal{D}_2} |\text{edgeLength} - p_i p_j| \quad (10)$$

and find  $p_2$  such that

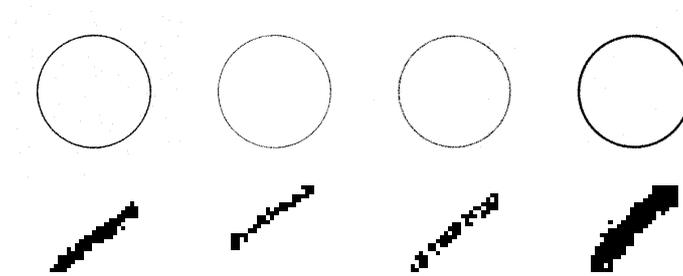
$$p_2 p_3 + p_2 p_1 = \max_{p_i \in \mathcal{D}_2} (p_i p_3 + p_i p_1). \quad (11)$$

If we consider that the initial arcs belong effectively to the same circle, this method constructs the equilateral triangle and gives three points of a same circle. The more the three points are distant from each other, the more accurate the construction of the new circle is.

In fact, we are likely to find a better distributed set of three points over a circle if instead of using  $\mathcal{D}_1$  and  $\mathcal{D}_2$  we use  $\mathcal{R}_1$  and  $\mathcal{R}_2$ .

## 4 Experiments

In collaboration with E. Barney Smith [10] we have been conducting an exhaustive survey of the influence of all possible internal parameters and external noise variations on the quality of the arc detection. Full analysis and report of this work is beyond the scope of this paper and will be published separately. Fig 7 shows some samples of used synthetic data for assessing the quality and precision of our approach.



**Fig. 7.** Various Degraded Synthetic Circles and Selected Zooms

The overall precision in circle detection is extremely high. Precision was measured using three different metrics:

**Circle Center Precision** is obtained by measuring the euclidian distance of the detected arc circle to the theoretical circle center.

**Circle Radius Precision** is obtained by measuring the absolute difference between the detected radius and the theoretical radius.

**Overlapping** is a metric correlating the two previous ones, and expresses the overlapping surface ratio of both circles. It is normalized to  $[0, 1]$ , where 1 signifies perfectly identical circles, and 0 perfectly disjoint circles.

Tested over a wide range of parameters ( $\tau$  outlier quantile, required coverage rates – *cf.* next section – ...) our method gives the following results:

	Worst Case	Best Case
Avg. Center Precision Error (pixels)	0.67	0.38
St.Dev. Center Precision Error	0.71	0.52
Avg. Radius Precision Error (pixels)	0.11	0.01
St.Dev. Radius Precision Error	0.37	0.12
Avg. Overlapping (%)	98.2	99.2
St.Dev. Overlapping (%)	6.0	2.3

This translates into estimation errors upto a pixel for center and radius. Coverage standard deviation might seem high for the announced detection precisions, but comes from situations where we tested on small circles (radius 8 pixels) where a single pixel shift accounts for a significant proportion of non overlapping.

Figures 2 to 3 show results on the GREC 2007 and 2009 contest images. The initial images are in black, while detected arcs are in Grey (right column).

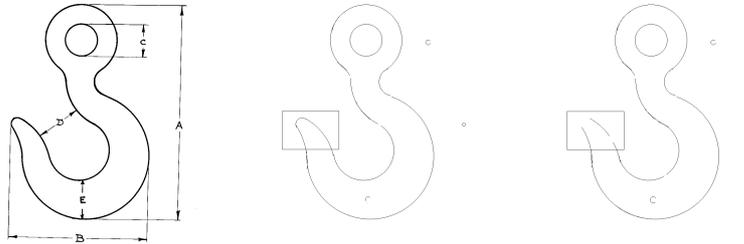
#### 4.1 Parameters and their Influence

All parameters mentioned here are either direct transpositions of the algorithm described in this paper, or are direct call parameters of the software available for download (*cf.* note below).

One parameter that has an influence on determining whether two arcs are partial estimates of a same global arc is `RatioRadiusError` (*cf.* section 3.2.2). To compare the radii of two arcs, experiments show that a value of 64% for `RatioRadiusError` makes the merge possible for most arcs having common black pixels and avoids merging arcs with significantly different radii as shown in Figure 8. For the image in Figure 8 65% was too high and resulted in a loss of precision as shown within the rectangle.

The `filterCoverage` parameter is used to keep only those arc candidates that have a sufficient percentage of pixels effectively lying on pixels of the image. This process uses the original image instead of the skeleton image. The coverage percentage of `filterCoverage` is set to 89% to ensure keeping only accurate estimates.

`preFilterCoverage` checks if the percentage of pixels of the discrete estimate of the given arc lying effectively on black pixels of the original image, oversteps a cover percentage and returns a boolean. This process uses the original image instead of the skeleton image. The cover percentage of `preFilterCoverage` can be lower than the cover percentage defined for `filterCoverage`, as some intermediate arcs that do not withstand the cover percentage of `filterCoverage`



**Fig. 8.** RatioRadiusError tuning; the original image; 64% filter; 65% filter

might be merged with another arc resulting in an arc that does withstand the cover percentage of `filterCoverage`. The percentage 89% proved to be a good value for the cover percentage of `preFilterCoverage`, while 90% removed some good circles. This is often due to the fact that in reality, the images have slight deformations, an perceived circles are actually ellipses.

The merging algorithm can be improved by increasing the aperture of the arcs “virtually”. In other words, each arc has two angles and two apertures. The “virtual” angle and aperture are those which are increased and used to check if two arcs have to be merged. The actual angle and aperture are not changed. In fact, while increasing the actual aperture, the discrete arc of an estimate might no longer withstand the `preFilterCoverage` processing, as it is likely to have more pixels that are not black. Thus by keeping the initial angle and aperture unchanged we maintain arcs that were not merged. Moreover when performing the merge, the points used to construct the new arc are from the closest points of the actual arc, which is more accurate especially if we choose the points from those who belong as well to the black pixels of the image. The “virtual” increase uses the threshold defined in 3.2.1, namely a value of 3.

## 5 Conclusion and Further Work

In this paper we have presented a highly efficient and complete arc detection algorithm that needs extremely few parameters or contextual knowledge to operate. We have validated it on quite difficult images, coming from the GREC 2007 and 2009 contest. Further work will include stroke width integration in order to obtain a more precise localization of the arcs, as well as a more quantitative assessment of the positioning and localization of the detected arcs.

## Acknowledgments

Authors acknowledge funding from the PROCORE-FRANCE/HONG KONG JOINT RESEARCH SCHEME (F-HK04/05T, 9050187): Knowledge represen-

tation issues for the performance evaluation of graphic symbol recognition methods. B. Lamiroy more particularly thanks Prof. Liu Wenyin for having received him at the City University of Hong Kong for finalizing this work. Authors acknowledge reporting snippet of yet unpublished work in collaboration with E. Barney Smith on evaluation of noise influence in section 4. Furthermore, the circle overlapping metric described in section 4, is the result of fruitful, informal discussions with E. Magagnin. B. Lamiroy was a visiting scientist at Lehigh University at the time of publication of this article.

### Access to Source Code

The source code of this work is available for download and evaluation under LGPL at <http://gforge.inria.fr/projects/visuvocab/>.

### References

1. Olson, C.F.: Constrained Hough Transforms for Curve Detection. *Computer Vision and Image Understanding* **73**(1) (March 1999) 329–345
2. Chen, T.C., Chung, K.L.: An Efficient Randomized Algorithm for Detecting Circles. *Computer Vision and Image Understanding* **83**(2) (August 2001) 172–191
3. Samiti di Baja, G.: Well-Shaped, Stable, and Reversible Skeletons from the (3,4)-Distance Transform. *Journal of Visual Communication and Image Representation* **5**(1) (1994) 107–115
4. Rosin, P.L., West, G.A.: Segmentation of Edges into Lines and Arcs. *Image and Vision Computing* **7**(2) (May 1989) 109–114
5. Hilaire, X., Tombre, K.: Robust and Accurate Vectorization of Line Drawings. *IEEE Transactions on PAMI* **28**(6) (June 2006) 890–904
6. Lamiroy, B., Gaucher, O., Fritz, L.: Robust Circle Detection. In: *Proceedings of 9th International Conference on Document Analysis and Recognition, Curitiba (Brazil)*. (2007) 526–530
7. Fischler, M.A., Bolles, R.C.: Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM* **24**(6) (1981) 381–395
8. Rousseeuw, P.J., Leroy, A.M.: *Robust Regression and Outlier Detection*. Wiley Series in Probability and Mathematical Statistics. John Wiley and Sons (1987)
9. Berman, M.: Large Sample Bias in Least Squares Estimators of a Circular Arc Center and Its Radius. *Computer Vision, Graphics and Image Processing* **45** (1989) 126–128
10. Smith, E.H.B.: Characterization of image degradation caused by scanning. *Pattern Recognition Letters* **19**(13) (1998) 1191–1197

# Symbol Recognition using Spatial Relations

Santosh K.C.<sup>a,\*</sup>, Bart Lamiroy<sup>b</sup>, Laurent Wendling<sup>c</sup>

<sup>a</sup>LORIA – INRIA, 615 rue Jardin Botanique, 54600 Villers-lès-Nancy, France

<sup>b</sup>LORIA – Nancy Université, BP 239 - 54506 Vandoeuvre-lès-Nancy Cedex, France

<sup>c</sup>LIPADE, Université Paris Descartes, 75270 Paris Cedex 06, France

---

## Abstract

In this paper, we present a method for symbol recognition based on the spatio-structural description of a ‘vocabulary’ of extracted visual elementary parts. It is applied to symbols in electrical wiring diagrams. The method consists of first identifying vocabulary elements into different groups based on their types (e.g., *circle*, *corner*). We then compute spatial relations between the possible pairs of labelled vocabulary types which are further used as a basis for building an Attributed Relational Graph that fully describes the symbol. These spatial relations integrate both topology and directional information.

The experiments reported in this paper show that this approach, used for recognition, significantly outperforms both structural and signal-based state-of-the-art methods.

*Keywords:* Vocabulary, Spatial Relations, Attributed Relational Graph, Symbol Recognition.

---

## 1. Introduction

### 1.1. Motivation

Symbol recognition – the core part of graphical document image analysis and recognition systems – plays an important role in a variety of applications such as automatic recognition and interpretation of circuit diagrams [[Okazaki](#)

---

\*Corresponding author

*Email addresses:* [Santosh.KC@inria.fr](mailto:Santosh.KC@inria.fr) (Santosh K.C.), [Bart.Lamiroy@loria.fr](mailto:Bart.Lamiroy@loria.fr) (Bart Lamiroy), [Laurent.Wendling@parisdescartes.fr](mailto:Laurent.Wendling@parisdescartes.fr) (Laurent Wendling )

et al., 1988], engineering drawings [Yang et al., 2007] and architectural drawings [Lladós et al., 2001; Valveny and Martí, 2003], maps [Samet and Soffer, 1996], musical notations [Rebelo et al., 2010], mathematical expressions [Chaudhuri and Garain, 2000], as well as optical characters [Yuen et al., 1998]. Therefore, a symbol can be defined as a graphical entity with a particular meaning in the context of a specific domain.

Research on graphics recognition has an extremely rich state-of-the-art literature, aimed to localise/recognise symbols depending on the applications. [Cordella and Vento, 2000; Lladós et al., 2002] show that these methods are particularly suited for isolated line symbols, not for composed symbols connected to a complex environment. In order to exploit the information embedded in those documents, one needs to be able to extract visual parts and formalise the possible links that exist between them. This combination of symbol localisation based on extracted visual parts is going to be the core of this paper and is very much inspired by a real world industrial problem [Tombre and Lamiroy, 2008; K.C. et al., 2009]. It consists in identifying a set of known symbols in aircraft electrical wiring diagrams, in order to bootstrap simulation algorithms. The main challenges come from the fact that the test symbols come in a wide variety of different forms. Symbols may either be very similar in shape, and only differ by slight details – or either be completely different from a visual point of view. Symbols may also be composed of other known and significant symbols and need not necessary be connected.

The rest of the paper is organised as follows. An overview of pertinent literature is given in Section 1.2, followed by a brief explanation of our proposed method in Section 2. We explain the way we describe symbols in Section 3, which mainly includes the concept of using spatial relations. We derive a symbol matching method from it in Section 4. Full experiments are reported in Section 5 and confront our method with current state-of-the-art algorithms. It includes a comprehensive experimental result analysis. We conclude in Section 6.

## 1.2. State-of-the-Art

### 1.2.1. Symbol Representations

Symbol recognition is a particular application of pattern recognition. Existing approaches, specifically those based on feature based matching, can be sorted into three classes: statistical, structural and hybrid. As respective

examples, among others, one can cite [Yang, 2005; Zhang et al., 2006; Lladós et al., 2001; Yang, 2005].

Under statistical approaches, global signal-based descriptors [Yuen et al., 1998; Kim and Kim, 2000; Tabbone et al., 2006; Belongie et al., 2002; Zhang and Lu, 2002, 2004] are usually quite fault tolerant to image distortions, since they tend to filter out small detail changes. This is unfortunately an inconvenient in our context. Moreover, they difficultly accommodate with connected or composite symbols. For instance, when symbols are combined, approaches that rely on centroid detection like [Yuen et al., 1998] tend to fail. Others, like Shape Context [Belongie et al., 2002] are sensible to occlusions on the symbol boundaries. Overall, they are generally not well adapted for capturing small detail changes, since they are specifically conceived to filter those out. In these statistical approaches, signatures are simple with low computational cost. However, discrimination power and robustness strongly depend on the selection of optimal set of features for each specific application.

Besides global signal-based descriptors, another idea is to decompose the symbols into either vector based primitives like points, lines, arcs *etc.* or into meaningful parts like *circles, triangles, rectangles etc.* These methods fall under structural approaches. They are then represented as Attributed Relational Graphs (ARG) [Bunke and Messmer, 1995; Conte et al., 2004], Region Adjacency Graphs (RAG) [Lladós et al., 2001], constraint networks [Ah-Soon and Tombre, 2001] as well as deformable templates [Valveny and Martí, 2003]. Their common drawback comes from error-prone raster-to-vector conversion. Those errors can increase confusions among different symbols. Furthermore, variability of the size of graphs leads to computational complexity in matching. However, structural approaches provide a powerful representation, conveying how parts are connected to each other, while also preserving generality and extensibility.

To describe the symbols, it is necessary to handle relations between the decomposed parts. The following paragraph gives an overview of existing work on spatial relations and their proper usages.

### 1.2.2. Spatial Relations

Effects of spatial relations on recognition performance have been examined comprehensively for scene understanding, document analysis and recognition tasks [Biederman, 1972; Bar and Ullman, 1993; Xiaogang et al., 2004; Pham and Smeulders, 2006]. Spatial relations can be either topological [Egenhofer and Franzosa, 1991; Egenhofer and Herring, 1991; Papadias

et al., 1995] directional [Bloch, 1999; Matsakis and Wendling, 1999; Wang and Keller, 1999] and metric in nature. For example, topological configurations are handled in [Xiaogang et al., 2004] with a few predicates like *intersection*, *interconnection*, *tangency*, *parallelism* and *concentricity* expressed with standard topological relations as described in [Egenhofer and Herring, 1991].

In a similar way, various directional relation models have been developed for a wide range of different situations.

- If the objects are far enough from each other, their relations can be approximated by their centres based on the discretised angle [Miyajima and Ralescu, 1994]. This approach is robust to small variations of shape and size.
- If they are neither too far nor too close, relations can be approximated by their *Minimum Bounding Rectangle* (MBR) [Lee and Hsu, 1992; E.Jungert, 1993; Papadias et al., 1995; Papadias and Theodoridis, 1997] as long as they are regular.
- Approaches like *Angle Histograms* [Wang and Keller, 1999] tend to be more capable of dealing with overlapping, something the previous ones have difficulties with. However, since they consider all pixels of a shape, their computational cost increases dramatically.
- Other methods, like *F-Histograms* [Matsakis and Wendling, 1999] use pairs of longitudinal sections instead of pairs of points, also at the cost of high time complexity.
- Another well-known approach uses fuzzy landscapes [Bloch, 1999], and is based on fuzzy morphological operators.

Previously mentioned approaches address only either topological or directional relations. Managing both comes at high computational costs. Even then, no existing model fully integrates topology. They rather have various degrees of sensitivity to or awareness of topological relations. While methods like [Xiaogang et al., 2004] focus on topological information only, our approach unifies both topological and directional information into one descriptor [K.C. et al., 2010] without any additional running time cost.

Placing spatial relations in the context of recognition and symbol description, one should note that spatial relations also have a language-based component (related to human understanding e.g., to the *right* of) that can be

formalised in a mathematical way (e.g., the 512 relations of the 9–intersection model [Egenhofer and Herring, 1991]). Therefore, qualitative and quantitative relations are another way to do categorisation of spatial relations. Consider an example, an object  $\mathbb{A}$  extending from *Right* (98%) to *Top* (2%) with respect to  $\mathbb{B}$  is expressed as *Right – Top*( $\mathbb{A}, \mathbb{B}$ ). This spatial predicate remains unchanged upto a reasonable change of the objects’ shape and position. Taking this into account, our work uses more natural relations than the all–or–none nature of standard relations [Freeman, 1975].

In the following section, we explain our proposed method by focusing on using spatial relations for describing and matching symbols.

## 2. Proposed Recognition Method

Our recognition method is based on a spatio–structural description of extracted visual parts that compose a symbol. This means that, to describe a symbol, we compute spatial relations between previously extracted visual parts. Without any other consideration, it is obvious that the size of the resulting relational graph is potentially very large and variable from one symbol to another. However, when grouping visual parts together according to their types (e.g., *circle*, *corner*) and by labelling them accordingly (see Section 3.1), we can eliminate all the combinatorial problems inherent to graph matching, without sacrificing recognition quality or expressive power.

We compute the spatial relations (see Section 3.2) between the distinct labelled attributes for building an Attributed Relational Graph (ARG – see Section 3.3), achieving at the same time integration of both topological and directional information.

Since each vertex represents a different class of visual parts, the graph has a uniquely and distinctly labelled vertex set. Vertex and edge matching thus becomes trivial and can be done in near–constant time.

## 3. Symbol Description

As mentioned in Section 2, we first define our visual vocabulary in Section 3.1. In Section 3.2 we explain the way we compute pairwise spatial relations and finally use both in Section 3.3 to build an ARG and completely describe the symbol.

### 3.1. Visual Vocabulary

We define a set of well controlled visual elementary parts as a *vocabulary* [K.C. et al., 2009]. While, in the general case, this vocabulary can be of any kind from any type of bag-of-features, related to what is visually pertinent in the application context under consideration, our current vocabulary is related to electrical symbols. It can be easily extended to adapt to other domains. Such visual elementary parts are extracted with the help of image treatment analysis operations as described in [Rendek et al., 2004]. Shortly, we discuss on how we accomplished it.

- *thick primitive*: We employ straight forward thin/thick separation by counting all *thick* connected components within the image. It simply uses standard skeletonisation using chamfer distance and computes the histogram of line thickness. An optimal cut value is computed from the histogram to distinguish between thick zones and thin zones.
- *circle primitive*: We use the algorithm as described in [Lamiroy and Guebbas, 2010] which is based on Random Sample Consensus minimization.
- *corner primitive*: We mainly consider four types of corners such as *North-East*, *North-West*, *South-East* and *South-West*. It uses simple template matching process *i.e.*, if the ratio of black and white pixels is greater than or equal to the template threshold, then the presence of corner is accessed.
- *extremity primitive*: We approach to detect loose end coordinates  $p_x$  from a given skeleton pixel where there is only a unique neighbouring pixel  $p_{c_1}$  connecting to the main skeleton, which itself is connected by a unique neighbouring pixel  $p_{c_2}$ .

Fig. 1 shows an illustration of visual primitives, extracted from two different symbols. Rather than using every detected element as a basis for expressing and computing spatial relations, we group them together by type as shown in Fig. 1. We denote the set of these generated groups as,  $\sum_{\mathbb{T}} = \{\mathbb{T}_{thick}, \mathbb{T}_{circle}, \mathbb{T}_{corner}, \mathbb{T}_{extremity}\}$ .

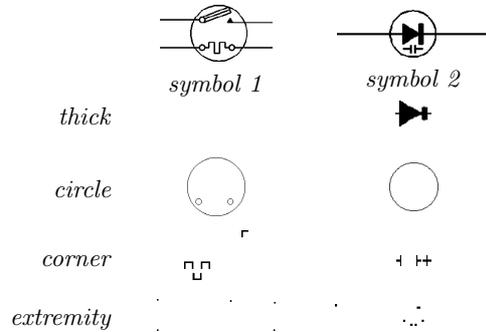


Figure 1: Illustration of vocabulary type.

### 3.2. Spatial Relation

In order to express the spatial distribution of the previously formed groups, we compute a spatial signature  $\mathfrak{R}$  (defined further in Eq. (1)), expressing the spatial relations between two sets of pixels  $\mathbb{A}$  and  $\mathbb{B}$ . This section explains in detail how it is computed.

Pairwise spatial relations are often expressed by using one of the objects as reference. For example,  $\mathbb{A}$  is to the *right* of  $\mathbb{B}$ :  $right(\mathbb{A}, \mathbb{B})$ , where  $\mathbb{B}$  is referenced. In our context, since the number of vocabulary types is not always the same for all symbols, it is difficult to take a particular type as a reference. To avoid such a difficulty, we first set up a unique reference point from each pair as shown in Step 1, hereafter. Then, we compute directional relations with respect to the reference point, thus avoiding potential ambiguity.

#### Step 1. Unique Reference Point Set

We consider a unique reference set  $\mathbb{R}$ , defined by the topology of the minimum bounding rectangles (MBR) of  $\mathbb{A}$  and  $\mathbb{B}$  and with the help of the 9-intersection model [Egenhofer and Herring, 1991]. In connection with [Renz and Nebel, 1998],  $\mathbb{R}$  is either the common portion of two neighbouring sides in the case of *disconnected* MBRs or the intersection in the case of *overlapping, equal* or otherwise *connected* MBRs. To do this, we simply check topological relations between them in a 9-dimensional binary space via the use of intersections of the boundaries, interiors and exteriors of two sets  $\mathbb{A}$  and  $\mathbb{B}$ .

Depending on the obtained topological configurations,  $\mathbb{R}$  can range from a point to a rectangular  $2D$  area. In what follows, we define its

centroid point  $\mathbb{R}_p$  as our reference point for computing spatial relation  $\mathfrak{R}$  between  $\mathbb{A}$  and  $\mathbb{B}$ .

**Step 2.** Directional Relation

For a given reference point  $\mathbb{R}_p$ , we cover the surrounding space at regular radial intervals of  $\Theta = 2\pi/m$ . As shown in Fig. 2 (a), a radial-line rotates over a cycle, and when intersecting with object  $\mathbb{X}$  ( $\mathbb{A}$  or  $\mathbb{B}$ ), generates a boolean histogram  $\mathcal{H}$ ,

$$\mathcal{H}(\mathbb{X}, \mathbb{R}_p) = [I(\mathbb{R}_p, j\Theta)]_{j=0, \dots, m-1}$$

where

$$I(\mathbb{R}_p, \theta_i) = \begin{cases} 1 & \text{if } \text{line}(\mathbb{R}_p, \theta_i) \cap \mathbb{X} \neq \emptyset \\ 0 & \text{otherwise.} \end{cases}$$

This boolean histogram expresses whether there are any black pixels in direction  $\theta_i$ . We extend this direction histogram, without loss of generality, to a histogram covering sectors defined by two successive angle values. Furthermore, rather than using boolean values, we can account for the percentage of pixels of the whole object lying in the general direction  $\theta_i$ . Fig. 2 (b) gives an example for both types of histogram, boolean and percentage.

Applying this to both objects  $\mathbb{A}$  and  $\mathbb{B}$ , our spatial relational signature  $\mathfrak{R}(\mathbb{X}, \mathbb{R}_p)$  is the set of both histograms

$$\mathfrak{R}(\mathbb{X}, \mathbb{R}_p) = \{\mathcal{H}(\mathbb{A}, \mathbb{R}_p), \mathcal{H}(\mathbb{B}, \mathbb{R}_p)\}. \quad (1)$$

It is important to understand that we know the visual vocabulary types to which  $\mathbb{A}$  and  $\mathbb{B}$  belong (*cf.* Section 3.1). Defining a fixed arbitrary order on the set of types  $\sum_{\mathbb{T}}$  solves the potential ordering problem when comparing two relational signatures.

Fig. 3 and Fig. 4 provide illustrations on hand-drawn and real-world examples respectively. The illustrations show how the reference point set  $\mathbb{R}$  is obtained and show the corresponding histogram  $\mathcal{H}$ . They are analysed in the following section.

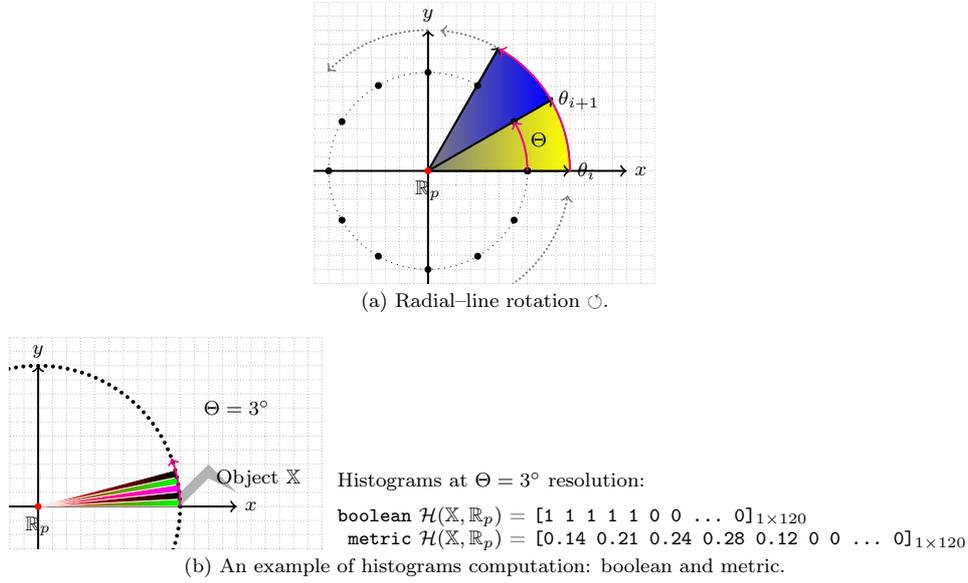


Figure 2: Computing spatial relations using radial-line rotation.

**Illustrations.** In Fig. 3 we show how our method adapts to different topological configurations. This section is not intended to give a full and formal evaluation of our approach, but rather to provide the user with an intuitive feeling on how it behaves. The figure shows various computed histograms in different configurations. These configurations were chosen to cover most topological relations between two objects one may encounter.

Let us consider the first three instances (a), (b) and (c). Keeping (a) as a reference image, we have changed a stroke thickness without changing relative positioning in (b) and moved objects closer while keeping identical topological configuration in (c). We observe that histograms do not show any significant difference. Scaling does not affect our method since  $\mathcal{H}$  is normalised. In addition, the line rotation does not consider distance (*far* or *near*) information as long as it does not change the angular positioning.

For false *overlapping* configurations, as shown in (d), the coverage angle of  $\mathcal{H}$  changes due to the change in structure (elongating horizontal limb in both objects).

For all *inclusion* configurations (like the false inclusion depicted in (f), but equally for full inclusion situations), our method does not produce any histogram for the component  $\mathbb{X}$  which is either *contained* in or *covered* by

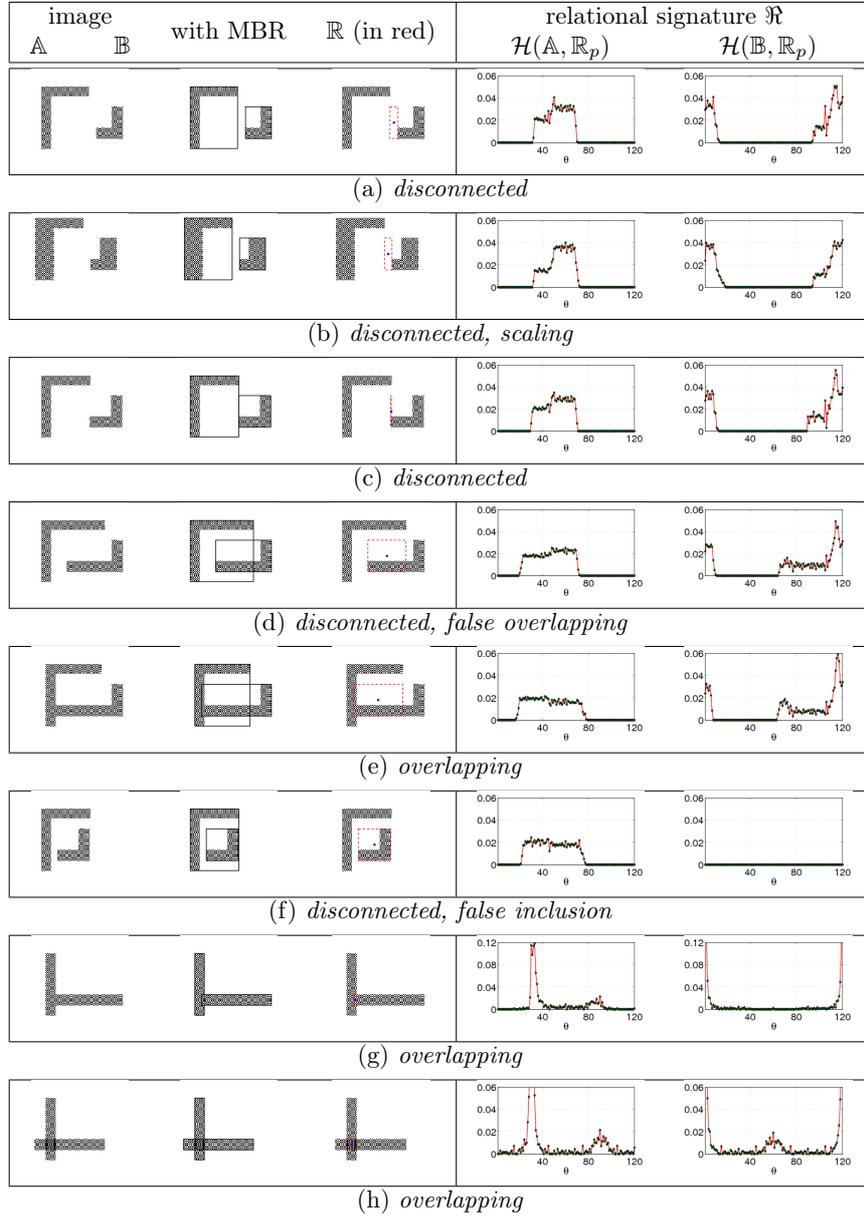


Figure 3: Histograms at  $3^\circ$  resolution for a few hand-drawn spatial objects  $\mathbb{A}$  and  $\mathbb{B}$ , having different topological configurations.

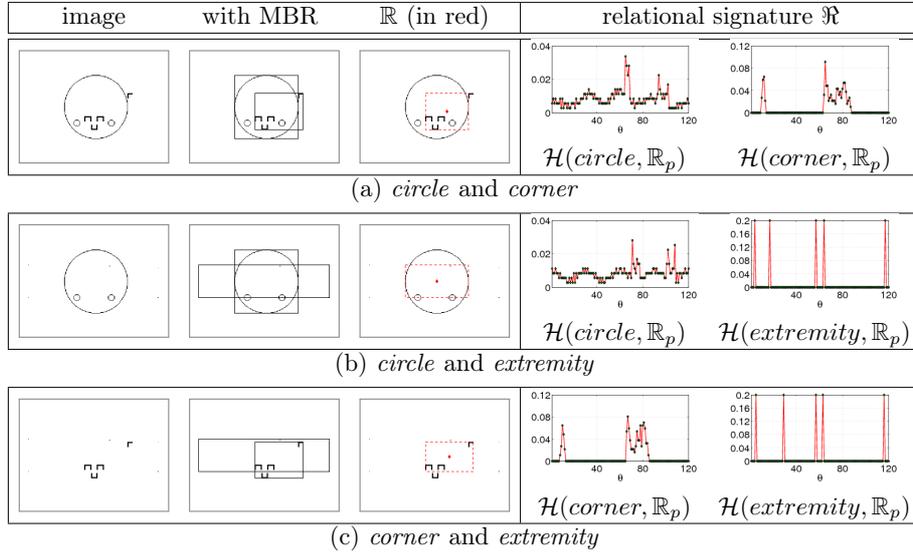


Figure 4: Histograms at  $3^\circ$  resolution for all possible pairs of vocabulary types from a *symbol 1* as shown in Fig. 1.

the other: it is simply  $\mathcal{H}(\mathbb{X}, \mathbb{R}_p) = \emptyset$ .

Besides, the difference of histograms between two *overlapping* cases in (g) and (h) can be observed in the middle of  $\mathcal{H}(\mathbb{B}, \mathbb{R}_p)$  (between  $40^\circ - 80^\circ$ ). This provides the fact that the method is able to discriminate slight changes in the object configurations even when identical topology exists.

Fig. 4 represents similar configurations taken from a real-world example, using the vocabulary extraction described in Section 3.1

**Remarks.** Our method captures the spatial information by the angular positions in the histogram. The magnitude of the histogram contains the structural information. Furthermore, running time does not depend on the size of the spatial objects as in the *Angle Histogram* approaches [Miyajima and Ralescu, 1994], for instance. Our method simply counts the number of pixels in every sector made by two consecutive radial-lines while rotating. However, running time is fixed and entirely depends on the parameter  $\Theta$  (rotation step) that defines the resolution of  $\mathcal{H}$  and the global size of the image. Its value is a trade-off between precision and execution time. We establish the optimal resolution for our application in Section 5.

### 3.3. Attributed Relational Graph

The vocabulary developed in Section 3.1 consists of a set of fixed label attributes, while the spatial relations between the attributes are the histograms described in Section 3.2. This gives us all the elements to express symbols as a complete ARG in which each vertex represents a distinct attribute type and the edges are labelled with a numerical expression of the spatial relations  $\mathfrak{R}$ .

More formally, we express the ARG as a 4-tuple  $G = (V, E, F_A, F_E)$  where

$V$  is the set of vertices;

$E \subseteq V \times V$  is the set of graph edges;

$F_A : V \rightarrow A_V$  is a function assigning labelled attributes to the vertices where  $A_V$  is the set of attributes type set  $\sum_{\mathbb{T}}$  (cf. Section 3.1) and

$F_E : E \rightarrow \mathfrak{R}_E$  is a function assigning labels to the edges where  $\mathfrak{R}$  represents the spatial relation of the edge  $E$  (cf. Section 3.2). Note that  $\mathfrak{R}$  does not provide symmetry,  $\mathfrak{R}(\mathbb{A}, \mathbb{B}) \neq \mathfrak{R}(\mathbb{B}, \mathbb{A})$ . But, this can be solved by fixed ordering of  $V$  and  $\mathfrak{R}$  is not affected.

For instance, using *symbol 1* in Fig. 1 as an example, and its corresponding spatial relations in Fig. 4 we obtain the following ARG representation:  $G = \{$

$$\begin{aligned} V &= \{\mathbb{T}_1, \mathbb{T}_2, \mathbb{T}_3\}, \\ E &= \{(\mathbb{T}_1, \mathbb{T}_2), (\mathbb{T}_1, \mathbb{T}_3), (\mathbb{T}_2, \mathbb{T}_3)\}, \\ F_A &= \{(\mathbb{T}_1, \mathbb{T}_{circle}), (\mathbb{T}_2, \mathbb{T}_{corner}), (\mathbb{T}_3, \mathbb{T}_{extremity})\} \\ F_E &= \{((\mathbb{T}_1, \mathbb{T}_2), \mathfrak{R}(\mathbb{T}_1, \mathbb{T}_2)), ((\mathbb{T}_1, \mathbb{T}_3), \mathfrak{R}(\mathbb{T}_1, \mathbb{T}_3)), \\ &\quad ((\mathbb{T}_2, \mathbb{T}_3), \mathfrak{R}(\mathbb{T}_2, \mathbb{T}_3))\} \end{aligned}$$

This forms a complete graph, and therefore has  $r = \frac{t(t-1)}{2}$  edges for  $t$  attribute types.

## 4. Symbol Recognition

Now that we have set up our ARG for symbol representation, we can define our recognition process. Recognition based on maximal similarity, measured by a matching score. The score is purely based on matching the corresponding relational signatures between the two given ARGs.

We then further extend the recognition by ranking database symbols based on the order of similarity, both of which will be explained in this section.

#### 4.1. Matching

Following the ARG description in Section 3.3, let us consider two graphs:

$$\begin{aligned} G^q &= (V^q, E^q, F_A^q, F_E^q) \text{ for the query symbol and} \\ G^d &= (V^d, E^d, F_A^d, F_E^d) \text{ for the database symbol.} \end{aligned}$$

Let us remind that the set of vertices  $V$ , with  $|V| = t$  and set of edges  $E$ , with  $|E| = r$ .

In order to explain our matching strategy, we are first taking the simplifying assumption that  $V^q$  and  $V^d$  are identical. In other words, both symbols contain items corresponding to identical vocabulary elements, but not necessarily sharing the same spatial arrangement. Since in our ARG every single vertex bears one distinct and unique attribute type, there is no cost in matching the vertices between  $G^q$  and  $G^d$ . As a consequence, matching edges is equally straightforward.

Since we have temporarily taken the assumption that  $V^q$  and  $V^d$  contain the same vocabulary elements, we can set up a bijective matching functions  $\varphi : V^q \rightarrow V^d$  and  $\sigma : E^q \rightarrow E^d$ . This bijection exists such that  $uv$  is an edge in graph  $G^q$  if and only if  $\varphi(u)\varphi(v)$  is an edge in graph  $G^d$ . Also we consider that ordering is preserved over the vertices sets  $V^q$  and  $V^d$ . *I.e.*  $v_1 < v_2 \Rightarrow \varphi(v_1) < \varphi(v_2)$ .

Thanks to our fixed labelling of attribute types, corresponding  $\mathfrak{R}$  alignment is possible between the two given graphs and we can provide a matching score between the two given graphs  $G^q$  and  $G^d$ ,

$$dist.align(G^q, G^d) = \sum_{r \in E} \delta(F_E^q(r), F_E^d(\sigma(r)))$$

where  $\delta(a, b) = \sqrt{\sum_{l=1}^L (a_l - b_l)^2}$ . This is actually a very simple and straightforward metric. Given the performances of our method reported in Section 5 there is no real need to have a more complex one, unless rotational invariance is needed.

Of course, the assumption that  $V^q$  and  $V^d$  share the exact same vocabulary is too strong. To generalise the previously described approach to any

situation, we define a binary (indicator) function  $\tau_A^V : \Sigma_{\mathbb{T}} \rightarrow \{0, 1\}$  to check the presence of vertices in the ARG, where the value of  $\tau_A^V(\mathbb{T})$  is 1 if  $\mathbb{T}$  is present in  $V$  and 0, otherwise. For example, for the *symbol 1* shown in Fig. 1,  $\tau_A^V = [0, 1, 1, 1]$ . This refers to the absence of *thick* components and the presence of *circle*, *corner* and *extremity* components.

We can then use a simple edit cost between  $\tau_A^{V^q}$  and  $\tau_A^{V^d}$  defined by the number of edge deletions/insertions or substitutions. To do this, we first note the number of vertices to be deleted/inserted or substituted. Then consider the number of adjacent links.

$$\text{dist.edit}(G^q, G^d) = \sum_{p=1}^P c(o_p), \forall (o_1, \dots, o_P) \in \Upsilon(G^q, G^d)$$

where  $\Upsilon(G^q, G^d)$  denotes the set of edit paths transforming  $G^q$  into  $G^d$  and  $c$ , the edit cost function for operation  $o_p$ . Once virtual links exist (*i.e.*, null relation) after insertion for instance, edit cost is carried out as if matching has been done with relational alignments.

The final matching score or distance in the generic case  $D(G^q, G^d)$  therefore is obtained from the fusion of edit cost and relational signatures alignment (reduced to the common node types between the two graphs). More formally, distance (matching score) between two given graphs is,

$$D(G^q, G^d) = \alpha \text{dist.align}(G^q, G^d) + (1 - \alpha) \text{dist.edit}(G^q, G^d)$$

where  $\alpha \in [0, 1]$ . The parameter  $\alpha$  provides weight while matching. In our experiments we use  $\alpha = 0.5$ . The excellent results obtained and reported in Section 5.2 have not required us to tune this parameter further.

In the following, we give an example of how distance is computed.

**Example.** Consider  $V^q = \{\mathbb{T}_1^q, \mathbb{T}_2^q\}$   in  $G^q$  and  $V^d = \{\mathbb{T}_1^d, \mathbb{T}_2^d, \mathbb{T}_3^d\}$   in  $G^d$ . Then matching score between them is,

$$D(G^q, G^d) = \alpha \underbrace{[\delta(\mathfrak{R}_{\mathbb{T}_1^q, \mathbb{T}_2^q}, \mathfrak{R}_{\mathbb{T}_1^d, \mathbb{T}_2^d})]}_{\text{dist.align}} + (1 - \alpha) \underbrace{\left[ \begin{array}{c} \delta(\mathfrak{R}_{\mathbb{T}_1^q, \mathbb{T}_3^q}, \mathfrak{R}_{\mathbb{T}_1^d, \mathbb{T}_3^d}) \\ + \\ \delta(\mathfrak{R}_{\mathbb{T}_2^q, \mathbb{T}_3^q}, \mathfrak{R}_{\mathbb{T}_2^d, \mathbb{T}_3^d}) \end{array} \right]}_{\text{dist.edit}}$$

where  $\mathfrak{R}_{x,y} = \mathfrak{R}(x, y)$ . It is clear that  $\mathbb{T}_3^q$  has to be inserted in  $G^q$  in order to transform it to  $G^d$ . As a consequence, virtual connections:  $\mathfrak{R}_{\mathbb{T}_1^q, \mathbb{T}_3^q}$  and  $\mathfrak{R}_{\mathbb{T}_2^q, \mathbb{T}_3^q}$

exists. Then matching is straightforward due to the labelled vertices in ARG. In addition, the weighting parameters are now useful to select either only *dist.align* or taking both with equal as well as with different weights.

#### 4.2. Ranking

The previously defined matching score conveys how similar/dissimilar a database symbol is with respect to a query. In order for similarity to be ranging from 1 to 0, we normalise  $D(\cdot)$  to  $[0, 1]$  by taking all database symbols:  $\overline{D}(\cdot) = \frac{D(\cdot) - D^{min.}(\cdot)}{D^{max.}(\cdot) - D^{min.}(\cdot)}$ . Now, the similarity is,

$$Similarity(G^q, G^d) = 1 - \overline{D}(G^q, G^d).$$

Ranking can therefore be based on the decreasing order of similarity.

## 5. Experiments

In this section, we first give an overview of the symbols in our dataset and explain how we have labelled them with ground-truth. Then we discuss the evaluation metric, clarifying its proper usage for this application. Based on the metric, we perform a series of experiments and confront our method with the existing ones.

In the very beginning of the experiment, we consider the influence of different resolutions  $\Theta$  in our relational signature. Once an optimal resolution is chosen, our spatial relation is compared with fundamental spatial relation models: *Cone-shaped* [Miyajima and Ralescu, 1994], *Angle Histogram* [Wang and Keller, 1999] and MBR [Papadias and Theodoridis, 1997]. Then we perform another assessment in order to make comparison of the complete method with the state-of-the-art approaches. For this, we first take a few representative global signal-based descriptors: region based *Zernike Moments* (ZM) [Kim and Kim, 2000], *Generic Fourier Descriptors* (GFD) [Zhang and Lu, 2002], *Shape Context* (SC) [Belongie et al., 2002] and *R-signature* [Tabbone et al., 2006], applied directly to the symbol. Then we take two recent pixel-based approaches: Statistical Integration of Histogram Array (SIHA) Yang [2005] and 2D kernel density Zhang et al. [2006] based symbol representation.

### 5.1. Dataset and Ground-truth Formation

**Dataset.** We work on a real world industrial problem to identify a set of different known symbols in aircraft electrical wiring diagrams as in [Tombre

and Lamiroy, 2008; K.C. et al., 2009]. Fig. 5 gives some examples of symbols in the database. Symbols may either be very similar in shape – and only differ by slight details – or either be completely different from a visual point of view. Symbols may also be composed of other known and significant symbols and need not necessary be connected. It is composed of roughly 500 different known symbols. Our dataset is completely unlabelled and imbalanced *i.e.*, neither ground-truth is given nor identical number of similar symbols exist for all queries.

**Ground-truth Formation.** Since there is no absolute ground-truth associated to our dataset, we have proceeded by using human validation, but by taking care of eliminating subjective bias. In order to achieve this we have asked 6 volunteers to manually select what they consider as “similar” symbols, for all queries executed in this section. Human evaluators have chosen the candidates which have similar visual overall appearance or which have significantly similar parts with respect to the chosen query. In our testing protocol, we consider that a result returned from an algorithm is correct if at least one human evaluator has selected the same result among the similar items. In more formal terms, for each query the “ground-truth” is considered to be the set of symbols formed by the union of all human selected sets. Fig. 5 provides a few examples. For instance, for query *a1*, evaluators have provided a list of symbols which they consider visually close, or containing parts that are visually close. The evaluators were not required to provide any ranking order nor degree of visual resemblance.

## 5.2. Experimental Protocol and Results

### 5.2.1. Experimental Protocol

**Evaluation Metric.** Our aim is not only limited to distinguish symbols but also extended to rank symbols in the provided lists. Ranking is related to similarity based on distance measure as described in Section 4.2. It becomes clear from Fig. 5 that there is a different number of pertinent documents in the database for each query. In order to be able to report retrieval results in a coherent way, we choose not to use classical precision and recall but to use the retrieval efficiency measure instead [Kankanhalli et al., 1995]. Retrieval efficiency has the advantage not to degenerate when the ranking parameter  $K$  grows bigger than the number of relevant items in the database, as would have been the case for precision and recall. For every chosen query, efficiency

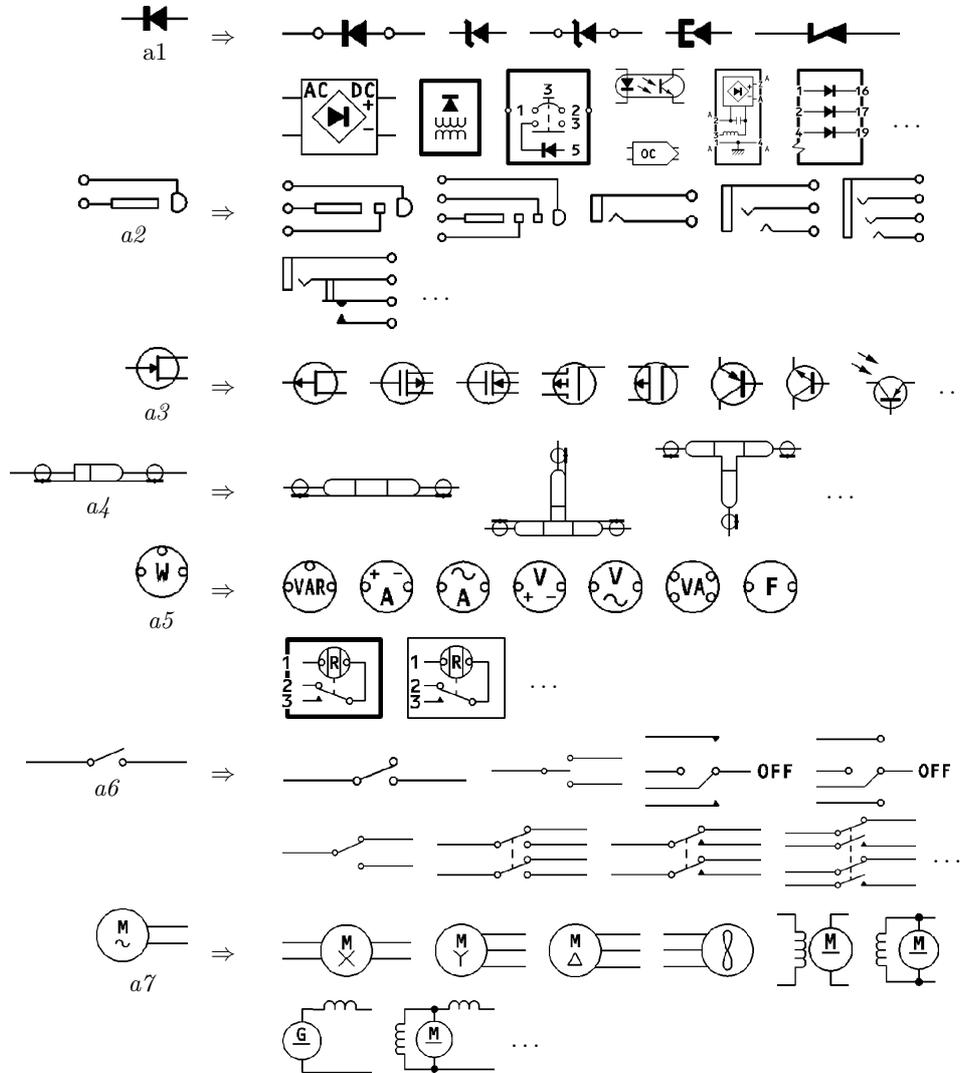


Figure 5: A sample of few electrical symbols. For every test symbol:  $a1$  to  $a7$ , a few relevant symbols are enlisted based on human evaluation. It consists of both linear as well as symbols in the composite form.

of retrieval for a given short list of size  $K$  is expressed as,

$$\eta_K = \begin{cases} n/N & \text{if } N \leq K \\ n/K & \text{otherwise,} \end{cases}$$

where  $n$  is the number of returned relevant symbols,  $N$  the total number of relevant symbols and  $K$  the number of ranked symbols requested. Note that  $\eta_K$  computes the traditional recall if  $N \leq K$  and computes precision otherwise. The main advantage of this is that the average retrieval efficiency curve is not biased even with different quantities of ground-truth for different queries, while it happens for precision measures when  $N < K$ .

**Matching Scope.** Because of the fact we have a fixed set of labelled vertices (*i.e.*, *vocabulary* types) in our symbol description, we are able to control the matching scope for every chosen query by using a parameter  $s$ . Using the notation introduced in Section 4.1, we define  $s$  as  $\delta(\tau_A^{V^q}, \tau_A^{V^d})$ . Depending on the value of  $s$  different matching strategies can be applied:

$s \geq 0$  : all candidates in the dataset are taken into consideration for matching.

$s \leq 1$  : matching is only done between candidates differing by at most one vertex (*i.e.*, one vertex can be absent or supplementary).

$s = 0$  : matching is done by candidates only having the exact same set of vertices (*i.e.*,  $V^q = V^d$ ).

Therefore, we have applied the three different matching strategies to evaluate the behaviour of different methods with scopes ranging from  $s \geq 0$  to  $s = 0$ . Our assumption is that candidates having same set of vertices as well as exact labels are similar either for their whole structure or part of it when in composite forms. This assumption has been experimentally validated.

### 5.2.2. Results

In this section we present a series of experiments establishing the performances of our approach. We address three specific issues:

1. What is the optimal set of parameters for our method?
2. How does our spatial relation model compare to other spatial relation models?
3. How well does our recognition model do with respect to state-of-the-art recognition models?

In all experiments, we have used *retrieval efficiency*, as described in Section 5.1. We compare the average retrieval efficiencies over the same 30 queries for all presented cases. These efficiency values have been computed for values of  $K = 1$  to 10.

**Resolution Parameter Determination.** Our method, besides depending on the choice of the vocabulary, uses one main parameter: the resolution at which the angular histogram is computed<sup>1</sup>. Its value represents the trade-off between the optimal choice of resolution – and thus precision of spatio-structural information capture – and time/space requirements. Fig. 6(a) shows the result of a series of experiments with  $\Theta$  varying over  $\{1^\circ, 3^\circ, 5^\circ, 9^\circ\}$ . For each of its values we have measured the retrieval efficiency on the same set of queries. Without surprise, the lower  $\Theta$ , the better the results, independently of the matching strategies used.

Based on these results, and given the relatively low gain of efficiency between  $3^\circ$  and  $1^\circ$ , we adopt the former for the rest of our experiments.

**Other Spatial Relation Models.** In order to compare our spatial relation model with others, we have adapted our ARG to function with those published in [Miyajima and Ralescu, 1994; Papadias and Theodoridis, 1997; Wang and Keller, 1999], and we have submitted them to the same testing protocol as described before. Fig. 6(b) shows their average retrieval efficiency. MBR outperforms all others in all situations. We shall further compare it to our method at the end of this section.

**Global Signal-based Descriptors.** In order to compare our method to other recognition methods, we have selected a set of major global signal-based shape descriptors described in Section 1.2 [Kim and Kim, 2000; Zhang and Lu, 2002; Belongie et al., 2002; Tabbone et al., 2006]. For GFD, we have tuned the parameters, and selected those values for radial and angular frequency that achieved the best recognition performance on our dataset: radial frequency 6 and angular 15. For *Shape Context*, only 70 sample points have been selected because of the presence of smaller size images in our database. In case of ZM, we have used 36 *zernike* functions of order less than or equal to 7. Also, we have taken radon image intensity over the projecting angle  $[0, \pi[$  by default, for *R-signature*. Unlike the methods based on spatial relations, we cannot establish different matching scopes, based on  $s$  as presented in Section 5.2.1 and used previously.

Again the same queries are presented and average retrieval efficiency is shown in Fig. 6(c). GFD seems to be performing the best among all tested

---

<sup>1</sup>The matching scope  $s$ , as introduced in Section 5.2.1 should not really be considered as a parameter, but as a measure of our method's robustness.

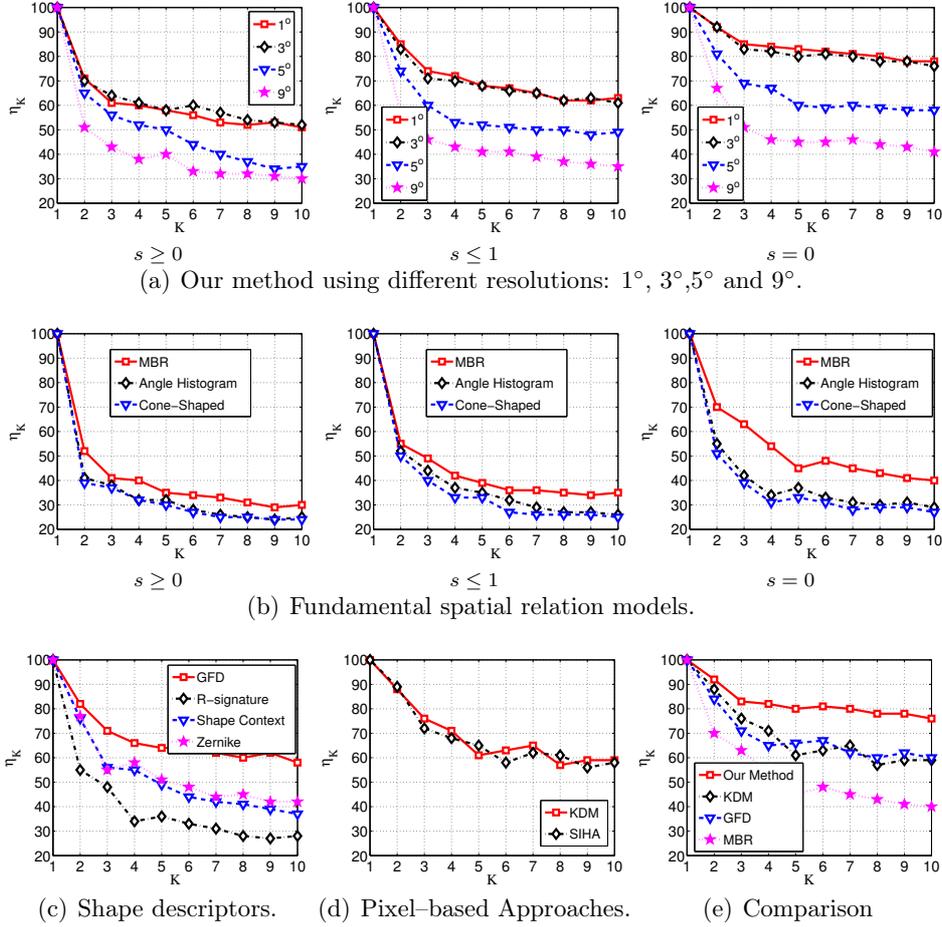


Figure 6: Average retrieval efficiency over requested list – 1 to 10: (a) Our method, (b) Fundamental spatial relation models, (c) Global signal-based descriptors, (d) Pixel-based approaches and (e) Comparison.

global signal-based descriptors in our setup.

**Pixel-based Approaches.** We have also compared our method with two pixel-based approaches specially designed for symbol recognition: Statistical Integration of Histogram Array (SIHA) [Yang, 2005] and Kernel Density Matching (KDM) [Zhang et al., 2006]. In SIHA, two different length-ratio and angle-ratio histograms are taken from every two pixels in reference to a

third pixel from the skeleton image. In KDM, skeleton symbols represent as 2D kernel densities and their similarity is measured by the *Kullback–Leibler divergence*. In Fig. 6(d), results are shown for both. In this test, we observe almost similar behaviour from the two. However, KDM performs slightly better, especially when also taking time complexity into account.

Overall, compared to our method, basic spatial relation models and global signal-based descriptors as well as recent pixel-based approaches have been lagging behind. Fig. 6(e) provides a comparison by taking the best of both classes: MBR from the spatial relation models, GFD from the global signal-based descriptors and KDM from the pixel-based approaches. Our method outperforms both with a significant difference in retrieval efficiency.

### 5.3. Discussions

In this section, the performance of the methods in response to the experimental results are analysed. Performance not only refers to retrieval efficiency but also to time complexity. In parallel, we discuss matching scope and its effect in ranking retrieved symbols.

To visually compare the results of our method with the best of breed solutions reported in Fig. 6(b), 6(c) and 6(d), we show a selection of queries in Fig. 7. They demonstrate the use of isolated as well as composed symbols as query. The first symbol on the top is always the chosen query and symbols are ranked from top to bottom (1 to 10) based on decreasing order of similarity. For query  $Q1$ , GFD and KDM come close to our method while MBR presents a notable difference. In case of query  $Q2$ , our method outperforms all others significantly. A similar situation happens for  $Q3$ .

Our method exploits spatio-structural description of the visual parts. The choice of the vocabulary types (*i.e.* collection of particular visual parts) is of course an important factor to its success. However, symbols like ,  *etc.* are retrieved for the query  due to the presence of *thick* patterns. This shows that our relational signatures do not provide or use any shape information. Therefore, symbols having any *thick* pattern like, , , , , , ,  *etc.* are selected for ranking. However, spatial organisation of *thick* patterns with respect to other primitives helps to rank the best one first.

Running time has been measured in all experiments. An average running time (in sec.) for all methods is given below.

1. Our Method	04
2. Basic Relation Models	
2.1 Cone-Shaped [Miyajima and Ralescu, 1994]	≤ 01
2.2 MBR [Papadias and Theodoridis, 1997]	02
2.3 Angle Histogram [Wang and Keller, 1999]	44
3. Global Shape Descriptors	
3.1 $\mathcal{R}$ -signature [Tabbone et al., 2006]	03
3.2 Zernike Moments [Kim and Kim, 2000]	13
3.3 GFD [Zhang and Lu, 2002]	09
3.4 Shape Context [Belongie et al., 2002]	32
4. Pixel-based Approaches	
4.1 SIHA [Yang, 2005]	64
4.2 2D KDM [Zhang et al., 2006]	24

We used MATLAB 7.8.0 in Linux platform.

Our method has benefited from the way we describe the matching strategy (*cf.* Section 5.2.1). Symbol matching between the candidates which share the same sets of vertices with exact labels (*i.e.*,  $s = 0$ ), is found to be the best among all. It sufficiently reduces time of matching to symbols which are obviously irrelevant. Similarly, this happens in those tests using basic spatial relations models. But for global signal-based descriptors as well as pixel-based approaches, running time increases with number of symbols in the dataset since matching scope does not exist.

## 6. Conclusions and Further Work

In this paper, we have presented a method to describe symbols using a specific Attributed Relational Graph via the use of spatial relations between the visual elementary parts. Each vertex represents all visual parts of a particular vocabulary type within the symbol. The edges represent the spatial relations between them. The proposed method is simple and flexible, and has the ability to express spatial relations between any number of visual parts. We have validated that such a description can be used for symbol recognition. Our method has proven to significantly outperform state-of-the-art basic spatial relation models, global signal-based descriptors and pixel-based approaches for symbol recognition.

	Q1				Q2				Q3			
	MBR	GFD	KDM	Our Method	MBR	GFD	KDM	Our Method	MBR	GFD	KDM	Our Method
1.												
2.												
3.												
4.												
5.												
6.												
7.												
8.												
9.												
10.												

Figure 7: Visual illustration of symbol retrieval and ranking for a few queries, showing the true () and false () retrieval.

Further work comprises the study of the influence of the weighting parameters in the matching score. Furthermore we are currently studying clustering techniques to enhance the discriminative power (and this enhance retrieval performance) of the *thick* component patterns. In addition, we are going to relate our work to [Bai et al., 2010] in order to see how both approaches can be combined to enhance overall performance.

### References

- Ah-Soon, C., Tombre, K., 2001. Architectural symbol recognition using a network of constraints. *Pattern Recognition Letters* 22 (2), 231–248.
- Bai, X., Yang, X., Latecki, L., Liu, W., Tu, Z., 2010. Learning context-sensitive shape similarity by graph transduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32 (5), 861–874.
- Bar, M., Ullman, S., 1993. Spatial context in recognition. *Perception* 25, 324–352.
- Belongie, S., Malik, J., Puzicha, J., 2002. Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 24 (4), 509–522.
- Biederman, I., 1972. Perceiving real-world scenes. *Science* 177 (43), 77–80.
- Bloch, I., 1999. Fuzzy relative position between objects in image processing: New definition and properties based on a morphological approach. *Uncertainty Fuzziness and Knowledge-Based Systems* 7 (2), 99–133.
- Bunke, H., Messmer, B. T., 1995. Efficient attributed graph matching and its application to image analysis. In: Braccini, C., Floriani, L. D., Vernazza, G. (Eds.), *Proceedings of International Conference on Image Analysis and Processing*. Vol. 974 of *Lecture Notes in Computer Science*. Springer-Verlag, pp. 45–55.
- Chaudhuri, B. B., Garain, U., 2000. An approach for recognition and interpretation of mathematical expressions in printed document. *Pattern Analysis and Applications* 3 (2), 120–131.

- Conte, D., Foggia, P., Sansone, C., Vento, M., 2004. Thirty years of graph matching in pattern recognition. *International Journal of Pattern Recognition and Artificial Intelligence* 18 (3), 265–298.
- Cordella, L. P., Vento, M., 2000. Symbol recognition in documents: a collection of techniques? *International Journal on Document Analysis and Recognition* 3 (2), 73–88.
- Egenhofer, M., Franzosa, R., 1991. Point-set Topological Spatial Relations. *International Journal of Geographical Information Systems* 5 (2), 161–174.
- Egenhofer, M., Herring, J. R., 1991. Categorizing Binary Topological Relations Between Regions, Lines, and Points in Geographic Databases. University of Maine, Research Report.
- E.Jungert, 1993. Qualitative spatial reasoning for determination of object relations using symbolic interval projections. In: *IEEE Symposium on Visual Languages*. pp. 24–27.
- Freeman, J., 1975. The modelling of spatial relations. *Computer Graphics and Image Processing* 4, 156–171.
- Kankanhalli, M. S., Mehtre, B. M., Wu, J. K., 1995. Cluster-based color matching for image retrieval. *Pattern Recognition* 29, 701–708.
- K.C., S., Lamiroy, B., Ropers, J.-P., 2009. Inductive logic programming for symbol recognition. In: *Proceedings of International Conference on Document Analysis and Recognition*. pp. 1330–1334.
- K.C., S., Wendling, L., Lamiroy, B., 2010. Unified pairwise spatial relations: An application to graphical symbol retrieval. In: Ogier, J.-M., Liu, W., Lladós, J. (Eds.), *Graphics Recognition. Achievements, Challenges, and Evolution*. Vol. 6020 of *Lecture Notes in Computer Science*. Springer-Verlag, pp. 163–174.
- Kim, W.-Y., Kim, Y.-S., 2000. A region-based shape descriptor using zernike moments. *Signal Processing: Image Communication* 16 (1-2), 95 – 102.
- Lamiroy, B., Guebbas, Y., 2010. Robust and precise circular arc detection. In: Ogier, J.-M., Liu, W., Lladós, J. (Eds.), *Graphics Recognition. Achievements, Challenges, and Evolution, 8th International Workshop, GREC*

- 2009, La Rochelle, France, July 22-23, 2009. Selected Papers. Vol. 6020 of Lecture Notes in Computer Science. Springer-Verlag, pp. 49–60.
- Lee, S.-H., Hsu, F.-J., 1992. Spatial Reasoning and Similarity Retrieval of Images Using 2D C-string Knowledge Representation. *Pattern Recognition* 25 (3), 305–318.
- Lladós, J., Martí, E., Villanueva, J. J., 2001. Symbol recognition by error-tolerant subgraph matching between region adjacency graphs. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (10), 1137–1143.
- Lladós, J., Valveny, E., Sánchez, G., Martí, E., 2002. Symbol Recognition: Current Advances and Perspectives. In: Blostein, D., Kwon, Y.-B. (Eds.), *GREC – Algorithms and Applications*. Vol. 2390 of Lecture Notes in Computer Science. Springer-Verlag, pp. 104–127.
- Matsakis, P., Wendling, L., 1999. A New Way to Represent the Relative Position Between Areal Objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 21 (7), 634–643.
- Miyajima, K., Ralescu, A., 1994. Spatial Organization in 2D Segmented Images: Representation and Recognition of Primitive Spatial Relations. *Fuzzy Sets and Systems* 2 (65), 225–236.
- Okazaki, A., Tsunekawa, S., Kondo, T., Mori, K., Kawamoto, E., 1988. An automatic circuit diagram reader with loop-structure-based symbol recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 10 (3), 331–341.
- Papadias, D., Sellis, T., Theodoridis, Y., Egenhofer, M. J., 1995. Topological relations in the world of minimum bounding rectangles: a study with r-trees. *SIGMOD Record* 24 (2), 92–103.
- Papadias, D., Theodoridis, Y., 1997. Spatial relations, minimum bounding rectangles, and spatial data structures. *International Journal of Geographical Information Science* 11 (2), 111–138.
- Pham, T. V., Smeulders, A. W. M., 2006. Learning spatial relations in object recognition. *Pattern Recognition Letters* 27 (14), 1673–1684.

- Rebelo, A., Capela, G., Cardoso, J. S., 2010. Optical recognition of music symbols: A comparative study. *International Journal on Document Analysis and Recognition* 13 (1), 19–31.
- Rendek, J., Masini, G., Dosch, P., Tombre, K., 2004. The search for genericity in graphics recognition applications: Design issues of the qgar software system. In: Marinai, S., Dengel, A. (Eds.), *Proceedings of International Workshop on Document Analysis Systems*. Vol. 3163 of *Lecture Notes in Computer Science*. Springer-Verlag, pp. 366–377.
- Renz, J., Nebel, B., 1998. Spatial reasoning with topological information. In: *Spatial Cognition, An Interdisciplinary Approach to Representing and Processing Spatial Knowledge*. Springer-Verlag, pp. 351–372.
- Samet, H., Soffer, A., 1996. Marco: Map retrieval by content. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 783–798.
- Tabbone, S., Wendling, L., Salmon, J.-P., 2006. A new shape descriptor defined on the radon transform. *Computer Vision and Image Understanding* 102 (1), 42–51.
- Tombre, K., Lamiroy, B., 2008. Pattern recognition methods for querying and browsing technical documentation. In: *Proceedings of Iberoamerican Congress on Pattern Recognition*. Springer-Verlag, pp. 504–518.
- Valveny, E., Martí, E., 2003. A model for image generation and symbol recognition through the deformation of lineal shapes. *Pattern Recognition Letters* 24 (15), 2857–2867.
- Wang, X., Keller, J., 1999. Human-Based Spatial Relationship Generalization Through Neural/Fuzzy Approaches. *Fuzzy Sets and Systems* 101, 5–20.
- Xiaogang, X., Zhengxing, S., Binbin, P., Xiangyu, J., Wenyin, L., 2004. An online composite graphics recognition approach based on matching of spatial relation graphs. *International Journal on Document Analysis and Recognition* 7 (1), 44–55.
- Yang, R., Lu, T., Cai, S., 2007. A dynamic-rule-based framework of engineering drawing recognition and interpretation system. In: *ICIC: International Conference on Advanced intelligent computing theories and applications*. Springer-Verlag, pp. 1006–1017.

- Yang, S., 2005. Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 27 (2), 278–281.
- Yuen, P. C., Feng, G.-C., Tang, Y. Y., 1998. Printed chinese character similarity measurement using ring projection and distance transform. *International Journal of Pattern Recognition and Artificial Intelligence* 12 (2), 209–221.
- Zhang, D., Lu, G., 2002. Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication* 17, 825–848.
- Zhang, D., Lu, G., 2004. Review of shape representation and description techniques. *Pattern Recognition* 37 (1), 1–19.
- Zhang, W., Wenyin, L., Zhang, K., 2006. Symbol recognition with kernel density matching. *Pattern Recognition* 28 (12), 2020–2024.

# How Carefully Designed Open Resource Sharing Can Help and Expand Document Analysis Research

Bart Lamiroy<sup>a</sup>, Daniel Lopresti<sup>b</sup>, Hank Korth<sup>b</sup> and Jeff Hefflin<sup>b</sup>

<sup>a</sup>Nancy Université – LORIA, Campus Scientifique, BP 239, 54506 Vandoeuvre Cedex, France  
Bart.Lamiroy@loria.fr

<sup>b</sup>Department of Computer Science and Engineering, Lehigh University, Bethlehem, PA 18015  
{lopresti,korth,hefflin}@cse.lehigh.edu

## ABSTRACT

Making datasets available for peer reviewing of published document analysis methods or distributing large commonly used document corpora for benchmarking are extremely useful and sound practices and initiatives. This paper shows that they cover only a very tiny segment of the uses shared and commonly available research data may have. We develop a completely new paradigm for sharing and accessing common data sets, benchmarks and other tools that is based on a very open and free community based contribution model. The model is operational and has been implemented so that it can be tested on a broad scale. The new interactions that will arise from its use may spark innovative ways of conducting document analysis research on the one hand, but create very challenging interactions with other research domains as well.

## 1. INTRODUCTION

It is commonly accepted that the sharing of reference benchmark material is an essential practice in science domains where reproducible experiments play an important role in the global peer review process. Many areas and communities have structured themselves around such kind of resources and have greatly benefited from doing so.

In document analysis, there have been numerous attempts and initiatives<sup>1-3</sup> to produce common datasets for many of the problems that are addressed by the community. Some of them have had great impact, others less. What is certain is that all of them have had their hour of glory, and have then more or less quickly declined. Not that their intrinsic quality changed, but very often datasets have become progressively obsoleted by technology advances, new research focuses, or lack of support by their creators.

There is, indeed, a very high cost attached to the creation, maintenance and diffusion of useful research tools, datasets and benchmarks for a given community. However, they are snapshots of research topics, problems and state-of-the art at the time of their creation, unless an equivalent amount of effort goes into keeping them in line with the continuous evolution of knowledge.

Another aspect, one that is particularly related to document analysis, is that the research in this domain is very application and problem driven. Be it invoice routing, building the semantic desktop, digital libraries, global intelligence, or document authentication, to name a few, they all tend to generate very specific datasets, and produce very focused software solutions, often integrating a complete pipeline of cascading methods and algorithms. This most certainly does not affect the intrinsic quality of the underlying research, but it does tend to generate isolated clusters of extremely focused problem definitions and experimental requirements. This often even increases the cost of producing and maintaining shared available benchmarking resources, since the burden has to be supported by a very small community and cannot easily be distributed to other potential contributors. This also makes it difficult to cross borders and agree on what kinds of tools, formats *etc.* are actually the most useful ...

In this paper we address the above mentioned issues by proposing a model that tries to minimize the cost and burden to deliver and maintain datasets and benchmarking tools, keeping the model open to any kind of new contribution and even making it flexible and open to evolution with the needs of changing technology and knowledge. Our work is structured as follows: first we address the fundamental questions that arise and need

to be addressed when sharing common datasets, benchmarks and methods. We shall study them point by point thus creating the requirements a perfect platform should meet in order to appeal to and be used by the research community. We then present the our implementation choices, and give concrete and extensive examples of how it is and can be used for document analysis problems.

## 2. SCENARIO, SCRIPT AND SCREENPLAY

In order to develop our ideas, let's consider a scenario where we can have access of a well identified resource that can provide us with any kind of data related to a specific document analysis problem we would like to solve. For instance: *"Provide me with 1,000 random documents in 300dpi bitonal .tif format, predominately containing printed material and of which a reasonable number of pages contain at least one handwritten annotation."*

### 2.1 Plot and Triggers

What we are going to describe, in essence, is the availability of a well identified, commonly available, resource (for convenience sake, let's assume this resource is centralized, we shall see further that this assumption can be greatly reduced) that offers the following services: storage and retrieval of document analysis data, complex querying of the document analysis data, collective, yet personalized markup, representation, evaluation, organization and projection of data, archival knowledge of uses and transformations of data, certified interaction with data.

The first item might, at a first glance, not look quite different from what already is available to some extent: document analysis datasets.<sup>1-3</sup> In the light of the second item, it proves to be completely different, however. Rather than to offer *monolithic* chunks of data and meta-data or *interpretations*, the envisioned resource treats data on a far finer grained level. This level of detail is illustrated in the scenario by the kinds of queries the system would be capable of answering: *"1,000 random documents in 300dpi bitonal .tif format predominately containing printed material and of which a reasonable number of pages contain at least one handwritten annotation."*

Furthermore, data representation (not only the document images, but extracted meta data, annotations and *interpretations* ...) needs not be conforming to a predefined format, but can be polymorphic, originating from both specific individual annotation initiatives as from collective agreed upon contributions or even algorithms. Not only is it stored, but it can also be retrieved and reprojected into any format. All these data are not necessarily human-contributed but can actually be the result of complete document analysis pipelines<sup>4</sup> and algorithms. As a result, the resource can hold apparently contradictory *interpretations* of identical documents, when these *interpretations* stem from different transformation and analysis processes. This means that there cannot be an absolute and unique notion of *ground-truth*, since the resource is hosting the interpretations for multiple contexts. This corresponds to recent debates on the existence (or rather lack thereof) of ground-truth or universal interpretations.<sup>5-8</sup>

The last item in our list describes a service where, certification set apart, not only data is provided as a resource, but also interactions with this data are available, as through documented benchmarks, state-of-the-art reference algorithms, *etc.* Not only is this interaction available, but its usage is monitored and registered such that the produced results are also stored in the resource, together with its full origin and provenance.

The next sections will describe in detail what specific roles and actors can be defined to realize the scenario we have depicted above.

### 2.2 Featured Roles

Before considering the more scientific and technical underlying key points of our scenario, here are the essential roles it has to fulfill. Each of these points will then be further developed in section 2.3.

**Formats and Representation** are to be examined before addressing the issue of storage and access to the data. In order for our scenario to have a chance of realization, it seems to us that it cannot, ever, take any assumption on what data representation or formats should be used. Past experiments have shown too often that an approach consisting in inciting or “coercing” a community into using a single set of representation conventions does not work or contributes to locking it into a limited sub-part of uses of the generated data, hampering creative and extended uses beyond their initial purpose. This is clearly contradictory to our standpoint with respect to *ground-truth* and our preferring of the term *interpretations*.

This clearly advocates for an as open as possible way of representing any kind of data. On the other hand, *à trop êtreindre, mal embrasse*, and abandoning any kind of imposed structure may prove rendering the querying part of our plot hard to realize. This will not be an issue if there is a widespread use and sharing of data, since it will eventually lead to the emergence of *de facto* formats, especially since we are targeting a rather well defined community. The less constraining representations formats are, the higher the chance there will indeed be a widespread use.

**Storage and Access** is one of the key elements to making our resource and associated services available and accessible. It seems, however, this is “merely” a question of material resources, rather a conceptual problem (this is an understatement, but availability, redundancy and consistency discussions are beyond the scope of this paper); especially when remaining under the assumption that the storage remains centralized (condition, as already mentioned, that can be alleviated). The quantity of data is potentially huge, and needs to be constantly available. The supporting infrastructure needs to be sufficiently dimensioned for handling multiple concurrent accesses and provide high enough bandwidth. The platform also needs to be able to host the versatile data-model implementing the previous point, allow for the rich semantic querying described in the next, and remain capable of hosting and executing the algorithms described after as well as providing the interfaces for commenting, evaluating, rating or re-projecting anything already stored.

**Querying and Retrieval** needs to allow an as broad as possible spectrum of interactions. Since we cannot really rely on a canonical representation of all stored data – mainly because there probably isn’t any universal representation of all possible annotations and interpretations of document analysis results and document contents – querying and retrieval cannot just be keyword based. We need a much more semantic way of interacting with our data ... raising “data” to the level of “information”, essentially. This task is challenging. The illustrative query mentioned before: “*1,000 random documents in 300dpi bitonal .tif format predominately containing printed material and of which a reasonable number of pages contain at least one handwritten annotation.*” raises quite some exciting challenges. While “bitonal .tif” might be the less controversial part of the query, it still may be subject to discussion: the .tif format may support compression, while given algorithms require their .tif input to be uncompressed; what exactly does bitonal mean ? and if we allow on-the-fly conversions of image formats into .tif, would we allow also conversions of graylevel or color into bitonal ? (but then what binarization algorithm should we use ?) *etc.* The further downstream use of the retrieved data may require the handwritten annotations markup and localization data need to be *compatible* with the algorithm is is going to be tested on. This implies that this *compatibility* be somewhere formally expressed, and that, at the same time, the input type of the algorithm be equally formalized.

This means that querying and retrieval most definitely need to operate on a semantic level. The corollary of that is that these semantics need to be represented somewhere. This is especially the case here, where we want to have automated interaction between data and algorithms. In order to be applied algorithms to any available document in our repository, the platform needs to have the capability to relate these algorithms to data, and therefore requires input formats and output semantics to be formalized and stored in some way, thus joining the point made previously.

**Provenance and Certification** are two supplemental benefits we get from correctly realizing the previous roles, and especially from provenance mentioned in section 2.1. They strongly relate to (and depend on) recording of the origin of all information that is stored, and can offer an interesting level of interaction and service. One

of the potential uses might be certification of experiments and algorithm results. These are, in fact, a specific instances of what global provenance tracking can offer. By assuring that all information is uniquely referenced, totally persistent (cannot be removed) and traceable, one can solve many data-algorithm-result interdependence problems. For instance, widely used benchmarks suddenly prove biased or flawed: provenance makes it possible to mark or invalidate depending data. Copyright claims may result in having datasets to be retracted or recomposed, but how about generated results? Provenance information provides tools to keep the level of availability as close to optimal as possible, by analyzing which data is impacted by the claim.

## 2.3 The Cast

Describing what a perfect and hypothetical environment should provide in order to assure a new and stimulating way of doing document analysis research is rather easy. Actually starting to implement it and making it available to the community is more of a challenge. The DAE server implementing a big part of what is described in this paper exists, and is continuously expanding its features. This section describes the more technical and scientific challenges that have been met, or need to be met, in order to completely transform above descriptions into reality.

### 2.3.1 The DAE Platform

Our platform is available at <http://dae.cse.lehigh.edu>. It runs on a 48TB storage, 12 core 32G RAM server,

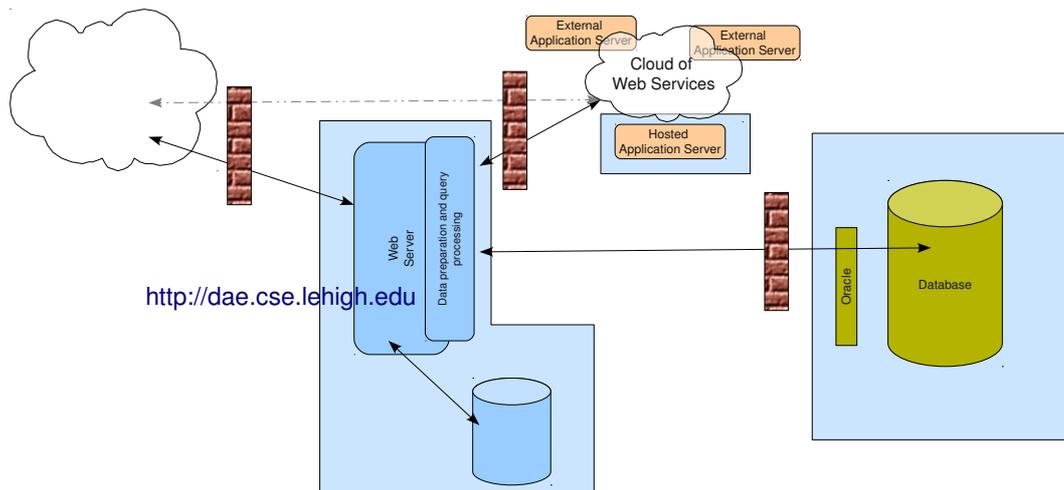


Figure 1. The DAE platform architecture

and as such, is to be more considered as a seriously endowed proof of concept\*, rather than as the ultimate infrastructure to support all claims and goals expressed in this paper.

Besides the technical architecture of the server, as depicted in Fig. 1, the DAE Platform is first of all the implementation of a *data model*.<sup>9</sup> The data model is based on the following claims:

- all data is typed; users can define new types;
- data can be attached to specific parts of a document image (but does not need to),
- both data and algorithms are modeled; algorithms transform data from one type into data of another type;
- full provenance of data history is recorded;

\*Ideally, the platform should evolve into a distributed community-managed resource, rather than remaining a centralized platform.

It has been fully implemented on an Oracle 11.2 back-end database management system (right – in green in Fig. 1). It is accessed by a web front-end that provides a Web 2.0-like interface (left – in blue in Fig. 1) and encapsulates SQL queries to the back-end. It also relies on independent “application” servers that are used for executing registered algorithms on the data (middle – in orange in Fig. 1). The data model can be downloaded from <http://dae.cse.lehigh.edu/Design/ER.pdf>. All source code is GPL licensed and freely available from <http://sourceforge.net/projects/daeplatform/> and are open to contributions and extensions by the community.

This platform fulfills most of the roles described in section 2.2:

**Formats and Representations** are transparently handled by the system, since the user can define any format, naming or association convention within our system. Data can be associated with image regions, image regions can be of any shape and format, there is no restriction on uniqueness or redundancy, so multiple interpretations are not an issue. Furthermore, in order to avoid generating a bulk mass of incomprehensible annotations and data-structures, data can be conveniently grouped together in sets. These sets can in their turn be named and annotated as well. Data items do not need to belong exclusively to a single set, so new sets can be created by recombination or combination of existing data sets.

The core elements of our solution consist of considering most of the stored information as `data_items` which in their turns instantiate as more specific subtypes. The principal ones are depicted in the partial data model view of Fig 2.

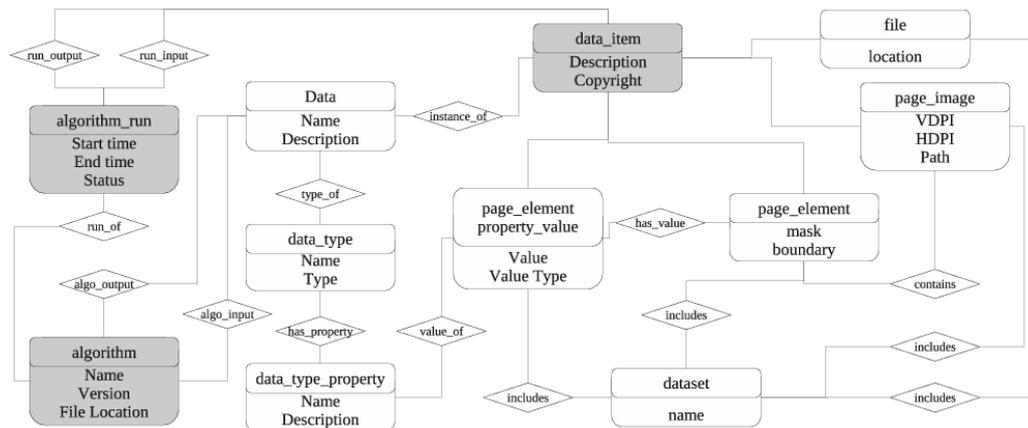


Figure 2. A partial view of the DAE platform data model

In our approach, `page_images` are considered as single, stand-alone, document analysis objects in a given context. They may be related to other images, known to represent the same physical object, but captured under other conditions (not represented in Fig. 2).

`page_images` can contain any number of `page_elements`. These are areas of pixels contained within the image. Our model is currently capable of representing these areas as bounding boxes or pixel masks, but can be extended in a very straightforward way to handle polygons or other geometric shapes. It is noteworthy to mention that there is no real need to extend beyond those types, however. One might be tempted (especially when considering graphical documents) to introduce complex vectorial representations, for instance, like ellipses, lines *etc.*. We argue that those belong to the domain of *interpretations* (and therefore `page_element_property_values`) rather than `page_elements`. `page_elements` are “physical” sub-parts of `page_images` and therefore just pixels.

`page_element_property_values`, on the other hand, can be any interpretation of a `page_element`: layout labels, OCR results, ...

Using this approach we have been able to incorporate a great amount of widespread and less widespread document analysis data formats into our platform. We are hosting and compatible with the native formats of the UNLV dataset,<sup>1</sup> Tobacco800<sup>2</sup> using the GEDI<sup>0</sup> format, and many others. We provide XSLT conversion tools to as well as a full XML Schema of our upload format.

The results are fully browsable and interactive datasets, as shown in Fig. 3.

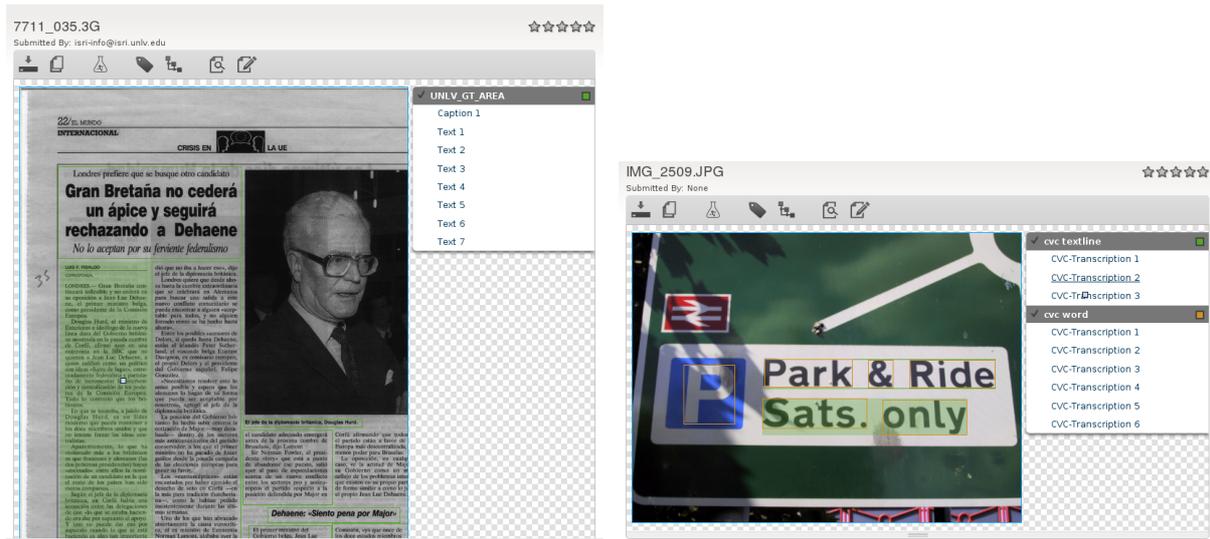


Figure 3. Example of UNLV (left) and custom (courtesy of CVC) imported `page_element_property_values` being consulted through our platform.

**Storage and Access** are covered by the system by the fact that the storage (48TB raw storage, equivalent to an effective 28TB once formatted and secured) and computing capacities are largely sufficient for making a actually operational proof of concept that scales to a few hundreds of concurrent connections, and capable of storing most currently available datasets. Furthermore its architecture makes it extremely easy to scale to higher demands, as both the storage and the computing infrastructures are conceived as physically separate entities. The current version already spawns some of its computing onto a separate high performance computing cluster for specific algorithms and offers access to the others as Web-services. This guarantees that the part of the platform that manages the execution of algorithm is totally scalable. Likewise, the storage of the data may also be distributed over multiple sites. Seamlessly integrating the meta data into a fully distributed environment is likely to require further investigation and research efforts, however.

**Querying and Retrieval** is the great gain that the DAE platform offers. Because of its underlying data model and architecture, everything is “SQL-queryable”. The standard datasets that can be downloaded from the platform are no longer monolithic .zip files, but actually potentially complex queries that generate these datasets on the fly. Because of the extreme degree of flexibility in annotating and supplementing existing data with meta-data, the potential uses are endless.

The first effect is that every single data item stored on our platform has a unique identifier, be it a `page_image`, a `page_element` or `page_element_property_value` (*i.e.* an interpretation), they all are referenced through a unified URL like this:

<http://dae.cse.lehigh.edu/DAE/?q=browse/dataitem/download/99262>.

This means our platform can be considered as an on-line storage engine and that algorithms can directly access the stored data for their use, rather than requiring it to be downloaded locally before use.

The second effect is that one can retrieve images on the one hand or other any information (like segmentation data, OCR results, recognition labels ...) on the other hand, or any combination of these. While this clearly allows for a classical approach to retrieving the data (*e.g.* like filtering out parts of a previously pre-packaged dataset) , it also allows for more creative uses and crossing information coming from various and different datasets, possibly even conceived for other application contexts than the one under consideration.

**Examples** of what our system offers are shown below.

- To find out what data sets that are declared on the platform (data sets may be recursively contain other datasets, so we restrict ourselves to top-level ones):

```
select distinct NAME, ID from DATASET where ID not in
(select ASSOCIATING_DATASET_ID from ASSOCIATE_DATASET);
```

- To retrieve the filenames of all document images in a dataset (here the dataset defined by its id 633):

```
select PATH from INCLUDES_PAGE_IMAGE, PAGE_IMAGE where
DATASET_ID = 633 and PAGE_IMAGE_ID = ID;
```

It is noteworthy to mention that the PATH attribute retrieved is actually the URL where the image can be downloaded from. This, in its turn means that the actual image can be stored virtually anywhere on the Internet thus providing the opportunity to create *virtual* datasets, consisting of collections of remote and distributed storage elements<sup>†</sup>.

- To retrieve all data produced by a particular algorithm (here algorithm 66):

```
select DATA_ITEM_ID from ALGORITHM, ALGORITHM_RUN_OUTPUT, ALGORITHM_RUN_OF where
ALGORITHM_ID = 66 and ALGORITHM_ID = ID and
ALGORITHM_RUN_OUTPUT.ALGORITHM_RUN_ID = ALGORITHM_RUN_OF.ALGORITHM_RUN_ID;
```

- To discover all user defined data types and their descriptions:

```
select * from DATATYPE_PROPERTY;
```

Although the previously mentioned query “*1,000 random documents in 300dpi bitonal .tif format predominantly containing printed material and of which a reasonable number of pages contain at least one handwritten annotation.*” requires some additional work in order to implement the randomization, and needs to handle some more subtleties of the underlying semantics (*cf.* section 2.3.2) an SQL query implementing an approximation of it is perfectly within the scope of our platform. There is a catch, however, as will immediately become clear from the following query. Without loss of generality, let us consider only one part of the previous query: *pages containing handwriting.*

Since our data model is quite straightforward, this means that we are looking for `page_images` that have some `page_element` that has somehow a `page_element_property_value` that is *handwriting*. The model perfectly handles this, and the corresponding `page_element_property_value` can be human provided or come from specialized detection algorithms (*e.g.* page segmentation<sup>11</sup>). The aforementioned catch lies in the fact that there is no unique representation – and therefore no unique query – that yields the answer to *handwriting*.

**Note:** from this point on, we assume the reader is sufficiently convinced of the validity of the data model, and that there is no need to actually give all SQL queries *in extenso*. We shall therefore only provide shorter, and more readable pseudo-queries, that focus on the core of the argument.

---

<sup>†</sup>Although this opens a broad new scope of problems related to archiving and guaranteeing perennity of the provided link, we are conveniently considering, for argument sake, that the platform can guarantee that the provided PATH is always valid, for instance by only serving data stored in its local storage.

Let's consider the two following pseudo-queries:

1. The following query returns pages having a property value labeled *handwriting*.

```
select PAGE_IMAGE such that PAGE_ELEMENT_PROPERTY_VALUE.VALUE = 'handwriting';
```

2. Similarly, the following query returns pages of which the property is of a type labeled *handwriting*.

```
select PAGE_IMAGE such that the PAGE_ELEMENT_PROPERTY_VALUE's
DATA_PROPERTY_TYPE = 'handwriting';
```

Indeed, as shown by queries like 1, some annotations or interpretations can be classification labels (*e.g. handwriting, table, logo ...*). Others, like 2 can actually be related to, for instance, handwriting recognition, and may yield the transcription of a `page_element` that is of type *handwriting*. As a matter of fact, queries like 1 may actually return unwanted results since they don't really take into account the type of the returned value. They might be OCR transcriptions of a text containing the printed word *handwriting*.

These differences may appear unnecessarily subtle for the average reader but they address a fundamental point related to the sharing of image interpretations. We shall discuss them in further detail in section 2.3.2.

**Interaction** with the data can be considered on multiple levels.

1. The first level is integrated in the data model, which represents algorithms, their inputs and outputs. Our platform goes further by actually providing access to binaries that can be executed on the stored data, thus producing new meta data and *interpretations* and expanding the range of available information. Queries like finding all OCR results produced by a specified algorithm, can either be used as an *interpretation* of a document, but can equally serve as a benchmarking element for comparison with competing OCR algorithms. Because of the fact that all are hosted in the same context, it becomes possible to "certify" (or *cite*, since they become unique references) evaluation or benchmarking runs.
2. The second level is more human-centered and concerns the interface that is provided to access the data. Besides the raw SQL access described previously, the platform offers a complete interactive environment that allows users to browse through stored document collections, upload data, run algorithms, tag, comment and rate data, *etc.* These functions are illustrated in Fig. 4.

### 2.3.2 Interpretations and Semantics

**Handling multiple interpretations** As mentioned before the design of our platform makes it possible to have multiple interpretations of the same set of data. This is a necessity due to multiple contexts in which the same data can be used, and the intrinsic difficulties to define absolute interpretations anyway.<sup>5-8</sup> The major side-effect is that semantically equivalent interpretations may be stored in various ways in our platform.

Let's return to the example developed in 2.3.1, where we considered the following two following pseudo-queries:

1. `select PAGE_IMAGE such that PAGE_ELEMENT_PROPERTY_VALUE.VALUE = 'handwriting';`

2. `select PAGE_IMAGE such that the PAGE_ELEMENT_PROPERTY_VALUE's
DATA_PROPERTY_TYPE = 'handwriting';`

Both are defining areas in `page_images` that contain the string *handwriting*. As already mentioned earlier, this does not really guarantee that the semantics of these areas actually cover handwritten visual information, since they might as well be the transcription of a printed text spelling the word *handwriting*.

Furthermore, the contributors of interpretations and labeling might have provided other labels, such as *manuscript* or *hand written* or *annotation*, *etc.* This is the cost we pay for a deliberate design choice. One might argue that using a thesaurus of authorized labels for annotation would solve the problem, but this is not necessarily true. This is related to the fundamental non-existence of absolute interpretations<sup>12</sup> and their relation with application contexts. One can easily conceive application contexts where *handwriting* is referring to totally

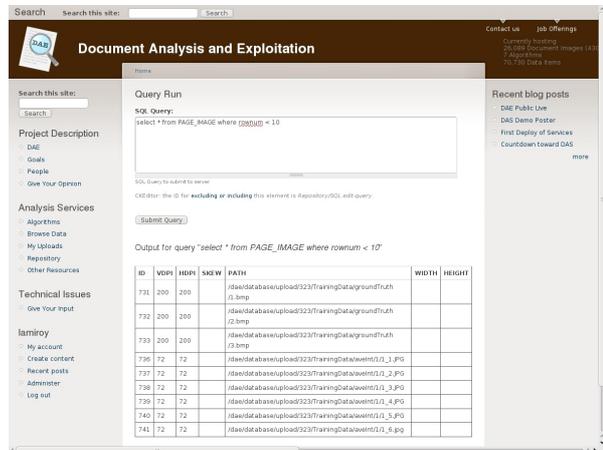
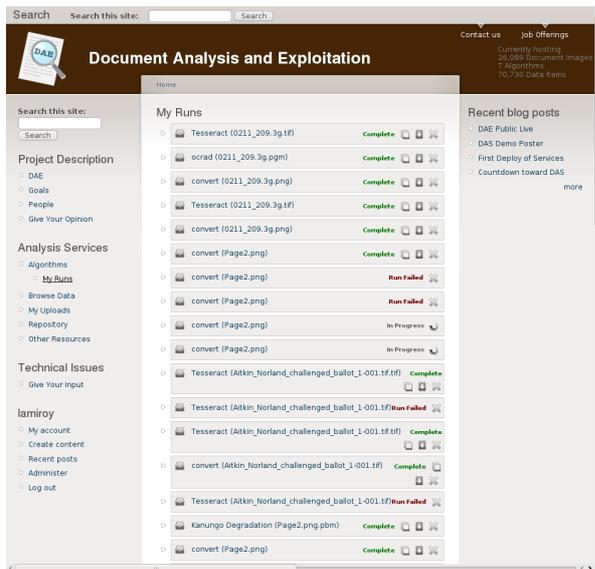
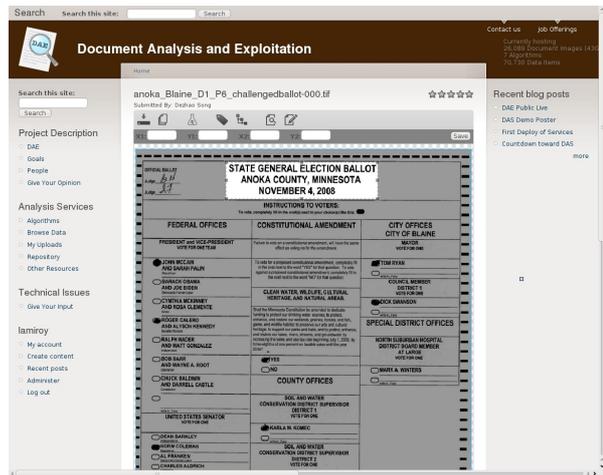
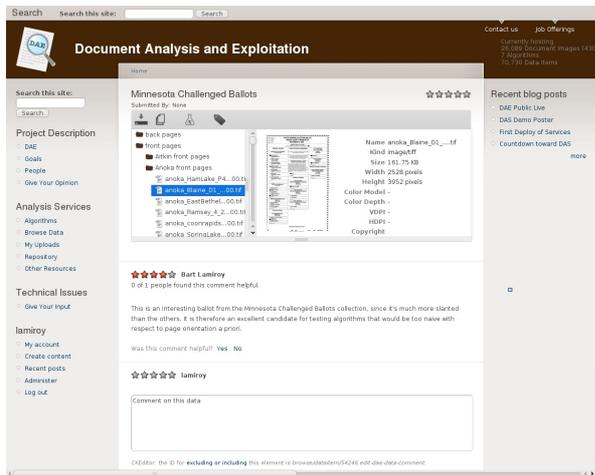


Figure 4. Data Interaction using the user interface: browsing data sets, commenting and rating (top left), image area markup creation (top right), algorithm execution (bottom left) and SQL querying (bottom right).

handwritten documents, while other applications only relate to handwritten annotations on printed documents. The latter category can even be split into contexts where the annotations are guaranteed to be actual written text, opposing others where they can also be markup glyphs or other non-textual signs and marks.

The current state of development of our platform does only allow retrieval or interaction with the data on the grounds of syntax-oriented queries. However, its design already integrates ways of extending its possibilities to a more semantic-oriented interaction.

1. Since all data provenance is recorded and accessible, one can easily retrieve data produced by specific algorithms using specific runtime parameters. This guarantees that the obtained data share the same interpretation context, and thus has some significant overlapping semantic value.
2. Since users are free to use and define their own, personal annotation tags, and since, again, provenance is recorded, one can assume that identical users, using identical tags, refer to identical semantic values.
3. Ongoing work currently consists in exporting our data model in OWL<sup>13</sup> and allowing contributors to explicitly define the semantics of their algorithms and data-relations.

**Semantic Web and Ontologies** Semantic Web<sup>14</sup> and ontologies are immediate future actors in our scenario. In order to make the querying really semantic in an as automated way as possible, and by correctly capturing the power of expressiveness as people contribute to the resource pool, the DAE platform will need to integrate adequate knowledge representations. This goes beyond the current storage of attributes and links between data. Since individual research contexts and problems usually require specific representations and concepts, contributions to the system will initially focus on their own formats. However, as the need for new *interpretations* arises, users will want to need to combine different representations of similar concepts to expand their experiment base. In order to allow them doing that, formal representations and semantic web tools are being developed in this framework.

When data needs to be shared, there are two common approaches: standardization the schema or design transformations. The first approach is very rigid, is slow to adapt to new needs, and may result in useful data being lost because there is no way to represent it. It has been tried in the past in the document analysis and its connected domains, without emergence of a global standard. The second approach can adapt to changes somewhat more easily, but because the transformations are often procedural in nature, it is difficult to reuse and/or compose them. Using semantic web technologies, ontologies can be used to provide not only schema information but additional semantic information that can aid in transformations. Also additional mapping ontologies can be used to provide declarative alignments between different domain ontologies (for instance different interpretations of the same data in other application contexts). Logical reasoners can then be used to check the correctness of the mappings and to compose them to produce indirect mappings between ontologies for which no direct mapping exists. A recent project has demonstrated how the technique works by integrating e-commerce taxonomies.<sup>15</sup>

### 2.3.3 SOA Architectures, Web-Services and the Cloud

The fact that the platform integrates the use of a fully SOA architected set of algorithms extends the data model in a similar way as the previous points. By opening the supporting architecture using carefully designed tools of remote, distributed storage and execution (aka *the Cloud*) the DAE platform may greatly reduce all risks related to scalability, availability, centralization and cost, and eventually become a community governed and distributed resource, not only hosted at Lehigh University, but shared by all its users, both in its use as in its infrastructure support.

Furthermore, the use of semantically labeled services can greatly increase the level of interaction and querying of all stored data, as hinted by in sections 2.3.2 and 2.3.2.

### 3. RELATED PROJECTS

The document-analysis community is not unique in its need for a managed resource repository. Various scientific communities maintain shared datasets that face issues similar to ours regarding the distributed sources of data, the derivation of data from other data, and the identification of the specific version of a specific dataset used in a specific publication.

The three most recent ACM SIGMOD Conferences have sought to gather the data and implemented algorithms behind accepted papers. Authors of accepted papers were invited (not required) to provide code, data, and experiment setup information to permit testing for repeatability of the reported experiments and the ability to run additional experiments with different parameter settings.

The problem of integrating multiple interpretations of a specific document is related to the problem of integrating scientific observations whether done by multiple scientists or gathered from multiple remote sensors. The Sloan Digital Sky Survey<sup>16</sup> was pioneering work that integrated world-wide astronomical observations into a single database. This work allows an astronomer to access data quickly without travel and to compare related observations made at a variety of times and from a variety of places. Microsoft's SciScope<sup>17</sup> is a tool that allows search of a wide variety of environmental databases and integrates them into a consistent format ([www.sciscope.org](http://www.sciscope.org)).

Standing in contrast to the two previously mentioned projects, which are specific to particular domains, Google Fusion<sup>18</sup> provides a framework for sharing data and visualizations of data, and for discussions pertaining to data. While Google Fusion is not domain-specific, it lacks our dual focus on both data and algorithms. However, the cloud-hosted, open framework of which Google Fusion is an example, might be a desirable long-term evolution path for our repository.

The digital-library community has studied citations to evolving documents and subparts thereof (see, e.g.,<sup>19</sup>). However, there is normally an underlying assumption in such systems that the hierarchical structure of a document (chapter, section, etc.) is well known. In document-analysis, the document structure itself may be a matter of interpretation.

Provenance, though having been studied in a variety of contexts over several years, is now emerging as a research area in its own right. While some work in provenance focuses on data and their derivation, other work focuses on workflows, or the process of data manipulation. The Panda project at Stanford,<sup>20</sup> is attempting to unify these two approaches and comes closest among prior work to capturing our goal of managing not only data but also analysis algorithms.

### 4. DISCUSSION

Apart from opening a whole new range of research practices that can be imagined from the availability of this platform, and by offering a set of fundamental tools to expand document analysis to interact with new communities like the Semantic Web and knowledge representation or databases and provenance,<sup>21</sup> to name only those, this work also addresses a number of basic concerns that have been around for a long time related to sharing common datasets (G. Nagy<sup>22</sup> acknowledges datasets dating back to 1961 in the OCR domain), “ground-truths” or *interpretations* and conducting objective benchmarking.

Upto now, datasets have always been homogeneous, either because having been produced in a common context,<sup>23-25</sup> targeted to solving a specific, well identified problem (like OCR training,<sup>1,26</sup> layout analysis<sup>27</sup> etc.) or both. Because of this, some of them have grown obsolete as technology has changed and evolved, and the notion of *document* itself has morphed over the years. Others have not been able to be kept available, since resources to do so have been consumed, or either the personal commitment and involvement of individuals or supporting groups has faded. This has always hampered reuse and combination of existing data to have it evolve.

Furthermore the costs of generating “ground truth” (or rather *reference interpretations*) remain extremely high, since, by construction, they need extensive human intervention. This makes them generally extremely and specifically problem focused and rarely generic. This makes very good sense, since they often serve as a basis for well defined contexts. However, their very specialized nature makes it difficult to re-use them in other contexts, and, because of their cost and their very specialized nature, they can produce a *lock-in syndrome*, simply because

there are no resources to re-engineer new, more appropriate ground truth data, as the domain and knowledge evolves.

These issues are largely solved by our approach in the sense that data sets are merely partial projections of the global data pool that can evolve naturally as new contributions, both document images as annotations and meta-data or *interpretations*, are added. Furthermore, since they can be used and reused in any context that is appropriate, they can more easily be used beyond their initial intended scope. Finally, since the platform is completely open, it can be reproduced, distributed and cloned so that global persistence and availability no longer are a problem.

Our current focus is also to provide interaction interfaces with a very smooth learning curve such that widespread use is not hindered by constrained and complex representation formats.

## 5. HOW THIS CAN IMPACT RESEARCH

The platform presented in this paper is a first step in a more global direction towards a more integrated and comprehensive vision of what research in Document Analysis, and more globally Machine Perception, benefit from. A more detailed description of it is open to contributions and is available at <http://dae.cse.lehigh.edu/WIKI>.

### 5.1 Improve Experimental Reproducibility and Evaluation

The main advantage of the DAE platform is that it can fulfill a role of independent certification authority, and thus contribute to enhancing reproducibility and evaluation of published results. For argument's sake, let's consider the case of a researcher developing a method operating on a specific class of documents (*e.g.* "300dpi bitonal .tif images predominately containing printed material and of which a reasonable number of pages contain at least one handwritten annotation."). Consider the following cases (in increasing order of both added value and complexity):

1. When publishing her first results the researcher can register her experimental dataset to the platform, and produce a unique and commonly accessible reference to the dataset she used to validate her work. This is not very different from current practices of publishing used datasets.
2. When her work becomes more mature and robust, she might be interested in not just using her own personal (and perhaps biased) datasets, but rely on the platform to provide her with a set of random documents corresponding to her context criteria. Hence the focus on "1000 random images ..." in the previously developed examples. Not only can the platform recall exactly what random images it provided, but it can actually provide the same set of images, as well as the original query, to anyone who requests it. This has several advantages:
  - other researchers can have access to the query that was issued; the previously potential and implicit bias to the dataset now disappears, since the operating conditions are made explicit (*e.g.* the method expects bitonal images);
  - the query result can still be stored for future reference and re-use;
  - in order to test the robustness of the published method, other, similar datasets can be produced, sharing the same set of properties.
3. Since the platform can easily keep track of uses and downloads of data, one can imagine a feature consisting in providing certified evaluation data sets that are guaranteed having never been provided to the user who requested them, while maintaining reproducibility and open referencing. This would allow researchers to publish "certified" evaluation results on data they are very likely never to have used before.
4. Given that the platform also hosts, or provides access to state-of-the-art algorithms, as well as certified data, it also becomes easy to imagine that this framework may serve as an independent benchmarking and comparison platform, allowing any user to compete against best of breed algorithms, potentially even integrated in complex end-to-end pipelines. Of course, as for the datasets, its provenance features would allow it to issue to guarantee the objectiveness and to certify the comparison results and rankings.

Furthermore, its open sourced code base, as well as its scalable features, notably its capacity to seamlessly integrate in a distributed environment, prevent it from being trusted by a single player in the field, and allow it to evolve in a truly community managed resource.

## 5.2 Open Pathways to Cross-Domain Research

As a last, but important item it is to be pointed out that, although originally concerning document analysis resources and research, many of the issues addressed in this paper open interesting research opportunities beyond: correctly and efficiently storing all provenance data is not completely solved, making and guaranteeing time-invariance of the queries mentioned in the previous section remains a challenge, making the data repository truly distributed even more so. The issues of multiple interpretations, the underlying questions of aligning the associated implicit or explicit ontologies are another major question; as is the one of automatically discovering or learning ontology-like structures from usage observation of data and algorithms. And what about the quality of added data and interpretations? Since the system both allows for user-provided ranking and commenting on the one hand, and monitors usage on the other hand, to what degree would social media inspired algorithms for reputation and ranking apply to assess the quality of crowd-sourced contributions to the hosted data ...

This platform offers the opportunity to leverage a large number of collaborative research initiatives around the concepts of Document Analysis and Interpretation in a very broad, yet concrete way.

## Acknowledgements

The DAE project and resulting platform is a collaborative effort hosted by the Computer Science and Engineering Department at Lehigh University and is funded through a Congressional appropriation administered through DARPA IPTO via Raytheon BBN Technologies.

The project has involved, over the year 2010, the following members (in alphabetical order) Chang AN, Sai Lu Mon AUNG, Henry BAIRD, Austin BORDEN, Michael CAFFREY, Siyuan CHEN, Brian DAVISON, Jeff HEFLIN, Hank KORTH, Michael KOT, Bart LAMIROY, Yingjie LI, Qihan LONG, Daniel LOPRESTI, Dezhaoh SONG, Pingping XIU, Dawei YIN, Yang YU and Xingjian ZHANG.

The authors would like to acknowledge their diverse contributions through discussions, shared thoughts, code development, *etc.*

## REFERENCES

- [1] “UNLV data set.” <http://www.isri.unlv.edu/ISRI/OCRtk>.
- [2] “Tobacco800 data set.” <http://www.umiacs.umd.edu/~zhugy/Tobacco800.html>.
- [3] “UW english document image database I: A database of document images for OCR research.” <http://www.science.uva.nl/research/dlia/datasets/uwash1.html>.
- [4] Lamiroy, B. and Najman, L., “Scan-to-XML: Using Software Component Algebra for Intelligent Document Generation,” in [*4th International Workshop on Graphics Recognition - Algorithms and Applications - GREC’2002 Lecture Notes in Computer Science*], Blostein, D. and Kwon, Y.-B., eds., *Lecture Notes in Computer Science* **2390**, 211–221, Springer-Verlag, Kingston, Ontario, Canada (10 2002).
- [5] Hu, J., Kashi, R., Lopresti, D., Wilfong, G., and Nagy, G., “Why table ground-truthing is hard,” *International Conference on Document Analysis and Recognition*, 0129, IEEE Computer Society, Los Alamitos, CA, USA (2001).
- [6] Lopresti, D., Nagy, G., and Smith, E. B., “Document analysis issues in reading optical scan ballots,” in [*DAS ’10: Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*], 105–112, ACM, New York, NY, USA (2010).
- [7] Smith, E. H. B., “An analysis of binarization ground truthing,” in [*DAS ’10: Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*], 27–34, ACM, New York, NY, USA (2010).
- [8] Clavelli, A., Karatzas, D., and Lladós, J., “A framework for the assessment of text extraction algorithms on complex colour images,” in [*DAS ’10: Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*], 19–26, ACM, New York, NY, USA (2010).

- [9] Korth, H. F., Song, D., and Heflin, J., "Metadata for structured document datasets," in [*DAS '10: Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*], 547–550, ACM, New York, NY, USA (2010).
- [10] David Doermann, Elena Zotkina, and Huiping Li, "GEDi - A Groundtruthing Environment for Document Images," in [*Ninth IAPR International Workshop on Document Analysis Systems (DAS 2010)*], (2010). Submitted.
- [11] An, C., Yin, D., and Baird, H. S., "Document segmentation using pixel-accurate ground truth," in [*20th International Conference on Pattern Recognition, ICPR*], 245–248, IEEE (2010).
- [12] Eco, U., [*The limits of interpretation*], Indiana University Press, Bloomington : (1990).
- [13] Bizer, C., "D2r map - a database to rdf mapping language," in [*WWW (Posters)*], (2003).
- [14] Feigenbaum, L., Herman, I., Hongsermeier, T., Neumann, E., and Stephens, S., "The semantic web in action," *Scientific American* **December** (2007).
- [15] Yu, Y., Hillman, D., Setio, B., and Heflin, J., "A case study in integrating multiple e-commerce standards via semantic web technology," in [*ISWC '09: Proceedings of the 8th International Semantic Web Conference*], 909–924, Springer-Verlag, Berlin, Heidelberg (2009).
- [16] Szalay, A. S., "The sloan digital sky survey and beyond," *SIGMOD Record* **37**(2), 61–66 (2008).
- [17] B. Beran, C. v. and Fatland, D., "Sciscope: a participatory geoscientific web application," *Concurrency and Computation: Practice and Experience* **22**, 2300–2312 (2010).
- [18] Gonzalez, H., Halevy, A. Y., Jensen, C. S., Langen, A., Madhavan, J., Shapley, R., and Shen, W., "Google fusion tables: data management, integration and collaboration in the cloud," in [*ACM Symposium on Cloud Computing*], 175–180 (2010).
- [19] Buneman, P. and Silvello, G., "A rule-based citation system for structured and evolving datasets," *Data Engineering Bulletin* (September 2010).
- [20] Ikeda, R. and Widom, J., "Panda: A system for provenance and data," *Data Engineering Bulletin* (September 2010).
- [21] Davidson, S. B. and Freire, J., "Provenance and scientific workflows: challenges and opportunities," in [*SIGMOD '08: Proceedings of the 2008 ACM SIGMOD international conference on Management of data*], 1345–1350, ACM, New York, NY, USA (2008).
- [22] Nagy, G., "Document systems analysis: Testing, testing, testing," in [*DAS 2010, Proceedings of the Ninth IAPR International Workshop on Document Analysis Systems*], Doerman, D., Govindaraju, V., Lopresti, D., and Natarajan, P., eds., 1 (2010).
- [23] Agam, G., Argamon, S., Frieder, O., Grossman, D., and Lewis, D., *The Complex Document Image Processing (CDIP) test collection*. Illinois Institute of Technology (2006).
- [24] Lewis, D., Agam, G., Argamon, S., Frieder, O., Grossman, D., and J.Heard, "Building a test collection for complex document information processing," in [*Proc. 29th Annual Int. ACM SIGIR Conference*], 665–666 (2006).
- [25] Thoma, G. R., "Automating data entry for an on-line biomedical database: A document image analysis application," in [*International Conference on Document Analysis and Recognition*], 370–373 (1999).
- [26] Rice, S. V., Nagy, G. L., and Nartker, T. A., [*Optical Character Recognition: An Illustrated Guide to the Frontier*], Kluwer Academic Publishers, Norwell, MA, USA (1999).
- [27] Wang, Y., Phillips, I. T., and Haralick, R. M., "Table structure understanding and its performance evaluation," *Pattern Recognition* **37**(7), 1479–1497 (2004).

# Computing Precision and Recall with Missing or Uncertain Ground Truth

Bart Lamiroy<sup>1</sup> and Tao Sun<sup>2</sup>

<sup>1</sup> Université de Lorraine – LORIA, Nancy, France – [Bart.Lamiroy@loria.fr](mailto:Bart.Lamiroy@loria.fr)

<sup>2</sup> Lehigh University – Computer Science and Engineering, Bethlehem, PA, USA

**Abstract.** In this paper we present a way to use precision and recall measures in total absence of ground truth. We develop a probabilistic interpretation of both measures and show that, provided a sufficient number of data sources are available, it offers a viable performance measure to compare methods if no ground truth is available. This paper also shows the limitations of the approach, in case a systematic bias is present in all compared methods, but shows that it maintains a very high level of overall coherence and stability. It opens broader perspectives and can be extended to handling partial or unreliable ground truth, as well as levels of prior confidence in the methods it aims to compare.

## 1 Introduction

Performance evaluation of information retrieval methods in a broad sense, *i.e.* globally any process associating high level information to a collection of weakly structured data often relies on comparing the output of the methods under evaluation to selected and verified data, for which the expected outcome of the methods is known (*cf.* [20] in graphical document analysis, for instance). These data are usually referred at as *ground truth*.

As long as the retrieval goals can be correctly captured and the scope of the data on which the methods must operate remains controllable, relying on ground-truth is possible [2, 7]. However, when the size of the potential data space becomes unmanageable or when it becomes more controversial to fully formalize the required outcome of the methods under investigation, fixing or obtaining ground truth becomes problematic to impossible. In some cases, especially when the data sets grow to a significant size, even when the retrieval process tends to favor *precision* rather than *recall* (*cf.* next section for definitions) performance evaluation approaches may rely on sampling and statistical extrapolation [8], rather than exhaustive validation. This still requires a sufficiently large set of ground-truthed data, however. Other approaches use higher level knowledge to assess coherence patterns in classified data [3].

In this paper we approach the problem differently, by making the assumption that there is either no ground truth available, or that the available ground truth may be unreliable (for instance, coming from crowd-sourced annotation processes, for which no post-processing has been done, or scenarios where human

feedback interferes with pre-established ground truth [19]). We show that by reformulating classical performance metrics like precision and recall in probabilistic terms we can establish a ranking between competing approaches that is comparable to the one that would be obtained in presence of reliable ground-truth. In that aspect, it shares some very interesting similarities with work related to classifier fusion using majority voting [11, 4]. This similarity will be addressed in Section 4.3.

Before that, and after a brief recall of the definitions of Precision and Recall in Section 2, we develop the theoretical framework of our approach in Section 3. Section 4 provides a series of experimental validations of our method and exposes some of its limitations. Further work and extensions are provided in Section 5.

## 2 Precision and Recall

### 2.1 General Definitions and Notation

Precision  $Pr$  and Recall  $Rc$  (and often associated F-measure or ROC curves) are standard metrics expressing the *quality* of Information Retrieval methods [15]. They are usually expressed with respect to a query  $q$  (or averaged over a series of queries) over a data set  $\Delta$  such that:

$$Pr_q^\Delta = \frac{|\mathcal{P}_q^\Delta \cap \mathcal{R}_q^\Delta|}{|\mathcal{R}_q^\Delta|} \quad (1)$$

$$Rc_q^\Delta = \frac{|\mathcal{P}_q^\Delta \cap \mathcal{R}_q^\Delta|}{|\mathcal{P}_q^\Delta|} \quad (2)$$

where  $\mathcal{P}_q^\Delta$  is the set of all documents in  $\Delta$ , relevant to query  $q$ , and where  $\mathcal{R}_q^\Delta$  is the set of documents actually retrieved by  $q$ . Although we can make a safe assumption by considering  $\mathcal{R}_q^\Delta$  known (*i.e.* the query  $q$  can actually be executed, and returns a known, manageable set of results), the same assumption does not always hold for  $\mathcal{P}_q^\Delta$ , as will be shown later. For ease of reading we will refer to respectively  $Pr$ ,  $\mathcal{P}$ ,  $Rc$ , and  $\mathcal{R}$ , when there is no ambiguity on  $\Delta$  and  $q$ .

Often both are combined in the  $F_\beta$  measure, where

$$F_\beta = (1 + \beta^2) \frac{Pr Rc}{\beta^2 Pr + Rc} \quad (3)$$

and where  $\beta$  expresses the importance of recall with respect to precision. Generally,  $\beta = 1$ , so that both are considered of equal importance.

### 2.2 Other Interpretations and Frameworks

Precision, Recall and the F-measure can also be defined with respect to *true positives*  $\tau_p$ , *false positives*  $\phi_p$ , *true negatives*  $\tau_n$  and *false negatives*  $\phi_n$ . In that

case, the corresponding formulas are:

$$Pr = \frac{\tau_p}{\tau_p + \phi_p} \quad (4)$$

$$Rc = \frac{\tau_p}{\tau_p + \phi_n} \quad (5)$$

$$F_\beta = \frac{(1 + \beta^2) \tau_p}{(1 + \beta^2) \tau_p + \beta^2 \phi_n + \phi_p} \quad (6)$$

Here again, it is necessary to know the values of  $\tau_p$ ,  $\phi_p$ ,  $\tau_n$  and  $\phi_p$  (as, previously, the sets  $\mathcal{P}$  and  $\mathcal{R}$ ) in order to be able to do the computations.

It is also possible to give probabilistic interpretations to  $Pr$  and  $Rc$ . In that case,  $Pr$  would be the probability that a random document retrieved by the query is relevant, and  $Rc$  that a random relevant document be retrieved by the query (taking as assumption that documents have uniform distributions). This is the interpretation we are going to use in the next sections.

### 3 Absence of Ground Truth

Previously enumerated metrics all made the assumption that the returns of queries can, in some way be qualified as “good” or “bad”. Most often, there even is the assumption that this can actually be quantified: belonging to set  $\mathcal{P}$ ,  $\tau_p$ , etc. This implies that there is some absolute knowledge of *ground truth* or an *oracle* function available for the assessment of these quantities. While it is very convenient to rely on established truth to further train or evaluate methods, it is often very costly to obtain in many cases, and even impossible in others. Furthermore, it generally requires some human intervention or validation of some sorts, which makes the ground-truthing process both difficultly scalable and error prone, and therefore costly.

This paper presents a way to estimate precision and recall using a probabilistic model, allowing either to compare algorithms operating on the same data, without the requirement of establishing ground truth, or, to leverage crowd-sourcing to establish ground truth in presence of noise, errors and mistakes. In order to achieve this, we shall first establish the underlying assumptions to our approach, in section 3.1, defining the context in which we have conceived our model. We then develop the mathematical foundations and tools in section 3.2.

#### 3.1 General Assumptions

In what follows we are assuming that the following general conditions and notations apply:

1. We are considering generic system  $\mathcal{S}$  that, given a query  $q$ , partitions<sup>3</sup> a set of documents  $\Delta = \{\delta_i\}_{i=1\dots d}$  into  $\mathcal{S}^{q+}$  and  $\mathcal{S}^{q-}$ .

<sup>3</sup> For the absent-minded reader, “partitioning”  $\Delta$  into  $\mathcal{S}^+$  and  $\mathcal{S}^-$  entails that  $\Delta = \mathcal{S}^+ \cup \mathcal{S}^-$  and  $\mathcal{S}^+ \cap \mathcal{S}^- = \emptyset$

The partitioning function  $\mathcal{S}^q$  is defined as

$$\begin{aligned} \mathcal{S}^q : \Delta &\rightarrow \{+, -\} \\ \delta_i &\mapsto \mathcal{S}^q(\delta_i) \end{aligned} \quad (7)$$

$\mathcal{S}^{q+}$  (resp.  $\mathcal{S}^{q-}$ ) is defined as the inverse image of  $\{+\}$  (resp.  $\{-\}$ ).

2. Other systems, similar to  $\mathcal{S}^q$  exist and their partitioning results are available. It is assumed that these systems operate in the same semantic context, and therefore aim to achieve the same partitioning as  $\mathcal{S}^q$ . We shall refer to the set of these systems as  $\Sigma^q = \{\mathcal{S}_i^q\}_{i=1\dots s}$

In what follows, and where it is obvious, parameter  $q$  will be omitted. Table 1 gives an example overview of what three different systems could produce for a given query over a particular document set  $\Delta$ .

$\Delta$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_1^+ = \{\delta_1, \delta_2, \delta_4, \delta_5\}$
$\delta_1$	+	+	+	$\mathcal{S}_1^- = \{\delta_3, \delta_6, \delta_7\}$
$\delta_2$	+	+	+	
$\delta_3$	-	+	-	$\mathcal{S}_2^+ = \{\delta_1, \delta_2, \delta_3\}$
$\delta_4$	+	-	-	$\mathcal{S}_2^- = \{\delta_4, \delta_5, \delta_6, \delta_7\}$
$\delta_5$	+	-	-	
$\delta_6$	-	-	+	$\mathcal{S}_3^+ = \{\delta_1, \delta_2, \delta_6\}$
$\delta_7$	-	-	-	$\mathcal{S}_3^- = \{\delta_3, \delta_4, \delta_5, \delta_7\}$

**Table 1.** Example of query systems  $\mathcal{S}_i$  operating on document set  $\Delta$

### 3.2 Performance Evaluation

The question that arises now is how to compare different  $\mathcal{S}_i$  and decide which one performs best. Traditionally, one would take an evaluation test set  $\Delta_*$  for which the ground truth of a query  $q_*$  is known and available. We shall refer to this ground truth as  $\Delta_*^+$  and  $\Delta_*^-$  (i.e.  $\Delta_*^+$  is the partition of  $\Delta_*$  containing the documents corresponding to  $q_*$ ,  $\Delta_*^-$  its complement). This knowledge then allows to compute precision and recall values, as described in Section 2, for all  $\mathcal{S}_i$  and establish a performance metric adapted to the context under consideration.

When  $\Delta_*^+$  and  $\Delta_*^-$  are unavailable, it is less obvious to compare the results of the different  $\mathcal{S}_i$ . One well documented approach is to use statistical estimators by considering each  $\mathcal{S}_i(\Delta)$  as the outcome of some random variable. What we are going to develop here, is very similar, but particularly focused on the expression of precision and recall.

**Simplified Case** First we're making the assumption that all  $\mathcal{S}_i$  are of equal importance, and that there is no *a priori* knowledge available allowing to presume

some of the systems are more reliable than others. This assumption will be alleviated in later work. We also assume all documents have equal frequency and occurrence probability.

For the arguments developed next, we need to introduce two “virtual” query systems,  $\mathcal{S}_\top$  and  $\mathcal{S}_\perp$ .  $\mathcal{S}_\top$  always returns all documents for any given query,  $\mathcal{S}_\perp$  never returns any. In other terms,

$$\mathcal{S}_\top^+ = \Delta, \mathcal{S}_\top^- = \emptyset \quad (8)$$

$$\mathcal{S}_\perp^+ = \emptyset, \mathcal{S}_\perp^- = \Delta \quad (9)$$

We are also slightly reconsidering the partitioning function defined in equation (7), such that it returns values in  $\{1, 0\}$  rather than in  $\{+, -\}$ .

Under these hypotheses, the probability that a document  $\delta_i$  belongs to  $\Delta_\star^+$  is

$$P(\delta_i) = \frac{1}{s+2} \sum_{k=1\dots s, \perp, \top} S_k(\delta_i) \quad (10)$$

The results of the application of this to the example in Table 1, is represented in Table 2.

$\Delta$	$P(\delta_i)$	$\mathcal{S}_\top$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_\perp$
$\delta_1$	0.8	1	1	1	1	0
$\delta_2$	0.8	1	1	1	1	0
$\delta_3$	0.4	1	0	1	0	0
$\delta_4$	0.4	1	1	0	0	0
$\delta_5$	0.4	1	1	0	0	0
$\delta_6$	0.4	1	0	0	1	0
$\delta_7$	0.2	1	0	0	0	0

**Table 2.** Example

Given the hypothesis of equidistribution of all documents  $\delta_i$  in  $\Delta$  and given the probabilistic definition of precision in Section 2.2, stating that  $Pr$  “is the probability that a random document retrieved by a query is relevant”, we can now define  $Pr(\mathcal{S}_k)$ :

$$Pr(\mathcal{S}_k) = \frac{\sum_{i=1\dots d} P(\delta_i) S_k(\delta_i)}{\sum_{i=1\dots d} S_k(\delta_i)} \quad (11)$$

Similarly,  $Rc$  was defined as “the probability for a random relevant document to be retrieved by the query”. In our case, however relevancy has no longer a binary value, but has been replaced by  $P(\delta_i)$ . By reformulating this conditional probability and using Bayes’ theorem (and using the fact that the inverse

conditional of  $Rc$  is  $Pr$ ), things smooth out elegantly.

$$\begin{aligned}
Rc(\mathcal{S}_k) &= Prob\left(\text{retrievedBy}_{\mathcal{S}_k}(\delta_i) \mid \text{isRelevant}(\delta_i)\right) \\
&= Prob\left(\text{isRelevant}(\delta_i) \mid \text{retrievedBy}_{\mathcal{S}_k}(\delta_i)\right) \frac{Prob(\text{retrievedBy}_{\mathcal{S}_k}(\delta_i))}{Prob(\text{isRelevant}(\delta_i))} \\
&= Pr(\mathcal{S}_k) \frac{\frac{1}{d} \sum_{i=1..d} S_k(\delta_i)}{\frac{1}{d} \sum_{i=1..d} P(\delta_i)} \\
&= \frac{\sum_{i=1..d} P(\delta_i) S_k(\delta_i)}{\sum_{i=1..d} S_k(\delta_i)} \frac{\sum_{i=1..d} S_k(\delta_i)}{\sum_{i=1..d} P(\delta_i)} \\
&= \frac{\sum_{i=1..d} P(\delta_i) S_k(\delta_i)}{\sum_{i=1..d} P(\delta_i)} \tag{12}
\end{aligned}$$

It is interesting to notice the resemblance between equations (1) and (11) as well as between (2) and (12). Table 3 shows the values obtained when applied to the examples of Table 2.

$\Delta$	$P(\delta_i)$	$\mathcal{S}_\top$	$\mathcal{S}_1$	$\mathcal{S}_2$	$\mathcal{S}_3$	$\mathcal{S}_\perp$
$\delta_1$	0.8	1	1	1	1	0
$\delta_2$	0.8	1	1	1	1	0
$\delta_3$	0.4	1	0	1	0	0
$\delta_4$	0.4	1	1	0	0	0
$\delta_5$	0.4	1	1	0	0	0
$\delta_6$	0.4	1	0	0	1	0
$\delta_7$	0.2	1	0	0	0	0
Sum	3.4	7	4	3	3	0
$\sum P\mathcal{S}_k$		3.4	2.4	2	2	0
$Pr$		0.49	0.6	0.67	<b>0.67</b>	$\infty$
$Rc$		1	<b>0.71</b>	0.59	0.59	0

**Table 3.** Example of precision and recall computations without established ground truth.

## 4 Experimental Validation

In order to experimentally validate the model developed we have taken two contexts. One consists in taking the results of experiments reported in [10] related to comparing standard symbol recognition techniques. A second is related to evaluation of binarization algorithms on downstream treatment and is very similar to the experiments conducted in [13].

#### 4.1 Symbol Recognition

In this section we use the experimental results reported in [10]. In this paper, the authors compare 5 different symbol recognition methods on a set of electrical wiring diagrams. Since their dataset has no known ground truth, they use a panel of human annotators to select and determine which ground truth corresponds to which query.

Since the authors in [10] report retrieval efficiency, as defined in [9], we have resampled their raw experimental data to extract precision and recall. The results, with respect to the human-defined ground-truth reported by the authors is shown in Figure 1.

Figure 2 reproduces the precision and recall values obtained using our method on the exact same data. It is interesting to note that, with one noteworthy exception, the ordering of the tested methods, with respect to precision or recall (*i.e.* when ordering methods from high precision/recall to low) is respected. Although not reproduced here, this also holds for the F-measure. What is even more compelling, is that the methods 'SC' and 'GFD' maintain their similarity in both cases, with and without consideration of ground truth.

The one exception is the 'ARG' method. While considered as a tie with 'SC' and 'GFD' with our method, it significantly outperforms all other approaches according to the ground truth. This is a very interesting result, and is currently under investigation.

#### 4.2 Document Binarization

The data used in this second study are the historical images collected from the Library of Congress on-line data set[1]. A total of 60 TIF format images with a resolution of 300 dpi. Various genres from official documents to private letters are included. The degraded quality of these images, such as uneven illumination, bleeding-through, handwritten marks, *etc.* are a great challenge for recognition algorithms. In this case, we are going to try and use our approach to evaluating binarization quality to downstream recognition, as in [13]. The document image analysis pipeline consists of three stages: binarization – OCR – named entity recognition.

Binarization is the first stage, and three thresholding methods are used in this stage respectively. They are Otsu [14], Sauvola [16] and Wolf [21]. Otsu's method is a global thresholding method while the latter two are local thresholding methods. After all the images are converted into binary images, the resultant binary images were converted to ASCII texts by the Tesseract-3.00 [17] open source software package in the second stage. Finally, Stanford Named Entity Recognizer [5] is used in the third stage. To sum up, we have three different pipelines this way. Although our method aims to calculate precision and recall without ground truth, we still need ground truth to evaluate if our method can achieve the goal proposed in Section 3.2. Since the ground truth of the historical images are not directly available, we generate the ground truth ourselves by manual typing the text and carefully proofreading.

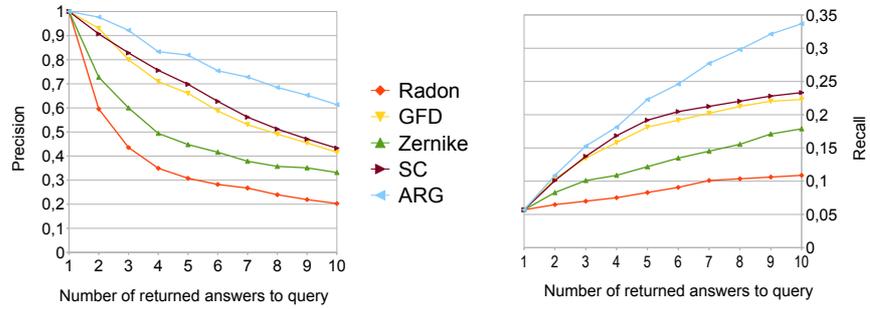


Fig. 1. Precision and Recall as reported in [10]

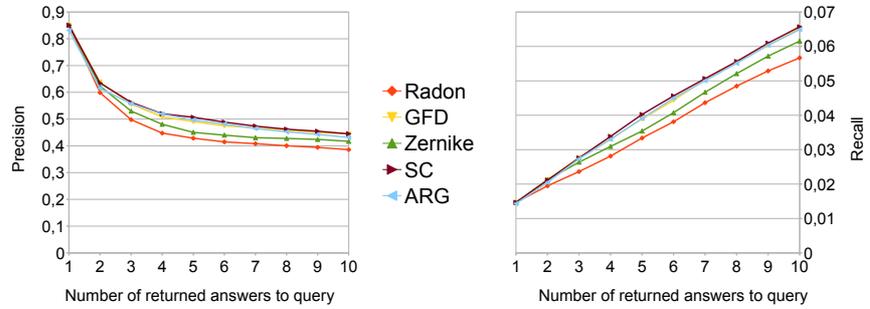


Fig. 2. Precision and Recall as computed without ground truth

Since the three different pipelines depend on three different thresholding methods, we use the names of them to stand for the three pipelines, respectively. The calculation of average precision and recall is based on the outputs of these pipelines, which are the named entity extraction results. When evaluating our method, we use two different ways to process the outputs of the three pipelines. Method I considers all the recognized named entities as ‘bag-of-words’, so they are organized in an alphabetical way. While Method II uses a multiple sequence alignment algorithm [18] to align the three outputs first, the original positions of these named entities are kept this way. The experiment results are shown in the following tables. From Table 4 we can see that Sauvola and Wolf beat Otsu thresholding method. The reason is obvious. Only one threshold is determined for the whole image by Otsu, while for the other two methods, different thresholds are calculated according to the grey distribution of their corresponding local windows. Table 5 and Table 6 show the results of our ground-truthless precision and recall measures using each of the metrics described before (Method I and

	Otsu	Sauvola	Wolf
Precision	0.6223	0.7715	0.7533
Recall	0.5915	0.7281	0.7230

**Table 4.** Average Recognition Accuracies with Ground Truth

	$S^\top$	Otsu	Sauvola	Wolf	$S_\perp$
Precision	0.4000	0.6327	0.6757	0.6722	$\infty$
Recall	1.0000	0.5153	0.5660	0.5662	0

**Table 5.** Method I: Average Recognition Accuracies without Ground Truth

	$S^\top$	Otsu	Sauvola	Wolf	$S_\perp$
Precision	0.5733	0.6035	0.6450	0.6416	$\infty$
Recall	1.0000	0.6550	0.6988	0.6957	0

**Table 6.** Method II: Average Recognition Accuracies without Ground Truth

II). We can see again the performance of Sauvola and Wolf is better than that of Otsu, while recognition accuracies between Sauvola and Wolf are similar. Both of them indicate that even if without ground truth, the precision and recall computed by our method is similar to those computed with ground truth.

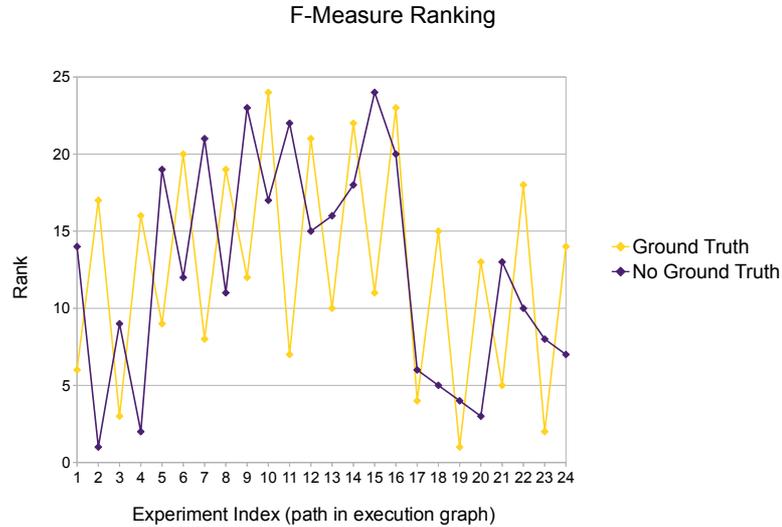
### 4.3 Limitations

It would be an error to consider the approach developed in this paper as a complete and equivalent replacement of ground truth. Since the approach consists in finding an overall consensus between the tested methods, it is sensitive to collective bias. This is illustrated in the following example, taken from the raw data of the ICDAR 2011 contest described in [13].

The contest setup is quite similar than the one used in the previous section where its general aim is concerned. The difference lies in the fact that 24 different 4-stage pipelines are compared to one another. The document analysis pipelines consist in binarization – text segmentation – OCR and named entity detection, using 3 different binarization algorithms, 4 text segmentation methods and 2 OCRs.

As reported in [13], the tested pipeline is very sensitive to the quality of the used OCR engine. The results obtained using the 24 different execution paths, where every other path uses one of the 2 tested OCR engines, show that one of them clearly outperforms the other.

In order to compare these results with the approach developed in this paper we are not going to use raw F-Measure values, since the previous results have shown that there may be a significant difference in range. Instead, we are going to look at the ranking of the different methods with respect to their decreasing F-Measure. Using the Method I of the previous section, we obtain the results represented in Fig. 3.

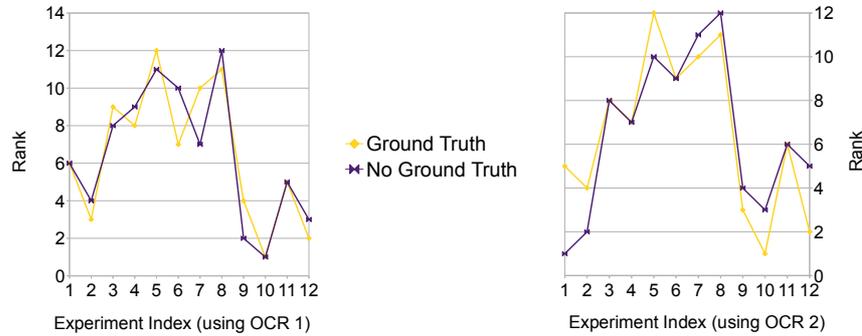


**Fig. 3.** Comparison of F-Measure ranking between ground truth based and ground truth-less measures.

There are two observations to be made regarding these results. The first, quite puzzling one, is that although both curves follow the same global trend, they are in complete opposite phase with respect to the oscillation induced by the OCR quality. Second, a closer look at the figures shows that there is an averaging effect operating. Since both OCR engines are consistent in their errors, they introduce a bias in the consensus values computed by our method, thus pulling the F-Measures toward an average value.

By separating the results in function of the OCR, we observe that we obtain much more coherent, and more encouraging results, in line with what we observed in the previous sections. Fig. 4 shows that the overall ranking pattern is preserved when projecting the F-Measures by OCR. It is clear, on the other hand, that there is no total equivalence between the ranking obtained with ground truth and the one obtained without. However, global ranking (top – middle – bottom tiers) is very consistent.

These results very much recall the experiments reported in [11] in the case of classifier fusion. Although there are some fundamental difference in combining binary classifiers by majority voting and the approach developed here, the underlying formalism is very much the same. The main differences are that on the one hand, we are not applying a full majority vote, in our case. Although the probability of an individual document being relevant depends on the number of systems having classified it as such, and therefore relates to a voting system, this probability is not truncated to either 0 or 1, as it would have been, in the



**Fig. 4.** Comparison of F-Measure ranking between ground truth based and ground truth-less measures in function of the underlying OCR method.

case of majority voting. On the other hand, the goal of classifier fusion is to obtain a new classifier, performing better than its individual contributors. This is not the aim in our case, where we just want to express a ranking between the different classifiers. One may argue, however, that the classifier obtained by majority voting may provide a theoretical boundaries to the reliability of the probabilistic Precision and Recall values presented in the previous sections. The math behind this assumption needs to be further developed and assessed.

## 5 Extensions

The probabilistic model developed in section 3.2 makes the simplifying assumption that both all data and all methods have uniform confidence values: no method is considered more reliable than the others, and all data either belongs or does not belong to the query results.

### 5.1 Method Weighting

Our model is capable of integrating ground truth, and may even handle uncertain ground truth (*e.g.* coming from reliable, but not fully verified human annotations). To that avail, the ground truth can be integrated as being the result of some “oracle” system  $S_{\mathcal{O}}$ , and the probability of a document  $\delta_i$  belonging to  $\Delta_{\star}^+$ , as expressed in (10) should be slightly modified.

$$P(\delta_i) = \sum_{k=1 \dots s, \perp, \top, \mathcal{O}} S_k(\delta_i) \kappa_{S_k} \quad (13)$$

Where  $\kappa_{S_k}$  is the confidence value associated to system  $S_k$ , and  $\sum_k \kappa_{S_k} = 1$ . In the case we previously developed, all systems had equal confidence, and

$\kappa_{S_k} = \frac{1}{s+2}$ . In case of one or more oracle systems  $S_{\mathcal{O}}$ , its confidence value can be adapted consequently. Setting  $\kappa_{S_{\mathcal{O}}} = 1$  would be equivalent to the commonly admitted use of (undisputed) ground truth. Moreover, in cases where multiple versions of reference interpretations exist [12] it now becomes possible to handle varying degrees of ground “truth”<sup>4</sup> by attributing appropriate values to the corresponding oracle systems.

## 5.2 Confidence Voting

Similarly, it is now possible to extend the approach beyond binary attribution of documents to queries, since systems can very well express their confidence of a document being relevant to a query with a probability value. All formulae and tables developed in section 3 remain valid in this context, and the probabilistic precision and recall computations are directly transposable to the case where individual documents for a given query have a probability of pertinence rather than a binary valuation. Furthermore, this can be combined with the method weighting expressed in the previous section.

## 6 Conclusion and Future Work

In this study we have presented how to compute precision and recall without presence of formally identified ground truth. Results indicate that this measure is coherent with real, ground truth based precision and recall measures, although it can obviously not infer ground truth and achieve the exact same performance as if ground truth were actually available.

On the other hand, the mathematical framework supporting the computation of probabilistic precision and recall has the interesting property to handle a continuum of situations ranging from perfectly known and available ground truth, over uncertain ground truth to total absence of it.

The major condition for this method to work, however, is that it has access to a number of competing systems that are providing multiple possible answers to the same queries, each of them supposedly trying to achieve the best possible result. This is particularly well suited for large scale performance evaluation contexts like the one experimented in [13] and formally developed in [12]. Its use in larger scale experiments will also contribute in further establishing the exact differences between full use of ground truth and the approximation presented in this paper.

Further work and development will consist in establishing how to rank or take into account user-contributed “partial” ground truth, especially considering “yes/no/unknown” information. Currently, our framework makes the assumption that all systems operate on the exact same set of queries and documents. There exist models that are capable of integrating overlapping or dissimilar

---

<sup>4</sup> Since there cannot exist varying degrees in truth, we prefer the term of “interpretation”.

query and document sets [6]. It would be interesting to confront them to our approach and to study how partial ground truth (for instance, resulting from crowd-sourced contributions) can be integrated and improve overall performance of our approach.

## Acknowledgements

The authors would like to acknowledge Dr. Santosh K.C. for having provided the experimental data, used in Section 4.1. They also thank Prof. Dan Lopresti for having pointed them to voting approaches in classifier fusion.

Bart Lamiroy was a visiting scientist at Lehigh University in 2010–2011. This work was conducted at the Computer Science and Engineering Department at Lehigh University and was supported in part by a DARPA IPTO grant administered by Raytheon BBN Technologies.

## References

1. Library of congress, <http://memory.loc.gov/>
2. Antonacopoulos, A., Karatzas, D., Bridson, D.: Ground truth for layout analysis performance evaluation. In: Bunke, H., Spitz, A. (eds.) Document Analysis Systems VII, Lecture Notes in Computer Science, vol. 3872, pp. 302–311. Springer Berlin / Heidelberg (2006)
3. Baraldi, A., Bruzzone, L., Blonda, P.: Quality assessment of classification and cluster maps without ground truth knowledge. *Geoscience and Remote Sensing, IEEE Transactions on* 43(4), 857 – 873 (april 2005)
4. Bauer, E., Kohavi, R.: An empirical comparison of voting classification algorithms: Bagging, boosting, and variants. *MACHINE LEARNING* 36, 105–139 (1999)
5. Finkel, J.R., Grenager, T., Manning, C.D.: Incorporating non-local information into information extraction systems by gibbs sampling. In: ACL. The Association for Computer Linguistics (2005)
6. Goutte, C., Gaussier, E.: A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. *Advances in Information Retrieval 27th European Conference on IR Research ECIR 2005 Santiago de Compostela Spain March 2123 2005 Proceedings* 3408, 345–359 (2005)
7. Grosicki, E., Carree, M., Brodin, J.M., Geoffrois, E.: Results of the rimes evaluation campaign for handwritten mail processing. In: Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on. pp. 941–945 (july 2009)
8. Hauff, C., Hiemstra, D., de Jong, F., Azzopardi, L.: Relying on topic subsets for system ranking estimation. In: Proceedings of the 18th ACM conference on Information and knowledge management. pp. 1859–1862. CIKM '09, ACM, New York, NY, USA (2009)
9. Kankanhalli, M.S., Mehre, B.M., Wu, J.K.: Cluster-based color matching for image retrieval. *Pattern Recognition* 29, 701–708 (1995)
10. K.C., S., Lamiroy, B., Wendling, L.: Spatio-structural symbol description with statistical feature add-on. In: The Ninth International Workshop on Graphics Recognition (2011)
11. Kuncheva, L., Whitaker, C., Shipp, C., Duin, R.: Limits on the majority vote accuracy in classifier fusion. *Pattern Analysis & Applications* 6, 22–31 (2003)

12. Lamiroy, B., Lopresti, D., Korth, H., Jeff, H.: How carefully designed open resource sharing can help and expand document analysis research. In: Agam, G., Viard-Gaudin, C. (eds.) Document Recognition and Retrieval XVIII. SPIE Proceedings, vol. 7874. SPIE, San Francisco, CA USA (January 2011)
13. Lamiroy, B., Lopresti, D., Sun, T.: Document Analysis Algorithm Contributions in End-to-End Applications. In: 11th International Conference on Document Analysis and Recognition - ICDAR 2011. International Association for Pattern Recognition, Beijing, China (Sep 2011)
14. Otsu, N.: A threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man and Cybernetics* 9(1), 62–66 (Jan 1979)
15. van Rijsbergen, C.J.: *Information Retrieval*. Butterworth (1979)
16. Sauvola, J.J., Pietikäinen, M.: Adaptive document image binarization. *Pattern Recognition* 33(2), 225–236 (2000)
17. Smith, R.: An overview of the tesseract ocr engine. In: ICDAR '07: Proceedings of the Ninth International Conference on Document Analysis and Recognition. pp. 629–633. IEEE Computer Society (2007), <http://www.google.de/research/pubs/archive/33418.pdf>
18. Thompson, J.D., Higgins, D.G., Gibson, T.J.: Clustal w: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* 22(22), 4673–80 (1994)
19. Tombre, K., Lamiroy, B.: Pattern recognition methods for querying and browsing technical documentation. In: Ruiz-Shulcloper, J., Kropatsch, W. (eds.) *Progress in Pattern Recognition, Image Analysis and Applications, Lecture Notes in Computer Science*, vol. 5197, pp. 504–518. Springer Berlin / Heidelberg (2008)
20. Valveny, E., Dosch, P., Winstanley, A., Zhou, Y., Yang, S., Yan, L., Wenyin, L., Elliman, D., Delalandre, M., Trupin, E., Adam, S., Ogier, J.M.: A general framework for the evaluation of symbol recognition methods. *International Journal on Document Analysis and Recognition* 9, 59–74 (2007)
21. Wolf, C., Doermann, D.S.: Binarization of low quality text using a markov random field model. In: *ICPR* (3). pp. 160–163 (2002)



## Chapter 5

# Research Project: the Limits of Image Interpretation

### Preamble

The methodology and reasoning in this document will try to convince the reader of a number of hard and unsolved issues related to machine perception and interpretation in general. At given key points, paragraphs will be marked as

**Research Proposal 1:** Quisque ullamcorper placerat ipsum. Cras nibh. Morbi vel justo vitae lacus tincidunt ultrices. Lorem ipsum dolor sit amet, consectetur adipiscing elit. In hac habitasse platea dictumst. Integer tempus convallis augue. Etiam facilisis. Nunc elementum fermentum wisi. Aenean placerat. Ut imperdiet, enim sed gravida sollicitudin, felis odio placerat quam, ac pulvinar elit purus eget enim. Nunc vitae tortor. Proin tempus nibh sit amet nisl. Vivamus quis tortor vitae risus porta vehicula.

These are unsolved problems for which a reasonable amount of tools and knowledge is available, and for which it is possible to formulate an approach towards resolution. They may typically result in medium term tasks, grant proposals or Ph.D. topics.

In some cases, they may have been the topic of proposals or submissions to calls for proposals by funding agencies. In that case they will be mentioned as such.

The rest of this document will develop a series of ideas and thoughts about the fundamentals of image interpretation, how it is considered in the current state-of-the art and why this conception contains a number of flaws, which, in our opinion, seriously constrain further significant development of research in the domains of visual recognition and interpretation of data. Many of the ideas expressed here directly stem from work described in chapter [3.3](#).

The overall goal of the rest of this document is to introduce and convince the reader of the need to develop and investigate the research directions and more technical projects presented at the end of the chapter. However, we feel it is necessary to sufficiently and explicitly introduce and develop the reasoning behind their justification. Therefore, § 5.1 relates our previous work to the global motivation of this chapter. It can be summarized as follows: Machine Perception (and document image analysis) problems are essentially concerned with interpretation and analysis, with as a general, overall goal, to match human performance; to measure quality and progress in this domain, it is necessary to benchmark and compare results; this is generally done by comparing results to *ground truth* or golden standards. While this is generally accepted good practice, reality shows it is very difficult to actually assess the global advance of the state-of-the-art on a broad range perception domains, and that there is a lack of tools and resources to correctly evaluate various approaches.

As a consequence, we analyse the concepts of ground truth (§ 5.2) and interpretation (§ 5.3) and come to the disturbing conclusions that ground truth is a fundamentally ambiguous concept, inherently related to interpretation context, and that neither can be defined in a formally satisfactory way. Section 5.4.3.0 therefore investigates a series of tools and associated paradigm shifts to considering machine perception that may lead to means of addressing the uncovered apparent contradictions.

## 5.1 Introduction

The general idea and research proposal developed in the following sections finds its origins in a series of thought experiments and introspective analysis of how to sift through the enormous amounts of produced experimental research in Document Image Analysis. More particularly, how can we decide, within the scope of all published available algorithms and approaches, which one best suits a particular problem, or in other terms, how far the state-of-the-art is from solving a specific problem, or what sub-problems can be considered solved [Lopresti and Nagy, 2011, Lopresti and Nagy, 2012]? These are essential questions, when trying to efficiently conduct research in a given field of study, and it would seem that the answers are far from trivial, as this chapter will show.

Addressing these questions, necessarily leads to connecting the general concept of data interpretation (mainly applied to Document Image Analysis, for practical reasons, and to Machine Perception in general) to performance analysis and experimental research reporting as developed in [23]. We started by formalizing this point in [41]:

The goal of document image analysis is to achieve *performance* using automated tools that is *comparable* to what a *careful* human expert would achieve, or at least to do *better* than *existing algorithms* on the same *task*.

Our use of terms like “performance,” “comparable,” and “better” indicate that there is an underlying notion of *quality* and therefore *measurement*. It sug-

gests a controlled process that continually improves toward perfection. However, we also make mention of “careful” humans, “tasks,” and “existing algorithms.” While humans may believe themselves to be expert and careful when performing a task, there are situations where they unavoidably disagree [Hu et al., 2001a, Lopresti et al., 2010, Smith, 2010, Clavelli et al., 2010], meaning that, at best, quality and improvement are subjective notions. It also strongly suggests that, depending on the task, measurements will differ, advocating again for multiple ways of measuring overall performance.

On the other hand, it is commonly accepted that shared reference benchmarks are essential in scientific domains where reproducible experiments are vital to the peer review process. For instance, there have been numerous attempts [UNLV, , Tobacco, , UW1, , UW2, , UW3, ] to produce common datasets for problems which arise in document analysis. It is important to note, however, that shared datasets are only a part of what is needed for performance evaluation, and since research in document analysis is often task-driven, specific interpretations of a dataset may exist. So whether the problem is invoice routing [Schulz et al., 2009], building the semantic desktop [Dengel, 2009], digital libraries, global intelligence [MADCAT, ], or document authentication, to name a few, the result tends to be application-specific, resulting in software solutions that integrate a complete pipeline of cascading methods and algorithms [Liang et al., 1997, Stefan Jaeger et al., 2006a]. This most certainly does not affect the intrinsic quality of the underlying research, but it does tend to generate isolated clusters of extremely focused problem definitions and experimental requirements. Crossing boundaries and agreeing on what kinds of tools, formats, measurements, *etc.* are the most useful is difficult and may, in fact, be impossible since the pursuit of goals may be prove orthogonal between domains.

Notwithstanding, it seems essential that collections of evaluation documents be annotated down to a fine level, the so-called “ground-truth” (*e.g.*, the location and identity of every character represented in the document). Even richer annotations may be desirable in some cases, *e.g.*, an interpretation that includes the type size and typeface for each character. The amount of manual intervention needed depends not only on the quality of the input document, but also on the requirements on the quality of the output.

It is generally assumed that there is a single, unambiguous annotation in every case and that it is recorded correctly in the ground-truth. While document image analysis may some day completely automate the task of creating such annotations, as of today, some manual intervention is required in all but the simplest of cases, or the lowest of expectations. In practice, such systems are “brittle” and a wide range of errors may arise. Some of these can severely impact intended uses of the acquired information. In practice, users must either tolerate a substantial amount of noise, or else forgo applications that

are predicated on the assumption the collection is noise-free.

Existing tools allow the user to indicate how he/she believes a document should be interpreted, but do little to help users understand differences in interpretations. Such differences might be called “errors” when there is a strong consensus about what constitutes the right answer. In many cases, however, there are legitimate differences of opinion [Hu et al., 2001b, Lopresti and Nagy, 2001] by various “readers” of the document, and these may differ from the *intention* of the author (which is usually hard or impossible to determine, although sometimes we can get access to it [Eco, 1990]).

The bottom line is that although standard document collections exist, their annotations or “ground truth” may be specific, recorded in pre-determined representations, incomplete or partially erroneous, while, on the other hand, there is a need to collect and manage annotations in ways that make it possible to construct more robust and general document analysis solutions.

This excerpt, although considering the topic from the angle of document image analysis, easily applies to broader machine perception research. It raises the following fundamental questions:

1. How can individual contributions to the state-of-the-art, solving machine perception problems, be objectively evaluated? Can they be compared to previous work? Can there be a set of measurable criteria establishing that it actually contributes to improving the state-of-the-art?
2. In how far are these contributions constrained to a specific context of use? What is a context of use? Can it be described, formalized or measured?
3. Is it actually possible to evaluate a contribution with respect to human perception performance? Does it make sense? What would be required to be able to do so?

While these questions seem to be naively simple and common sense, there does not seem to be any thoroughly established framework addressing them. They actually seem to be taken for granted and “*obvious*”. We shall prove in what follows that they are far from being so, and that considering them lightly actually leads to severely distorted perceptions of the quality of research in many ways.

This is far from being a new problem, and has been considered before. For instance, part of it is very closely related to one of the very basic aspects of experimental research: the level of “*verifyability*” and traceability of claims and results published by their authors. In [21] we already wrote that:

The question on how to produce and report scientifically sound and valid experiments is not new [Popper, 1992], it is an on-going debate in all disciplines, e.g. [Moret and Shapiro, 2001, Schwab et al., 2000].

The basics are:

1. reporting of clearly set goals and defined interpretation framework,
2. full access to all experimental data,
3. reporting of the experimental apparatus, setup and protocol, in such a way that it becomes fully reproducible,
4. all parameters defining the data (if applicable) and those related to the experimental process.

While these seem obvious, [...] the currently available resources fail to produce the effect of fully reproducible open experiment reporting. The reasons for this include:

**Full disclosure and complete reporting** is often difficult to achieve for methods and algorithms, notably because of space constraints on publications. Even if the reporting is completely transparent, it still may be hard to reproduce complex algorithms and obtain identical behavior due to implementation choices, bugs, *etc.* Making source code available, or using shared development or execution platforms [Breuel, 2008, Rendek et al., 2004, Stefan Jaeger et al., 2006b] is helpful, but practice shows that this only rarely yields comparative studies. The reason for that is that the platforms are very much technology dependent (choice of specific programming languages, operating systems, data structures or other constraining paradigms) that often require a time investment that discourages others from using it. They also may suffer from progressive obsolescence when not actively maintained. Releasing source code can also be problematic when private funding, IP or patents come into play.

**Full access to all experimental data** should not really be an issue, given today's ubiquitous access to storage and bandwidth (although this does become an issue when the amount of data becomes too large [Smeaton et al., 2006]), but there are more subtle difficulties. The way benchmark datasets are currently conceived and made available is rather "monolithic" in the sense that they have usually been created for a specific experimental context, and that their intrinsic parameters (*e.g.* type of images, resolution, content, frequency ...) and associated interpretations are those that suit this context. In order to adapt to these implicit constraints, re-use of existing datasets often comes with recomposition, selection and filtering of the original data, blurring the exact boundaries of the effectively used data.

**Exact description of all parameters** is difficult to provide, especially for data, since they often reflect a mix of arbitrary design decisions (my method does only take `.tif` images) and more subjective ones ("reasonably" good

scan quality so that the OCR doesn't fail). Because of different experimental contexts, it is rarely the case that exhaustive experiments are reported over complete datasets, without the selection and filtering mentioned beforehand. This also sometimes holds for contests where training data is not always formally characterized, for instance.

However, before reaching these more fundamental considerations, it is necessary to trace back the origins of their development. This requires to look into to performance analysis, and how experimental research needs to be conducted in order to be valid, as we already outlined in chapter 3.3.

The rest of this document is organized as follows: section 5.2 starts with considering some more formal approaches to look into Machine Perception algorithm performance analysis, and how it relates to the limits (and, as we shall see, unavoidable subjectivity) of ground truth specification. It will establish the intrinsic ambiguity of *interpretation* and *analysis* when used in conjunction with Machine Perception; section 5.3 therefore takes a closer look to these concepts, mainly in relation to Document Image Analysis, so that in section 5.4 we can introduce some approaches to actually model notions of interpretation, context and relate them to performance analysis.

Since the arguments in this thesis are constructed in increasing order of abstraction, we start with a set of basic considerations, using commonly admitted definitions and conditions. As limitations and contradictions will start to appear, the need of more formally address various issues will start to appear.

Formulated differently, the rest of this chapter will outline and justify the following general proposal in detail.

After establishing the fact that interpretation is open to ambiguity and that most of this ambiguity comes from inconsistent or different interpretation contexts our overall goal is to investigate whether one can:

- establish a form of context description that is appropriate for machine perception (and document image analysis in particular) and whether it can be obtained automatically by statistical or formal learning techniques?
- use this context description to evaluate algorithm performances?
- use this context description to describe data, so that it can be used for information retrieval purposes?
- establish formal boundaries or limitations for the previously described descriptions and establish whether there are interpretations that are provably impossible to be obtained through an algorithm. If there is indeed a class of interpretation problems that cannot be solved by an algorithm, the second question would be whether this class can be characterized in some sorts.

## 5.2 Comparing Machine Perception Algorithms

The first question we are considering is related to evaluating different approaches trying to solve a same identification or interpretation problem (*cf.* work presented in section 3.3 like [8]). As mentioned in the introduction, one of the major goals of machine perception consists in achieving near-human performance in identifying perceptual input.

Since comparing algorithms to human performance is far from being obvious, and is something that is very rarely actually measured, most published contributions to the state-of-the-art contain claims of improvement with respect to existing approaches. How are these claims measured, and what do they actually learn us? Does actually make any sense to try and relate human performance to an algorithm, with respect to the Church-Turing thesis [Jones, 1997] (*i.e.* can human performance be calculated by an algorithm)?

### 5.2.1 Notations and Definitions Relating to Reference Data

One of the most basic comparison approaches in Machine Perception is to use a standard set of reference data  $\Delta$  for which a (supposedly human) oracle  $\mathcal{O}$  has provided the expected correct interpretations within a set of allowable interpretation values  $\mathcal{I}$ . We can therefore consider the oracle  $\mathcal{O}$  as a function associating an interpretation to each of the items in  $\Delta$ .

$$\begin{aligned} \mathcal{O} : \Delta &\rightarrow \mathcal{I} \\ \delta &\mapsto \mathcal{O}(\delta) \end{aligned} \quad (5.1)$$

In what follows we shall be using the slightly debatable but more convenient notation below, expressing the set of interpretations according to  $\mathcal{O}$  as follows:

$$\mathcal{O}(\Delta) = \mathcal{I}_{\mathcal{O}} = \{(\delta, \mathcal{O}(\delta))\}_{\delta \in \Delta}$$

Depending on the experimental domain, this set of reference interpretations is called *ground truth* or *golden standard*, and the claims of published approaches generally consist in trying to minimize the discrepancy between it and the results produced by various algorithms. In what follows we will be using the term *ground truth* rather than *golden standard*.

Let us consider two algorithms,  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , trying to solve the same interpretation problem on data set  $\Delta$ , and producing results  $\mathcal{I}_{\mathcal{A}_1}$  and  $\mathcal{I}_{\mathcal{A}_2}$ :

$$\mathcal{A}_1(\Delta) = \mathcal{I}_{\mathcal{A}_1} = \{(\delta, \mathcal{A}_1(\delta))\}_{\delta \in \Delta}$$

$$\mathcal{A}_2(\Delta) = \mathcal{I}_{\mathcal{A}_2} = \{(\delta, \mathcal{A}_2(\delta))\}_{\delta \in \Delta}$$

These result sets are usually reported through *precision/recall* or *true/false positive/negative* matrices [van Rijsbergen, 1979]. Both, of course, are essentially the same, and capture a

distance measure between the produced interpretation sets and the reference set  $\mathcal{I}_{\mathcal{O}}$ . In our case, for instance, precision  $p$  and recall  $r$  could be expressed as

$$p = \frac{|\mathcal{I}_{\mathcal{A}_1} \cap \mathcal{I}_{\mathcal{O}}|}{|\mathcal{I}_{\mathcal{A}_1}|} \quad r = \frac{|\mathcal{I}_{\mathcal{A}_1} \cap \mathcal{I}_{\mathcal{O}}|}{|\mathcal{I}_{\mathcal{O}}|}$$

It is an incomplete measure, however, when it comes to comparing two algorithms to each other and using it as an element to characterize the improvement over the state-of-the-art. Figure 5.1 represents the results of both algorithms and the reference oracle as sets and their mutual intersections.

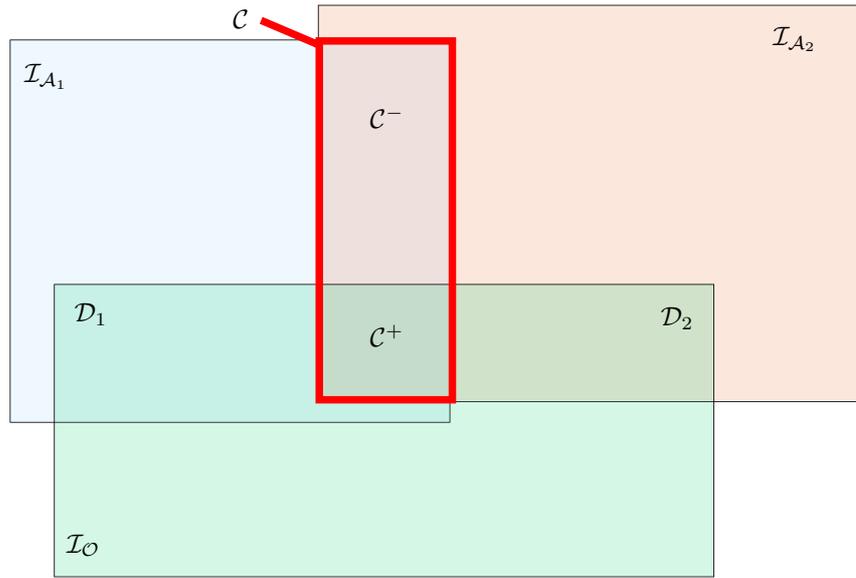


Figure 5.1: Various levels of possible overlapping of interpretation algorithm results

Where precision and recall measures mainly focus on the intersection area between two sets (and its coverage with respect to the rest of the set) it becomes clear, from the observation of the diagram, that similar scores for  $\mathcal{A}_1$  and  $\mathcal{A}_2$  may actually hide different situations, depending on how the various results overlap. We therefore introduce the following definitions:

**Consensus  $\mathcal{C}$**  (divided into *hard* consensus  $\mathcal{C}^+$  and *soft* consensus  $\mathcal{C}^-$ ) are the areas where two algorithms agree.

$$\mathcal{C} = \mathcal{C}^+ \cup \mathcal{C}^- = \mathcal{I}_{\mathcal{A}_1} \cap \mathcal{I}_{\mathcal{A}_2}$$

The hard consensus is where, moreover, they both agree with the oracle; the soft consensus is where they agree with each other, while in disagreement with the oracle:

$$\mathcal{C}^+ = \mathcal{I}_{\mathcal{A}_1} \cap \mathcal{I}_{\mathcal{A}_2} \cap \mathcal{I}_{\mathcal{O}}$$

$$\mathcal{C}^- = (\mathcal{I}_{\mathcal{A}_1} \cap \mathcal{I}_{\mathcal{A}_2}) \setminus \mathcal{I}_{\mathcal{O}}$$

**Disagreement** is when two algorithms provide different interpretations on the same input data.

**Differentiation**  $\mathcal{D}_i$  corresponds to  $(\mathcal{I}_{\mathcal{A}_i} \cap \mathcal{I}_{\mathcal{O}}) \setminus \mathcal{I}_{\mathcal{A}_j}$ . *I.e.* where both algorithms disagree, but where  $\mathcal{A}_i$  is in agreement with the oracle.

In order to understand what these fairly obvious and straightforward definitions provide us, we need to consider in some more detail how *ground truth* is commonly acquired, and how it is used to guide the development of new algorithms.

### 5.2.2 On the (Correct) Use of Ground Truth

By definition, *ground truth* consists of the correct expected outcome for a process conducting a defined task. It is largely used in Machine Perception to benchmark and evaluate algorithms and methods, under the assumption that, the more an algorithm's results approach the ground truth, the better it solves the underlying task. Hence the elements of discussion mentioned in the introduction.

There are a number of issues with this paradigm, however. While it is obviously right when the task at hand is to reproduce the required results on a known and fully outlined dataset, it is not clear how to

- evaluate or extrapolate the quality of results on data beyond the initial data,
- interpret the situations of disagreement or differentiation (*cf.* previous definitions) between partial solutions that do not entirely meet the mark,
- reuse and evaluate solutions for one set of ground truth in the context of other, similar experimental data sets.

These points are quite essential to what will follow later (*cf.* § 5.2.3, p. 151 and § 5.4, p. 164) and find their origin in the way ground truth is used in Machine Perception. One of the reasons ground truth has been rendered necessary is the fact that most addressed problems are formulated in an *ostensional* way, while algorithms and programs are essentially *intensional* formulations. “*Ostensional*” and “*intensional*” refer to WITTGENSTEIN’s semantic definition classes [Wittgenstein, 2001]. A concept is intensionally defined if one can define a set of sufficient and necessary properties that characterize its instances. It is ostensional if no such set of properties exists (or is known) but if one can enumerate representative examples and/or counter-examples of instances.

Many of the current trends in Machine Learning and *Big Data* (many of which are applied to Machine Perception) are approaches to convert ostensional concepts to intensional definitions. This is not a fundamentally new approach, and, as a matter of fact, is how Science is generally conducted: from Ockham’s razor [Ockham, 1323] to the *abduction* process coined by PEIRCE [Peirce, 1998]. For what will follow, it is interesting to elaborate

on the underlying assumptions and reasoning processes. Peirce’s abduction actually relies on “erroneous” reasoning when expressed in first-order logic, since it follows the

$$\text{Axiom: } A \Rightarrow B \quad (5.2)$$

$$\text{Observation: } B \quad (5.3)$$

$$\text{Conclusion: } A \quad (5.4)$$

In other terms, when knowing that  $A$  causes  $B$ , and without any other contextual knowledge, we are inclined to assume that, when we observe  $B$ ,  $A$  must have caused it, although we perfectly well know that  $\neg A \vee B$  and  $A \Rightarrow B$  are equivalent expressions, and that, therefore,  $B = \text{true}$  and  $A = \text{false}$  result in both expressions being **true**, thus formally invalidating the abduction with a counter-example.

This is, however, a purely formal and theoretical approach, in practice, abduction is used constantly, in a continuously improving iterative process  $\mathcal{G}$ :

1. observations  $B$  give rise to a set of hypotheses  $A$ ;
2. given no known counter-examples violate the hypotheses, we accept the hypotheses as a model for our observations;
3. subsequent observations  $B'$  either reinforce the model, or invalidate it (totally or partially) by providing counter-examples;
4. new hypotheses are emitted and the model is updated.

Transposing this generic framework to the topic that is of our concern: use of ground truth for Machine Perception, sheds another light on its role. The iterative process described above does not differentiate between specifically annotated or particular testing data: it just considers experimental evidence in support of or in opposition with a given model. This is not quite the way ground-truth is generally used for the evaluation of perception methods. The most commonly observed use reflects the following process  $\mathcal{P}$ :

- a) annotated ground-truth is provided,
- b) it is partitioned in training and testing data (according to the experimental evaluation protocols the partitioning schemes may vary, as may the availability of the testing data),
- c) perceptions methods are developed to conform to the training data,
- d) these approaches are applied to the testing data, and their adequacy to the ground-truth is measured.

When confronted to  $\mathcal{G}$ , we can make the following observations:

- The ground-truth used in  $\mathcal{P}$  can be assimilated to  $B \cup B'$  used in  $\mathcal{G}$ . Furthermore  $B$  is represented by the training data, and  $B'$  by the testing data.
- Step **c** in  $\mathcal{P}$  relates to steps **1–2** in  $\mathcal{G}$ .

One can also argue that, once the adequacy with the ground-truth has been measured in step **d** of  $\mathcal{P}$ , the natural progress of the state-of-the-art takes care of “updating” the model and reiterating as described in step **4** of  $\mathcal{G}$ . There are however several points on which both processes diverge.

First of all, step **2** in  $\mathcal{G}$  assumes  $A$  is in total adequacy with  $B$ <sup>18</sup>. This is not something that is always formally established in the case of  $\mathcal{P}$ . It is implicitly assumed, in most cases, that methods achieve the best possible performance on the training data  $B$ .

**Research Proposal 1:** It would be interesting to conduct a survey of existing benchmark and evaluation methods and initiatives that explicitly take into account the *a priori* performance on training data of the methods under consideration. This information is generally assumed uninteresting with regard to the performances on testing data, and is assumed to be near 100%. With respect to our thesis, it is essential to formally establish whether this is true under all circumstances, or identify those where it isn’t.

Second, the iterative process  $\mathcal{G}$  explicitly assumes that the series of observations  $B_i$  formed over all iterations  $i$  is strictly monotonous in the sense of inclusion:

$$B_o \subset B_1 \dots \subset B_i$$

Consequently, claims of performance increase or state-of-the-art improvement are only valid if they respect this criterium. This consideration is not necessarily absent in  $\mathcal{P}$ , but it is not always explicitly put forward all evaluation contexts.

This clearly echos with a keynote talk by G. NAGY at DAS 2010 [Nagy, 2010] in which the speaker actively defended the case that experimental validation data could be considered once and only once as new and “unknown”, claiming that subsequent use of the same data could have a polluting effect and introduce experimental bias, and should henceforth be assimilated to the available ground truth. This is exactly what the DAE platform (*cf.* § 5.4.3.0) defends also.

Third,  $\mathcal{G}$  is based on the explicit identification of counter-examples in  $B'$  invalidating the mode  $A$ , in order to improve the latter. To the best of our knowledge, no reports of explicit investigations on characterization or analysis on these counter-examples exist in the literature. They are necessarily taken into account in subsequent iterations of  $\mathcal{P}$  at some implicit level, but rarely in a clearly documented and reproducible way.

Finally, we have implicitly taken for granted, in the previous paragraphs, that the observations  $B$  – or their avatar: ground-truth – were unambiguously established, and that the subsequent derived models are indisputable interpretations if they perfectly cover

<sup>18</sup>In the general case, it often occurs that  $A$  doesn’t entirely cover  $B$ . In those cases, however, one identifies  $B^+$  (observations within  $B$  in accordance with  $A$ ) and  $B^-$  (observations within  $B$  not in accordance with  $A$ ) and shifts the research focus on  $B^-$ , thus returning to the configuration described above.

them. The next section will establish that ground-truth is not necessarily unambiguous. More, by essence, it necessarily carries a certain level of ambiguity. Further sections will then prove how this impacts overall performance analysis and measurement of perception algorithms in general.

### 5.2.3 On the Subjectivity of Ground Truth

As stated before, the need for ground truth comes from the necessity of measuring quality of models claiming to solve a given problem, and for which no formal description is yet available. In Machine Perception, most of the problems are defined in an ostensional way [Wittgenstein, 2001], meaning that there is no set of available or useful (sufficient and necessary) properties that describe the solutions of the perception problem. Most attempts to try and describe the problems eventually fall back on partial descriptions resorting to comparisons with what human perception would achieve under similar experimental conditions.

The search in Machine Perception to find algorithms performing as good as humans, is a tentative to express the ostensionally defined problems in an intensional way. The fundamental question remains how to measure whether the various available solutions in the state-of-the-art actually do “as good”, better or worse (than humans, or than competing solutions). The answer depends on the class of problems at hand:

- either the problem is ostensional, and exhaustively defined (also called *extensional* by WITTGENSTEIN); this means that all instances of the problem are represented by the ground truth, or in other terms, the ground truth entirely covers the problem domain;
- either the problem is ostensional, but the ground truth only partially covers the problem domain; one considers there exists an oracle that can assess output a posteriori;
- either the problem is intensional, in which case there is a set of sufficient and necessary properties that define the correct output, and thus is, by definition, already expressed as an algorithm.

The previous classification sheds an interesting light on the problem of data driven problem expression. Either the solution consists in writing an algorithm that produces the exact same output as the ground truth, on the same data, without any other further requirements or expected behavior beyond those limits (*i.e.* we don’t care about the output on other data). This is the case for extensional problems. Either the ground truth is a mere indication to the results the algorithm is supposed to produce in a broader context, on other data than the one presented (the ostensional, non exhaustive case). In the first case the solution is trivial: the algorithm simply needs to reproduce the same mapping, without any further “intelligence” or interpretation of the data. In this case, it is hardly appropriate to talk about Machine Perception, since the problem is more re-

lated to storing and retrieving pre-existing and referenced information. The second case is, of course, the most relevant situation, and by far the one most widely encountered in machine perception nowadays.

The main questions remain, however. Since we are trying to evaluate perception algorithms that are defined by a set of ground truth examples, and that this evaluation depends on the presence of an oracle, we need to know: “How to express the general interpretation problem of which the ground truth is an instance, as to produce the ground truth?” and “How to evaluate the behavior of the produced algorithms beyond the ground truth?”.

### Formalizing Ground-Truth Descriptions

As already mentioned, if the general interpretation problem could be expressed formally, the problem itself would very likely be solved, abstraction made of remaining decidability or tractability issues<sup>19</sup>. However, most Machine Perception problems are difficult to express in a formal way, and tentatives to do so have either long been abandoned or are notoriously limited or complex [Marr, 1982b, Biederman, 1987]. As hinted by in the introduction, they are most often described in terms of, or in comparison with human performance (“Find all zebras in a collection of pictures”, “Transcribe the handwritten text”, “Identify the language of the speaker”, “Segment a video into sequences and shots” ...) and accompanied by ground truthed data sets that are intended to define their scope of operation.

This section will elaborate on the reasons why it is difficult to formally establish the scope of a Machine Perception problem and that it necessarily carries a certain level of ambiguity that cannot be resolved.

Meanwhile, one can observe that this fundamental difficulty to express Machine Perception problems is one of the reasons for the success and breakthrough of machine learning approaches in this domain; statistical learning in particular. The implicit assumption and reasoning behind most machine learning approaches, applied to perception, is that, even though we have no means to formally describe a given machine perception problem, we can assume it can be solved in some way, and therefore there exists a mapping  $\mathcal{A}$  between the input data  $\delta \in \Delta$  and their possible interpretations  $\mathcal{I}$ . Many statistical approaches assume some local continuity properties exist (generally related to tolerance to noise and small deformations of the input) and that it is likely that  $\Delta|_i$  (*i.e.* the class of all data for which the problem should return  $i$ ) constitutes a sub-manifold of some kind, concluding that, in that case, If sufficient sampling points are chosen from  $\Delta|_i$ ,

---

<sup>19</sup>Indeed, being able to fully formalize a problem and its solution doesn’t make it *feasible*. For instance, the essence of modern public key cryptography consists in finding easily expressed formal interpretation problems that are of extreme complexity or intractable. *E.g.* whether a given number is the product of two large prime numbers, or to find the discrete logarithm of a random elliptic curve element [Hankerson et al., 2003]

appropriate numerical estimation techniques can provide a fairly accurate formalization of it.

The fact that these working hypotheses are based on the capacity of learning from valid sample data, brings us to the core of this section. How valid is the sample data? In a broader sense, the underlying question is whether the data is fully appropriate and sufficiently representative of the problem being addressed. Part of this question can be addressed from a purely statistical point of view and one can formally establish the limits of what can be inferred from measures by a variety of well known approaches. The question remains, however, whether the learning data is void of ambiguity and/or whether there is a potential overlap between what is considered as noise in the data, and what is due to a semantic shift.

For instance, in [Everingham et al., 2010], reporting on the Pascal Visual Object Classes (VOC) Challenge, a significant part of the publication consists in an account of the various conditions to acquiring, annotating and validating the data and refers to explicit annotation guidelines [Winn and Everingham, 2007]. There is no doubt that the effort of constructing the evaluation data for VOC and the benefit of making it available to a whole community has greatly contributed to creating a well defined reproducible evaluation environment for a complete research domain. However, the annotation guidelines contain descriptions of the data like “*Bounding box should contain all visible pixels, except where the bounding box would have to be made excessively large to include a few additional pixels (<5%)*”, “*Images which are poor quality (e.g. excessive motion blur) should be marked bad. However, poor illumination (e.g. objects in silhouette) should not count as poor quality unless objects cannot be recognised.*” ... as for the categories, “Bus” includes minibus, and “Car” includes cars, vans, people carriers *etc.* but should not be labeled when only the vehicle interior is shown. Obviously, images like the one in Fig. 5.2 fall in an ambiguous category both whether they should be labeled as “Car” (as van) or “Bus” (as minibus) on the one hand, and whether they *only* depict the interior of the vehicle.

This shows that it is premature to investigate whether some ground truth sampling (especially the training data) is representative and sufficient, and if it captures the complete scope of the interpretation context it is supposed to cover from an information theory, Shannon-Nyquist or linear algebra point of view, to name a few. While these remain essential and fundamental questions, they would obviously require a much larger and elaborate study than what can be reported, here. From here on, and for argument’s sake, we are going to assume that ground truth used in Machine Perception, is generally a sufficiently representative sampling for the intended interpretation context.

But what about the *interpretation context*, the set of rules, conditions and constraints that define whether a given interpretation  $i \in \mathcal{I}$  applies to some given input data  $\delta \in \Delta$ ? We previously associated this set to an oracle  $\mathcal{O}$ . We can reasonably assume there exists no known algorithm for  $\mathcal{O}$ , otherwise the corresponding Machine Perception problem would be solved (except, perhaps, for some performance issues). Although this sounds trivial, it actually leads to a very interesting paradox we are going to make explicit,

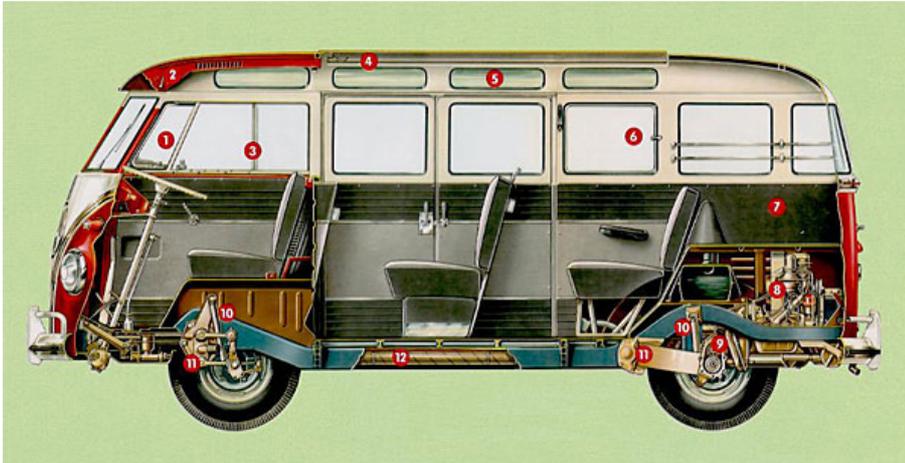


Figure 5.2: Example of ambiguous image category<sup>a</sup>

<sup>a</sup>Taken from <http://www.doobybrain.com/2008/01/30/car-cut-away-gallery/>

here.

### The Case of Human Annotated Ground Truth

The most common approach to generating ground truth is to use human annotators. In this configuration, the annotators serve as instances of the oracle  $\mathcal{O}$  and are provided with input data, for which they are to produce the corresponding interpretations, following clear instructions. These instructions correspond to the interpretation context and are defined as precisely as possible using both natural language and mathematically formalized criteria. The paradox arises immediately: either the instructions are totally unambiguous, and identically interpreted by all human annotators; either the instructions are ambiguous at some point, and may create legitimate different interpretations, depending on the annotators' viewpoints. Yet, totally unambiguous, fully formalized and totally reproducible instruction sets bear a name: algorithms. Hence, if the interpretation context can be formalized, the Machine Perception problem is solved. Consequently, in the case of human annotated ground truth, it is impossible to avoid a certain level (may it be minimal) of ambiguity, and therefore legitimate differences of interpretation will persist.

This is actually supported by many findings. [Smith, 2010] reports an experiment of pixel-level human annotation for document binarization, for instance, and we already addressed the case of the Pascal Visual Object Classes (VOC) Challenge, just before [Everingham et al., 2010, Winn and Everingham, 2007]. This is also in line with the distinctions made by WITTGENSTEIN concerning *intensional*, *extensional* and *ostensional* definitions for semantic classes, we already referred to before.

### The Case of Synthetic Data

Synthetically generated ground truth is the dual configuration of human annotated ground truth, with respect to interpretation context. Indeed, in this case, there exists an algorithm  $\mathcal{S}$  that is capable of generating data that is conforming to a given interpretation context. Formally speaking,

$$\begin{aligned} \mathcal{S} : \mathcal{I} \times \mathbf{P} &\rightarrow \Delta \\ i, p &\mapsto \delta \end{aligned} \tag{5.5}$$

where  $\mathbf{P}$  is the parameter space of  $\mathcal{S}$ . Under those conditions, trying to determine an algorithm for  $\mathcal{O}$  becomes an *Inverse Problem*, which is class of reputedly hard, ill-posed problems, introducing a high level of ambiguity [Tarantola, 2005] in the general case.

It is interesting to consider the cases where  $\mathcal{S}$  either is injective, surjective or bijective (other situations can, without loss of generality, be reduced to these three).

1.  $\mathcal{S}$  is injective (and not bijective): this means that the generated ground truth does not cover the entire set of possible interpretation configurations, and therefore is not an appropriate, nor a representative tool for performance evaluation<sup>20</sup>.

Given that  $\mathcal{S}$  can still be used for addressing a sub-part of the interpretation problem by restricting  $\mathcal{O}$  to  $\Delta' = \mathcal{S}(\mathcal{I}, \mathbf{P})$ , the derived use comes down to considering the surjective or bijective case.

2.  $\mathcal{S}$  surjective (and not bijective): this means that the interpretation problem is potentially ambiguous. If

$$\exists (i, p), (i', p') \in \mathcal{I} \times \mathbf{P} : i \neq i' \wedge \mathcal{S}(i, p) = \mathcal{S}(i', p')$$

then there is a  $\delta$  for which both interpretation  $i$  and  $i'$  hold<sup>21</sup>. However if, independently of any  $p, p'$

$$\forall i \neq i' \in \mathcal{I} : \mathcal{S}(i, p) \neq \mathcal{S}(i', p')$$

then the surjectivity is only due to an over-parametrization of the generative function, and has no impact on interpretation ambiguity. In that case the problem can be reduced, using an alternative  $\mathcal{S}'(\mathcal{I}, \mathbf{P}')$ , to the bijective case.

3.  $\mathcal{S}$  is bijective: in that case  $\mathcal{O} = \mathcal{S}^{-1}$ .

---

<sup>20</sup>This is a somewhat strong statement, and in many cases it can be helpful to use these functions anyway, as an instance of common practice in experimental research: “If we cannot immediately solve the global problem, let’s try and solve a more manageable sub-problem.”

<sup>21</sup>We are making the implicit assumption that interpretations are mutually exclusive. Although this may seem restrictive, it is not. In cases where multiple interpretations are acceptable, one can simply replace  $\mathcal{I}$  by  $\{0, 1\}^{|\mathcal{I}|}$ .

Besides the fact that most of the synthetic ground truth generating methods have not been categorized into one of the above classes, and that, consequently, performance evaluation based on their use cannot be considered totally reliable (if not seriously flawed) they introduce a similar paradox as in the previous case: either the problem is well posed ( $\mathcal{S}$  is bijective) but then it should be theoretically possible compute  $\mathcal{O}$  as  $\mathcal{S}^{-1}$  and the problem is solved by posing it; either the problem is ill-posed and any proposed solution will either be irrelevant ( $\mathcal{S}$  is injective) or non-unique or ambiguous ( $\mathcal{S}$  is surjective).

### Standoff

The infamous Semantic Gap is here to stay, and seems to be a fundamentally intrinsic part of interpretation: either one is capable of very precisely state an interpretation problem, in which case the mere fact of stating it lifts any possible ambiguity and consists in solving it; either the problem is open to interpretation, and multiple contradictory solutions may fit the problem.

This is not really surprising, and is in line with post-modernist philosophic considerations on truth and interpretation [Heidegger, 1967, Peirce, 1998]. While this does not mean that interpretation is impossible, it does conclude that multiple possible interpretations coexist and cannot be compared to one another. In the following section we shall be developing a set of computational tools to accommodate to this paradigm shift, very much in line with Eco's idea that only a limited number of all possible interpretations are worthwhile to consider [Eco et al., 1992, Eco, 2007].

### 5.2.4 Interpreting Ground Truth

Given the elements developed in the previous section, we need to reconsider the questions asked p. 151: "How to express the general interpretation problem of which the ground truth is an instance?" and "How to evaluate the behavior of the produced algorithms beyond the ground truth?", since the ground truth itself is open to interpretation, and thus to controversy.

We have started to address this problem in [8,49,47] but we can go further. Fig. 5.3 provides a graphical representation of the consensus and disagreement of two algorithms and an oracle, similar to the one in Fig. 5.1, but showing all possible configurations on ground truth data (center) and unlabeled data (periphery).

Before proceeding, it should be noted that, given the fact that ground truth is subject to controversy, there is no compelling reason to continue to use the concept of *oracle* anymore. From hereon we shall only be considering various interpretation results coming from different sources (without differentiating whether these sources are algorithms/methods or humans)<sup>22</sup>.

---

<sup>22</sup>This is a fundamental paradigm shift that opens opportunities to innovative new interactions with

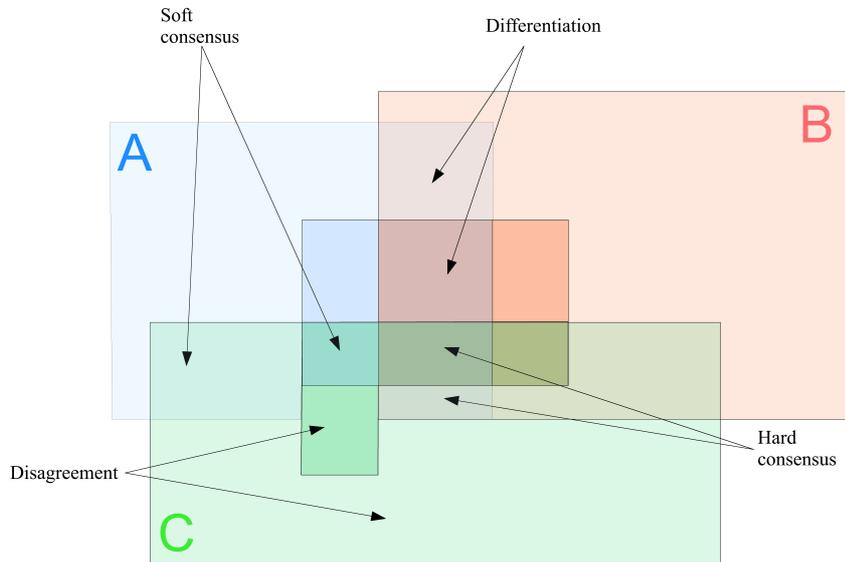


Figure 5.3: Various levels of possible overlapping of interpretation algorithm results with ground truth (center) and without

As a consequence, one can consider that evaluating the performance of a method uniquely with respect to its conformity to a specific source consists in verifying whether it has the exact same interpretation of the ground truth. By studying the data where consensus and differentiation occurs (either with respect to annotated ground truth or with respect to a broader scope of data) we may discover whether these differences in interpretation are legitimate or not (*i.e.* resulting from an acceptable, yet different interpretation of the ground truth or from an error in the method).

**Research Proposal 2:** The main idea is to try and capture the possible structure underpinning the consensus and differentiation areas of a set of competing algorithms on the same data. By using Formal Concept Analysis [Ganter and Wille, 1999, Ganter et al., 2005] on the one hand, and possibly statistical clustering techniques on the other hand, we expect to be able to characterize their differences in interpretation. The general idea is developed below. This proposal has also been formalized for funding by the CNRS in 2013 (unfortunately without success). Its description can be found p. 190.

other research domains. Considering any interpretation can potentially correspond to a valid context, it becomes necessary to try and evaluate confidence levels of these perceived contexts. This is something we have started to investigate with in our collaborations with Lehigh University (*cf.* PICS project proposal p. 197, and social networking concepts, developed p. 173).

	$M_1$				$M_2$				...	$M_m$			
	$\mathcal{A}_1$				$\mathcal{A}_2$					$\mathcal{A}_m$			
	$i_1$	$i_2$	...	$i_n$	$i_1$	$i_2$	...	$i_n$		$i_1$	$i_2$	...	$i_n$
$\delta_1$	0	1		0	1	0		0		0	1		0
$\delta_2$	1	0		0	0	0		1		0	1		0
$\vdots$													
$\delta_d$	0	1		0	1	0		0		0	0		0

Table 5.1: Representation matrix  $\overline{M}$  for FCA-based ground-truth interpretation and performance evaluation

We are assuming that the task at hand can be expressed as a discrete set of expected results  $\mathcal{I} = \{i_1 \dots i_n\}$ . If not, discretization techniques like those developed in [Coustaty et al., 2011] can be used or adapted to fit the specific kind of descriptors used.

In that case, as in [8,49], we consider  $m$  algorithms  $\{\mathcal{A}_k\}_{1\dots m}$  and a data set  $\Delta = \{\delta_1 \dots \delta_d\}$ . This allows us to construct a family of  $m \times d \times n$  matrices  $M_k$  such that

$$M_k(s, t) = \begin{cases} 1 & \text{iff } \mathcal{A}_k(\delta_s) = i_t \\ 0 & \text{otherwise} \end{cases}$$

Let the final matrix  $\overline{M}$  be the concatenation of the matrices  $M_k$  such that  $\overline{M} = [M_1 \dots M_m]$  as represented in Tab. 5.1.

By conducting a Formal Concept Analysis on these data, the appropriate clusters of coherent interpretations can be uncovered and compared with the “natural” concepts underpinning them. This will eventually result in a better understanding of how Machine Perception methods compare to one another in a more semantic sense.

An interesting side-effect associated to this approach, which we also already discovered in [8], is that it contains a duality between data and methods, in the sense that it cannot only be seen as a tool for comparing and studying different algorithms and methods, but that it can also be considered as a way to assess the appropriateness of data with respect to the methods. This property will be developed further in section 5.4.3, p. 169.

It should be clear to the reader, by now, that “ground-truth” is intrinsically ambiguous, and is that, at most, it is a representation of a non-formalized interpretation context, eventually connected to some level of human interpretation. The next section will therefore take a deeper look into the concepts of interpretation and analysis (as in “Document Image Analysis” and “Image Interpretation”).

## 5.3 Analysis and Interpretation

Often, “analysis” and “interpretation” are put the same level, as if they were necessarily related. Indeed analysis can be seen as the process resulting in an interpretation. It is important to spend some time describing the actual meaning of both words and to distinguish the differences between them before actually relating them to the more specific case of graphical documents. As it will become clear in the next sections, we consider “interpretation” as the result of an analysis, bearing a significant meaning in a specific, precisely and well defined context. The “analysis” itself is therefore a process, combining a wide range of knowledge sources and data processing tools defined by this context.

### 5.3.1 Interpretation

The concept of interpretation is open to controversy and debate and has a longstanding history of intense research in a wide area of domains, ranging from philosophy<sup>23</sup> and metaphysics [Wittgenstein, 2001, Eco, 2007] to linguistics and pattern recognition. It is important to bear in mind some fundamental discussion points related to more philosophical debates, since they have repercussions on whether certain kinds of interpretation are actually computationally feasible or decidable.

In theoretical computer science, the notion of interpretation is usually related to formal languages, logic and model theory, and consists in a mapping from a formal system into another domain. Without going into further detail, the main point to be retained is that this requires that there is some initial formalized model containing operators, functions, theorems, axioms or properties, operating on one or more formal domains. The act of interpretation consists in instantiating this formal model by mapping it to the real-world, in such a way that all its properties are preserved. Although there have been tentatives in the past to formalize interpretation of visual information, none of them have been as far as establishing the formal mapping of all properties of the formal domain in the perception domain or vice versa.

Since the current state of the art in Pattern Analysis and Machine Perception does not contain any documented formal study of interpretation, this section will adopt a more pragmatic viewpoint, by giving an overview of what general interpretation contexts are implicitly admitted in Graphics Analysis, how they condition the various kinds of resulting interpretations. Interpretation is the deliberate choice of associating a representation to a concept, or considering one concept as a representation of another concept. This link between syntax and semantics, or “low level” semantics to “higher level” semantics is to

---

<sup>23</sup> “Every word instantly becomes a concept precisely insofar as it is not supposed to serve as a reminder of the unique and entirely individual original experience to which it owes its origin; but rather, a word becomes a concept insofar as it simultaneously has to fit countless more or less similar cases — which means, purely and simply, cases which are never equal and thus altogether unequal” [Nietzsche, 1873]

be considered as arbitrary to a certain extent, in the sense that the relationship between one and the other is purely conventional and by no means unique or exclusive.

### 5.3.2 Context

The first essential element to point out is that interpretation is totally dependent on “context”. For example, determining if a set of pixels is to be interpreted as a straight line will depend on whether we are considering printed lines or handwritten lines. Formalizing the context is not something that comes naturally in the general domain of document image analysis and pattern recognition in general. It is often something that is partially expressed (but not formalized) as implicit domain knowledge, embedded within the analysis process when it is formalized as an algorithm, and sometimes partially represented by the input parameters of the process [Wittgenstein, 2001]. The level of tolerance to noise or deformation, is one of those items that contribute to the context, for instance, but it is far from being the only one.

### Interpretation of Images

In the context (no pun intended) of image interpretation, and more specifically document images, we can define three main classes of context, each of which give a rather different meaning to “interpretation”. Most of them relate to image and document analysis, but can easily be extended to other Machine Perception applications. These three classes, obviously, have fuzzy boundaries, but they do correspond to rather well established research problems and their associated communities.

1. The first category of contexts considers the image as a pixel matrix, resulting from a complex acquisition process. In that case, interpretations are mostly related to the image acquisition process and its effects on the produced graphics. Most of these issues have been addressed in Chapter 2, and, generally speaking, they fall in the domain of general image analysis. They are very often only related to graphics analysis by the fact that higher level interpretations may depend on the interpretations resulting from these contexts. These interpretations are blur estimation, grid calibration, image restoration, binarization, *etc.*

There are however contexts where explicit knowledge about the origin of the documents under consideration, and the fact that they contain specific graphical elements, are needed. Examples are [Drevin, 2011, Oliveira et al., 2011].

In [Drevin, 2011], for instance, the image acquisition process is quite unusual, compared to what is generally observed in the (graphical or not) document analysis domain. While the underlying interpretation goal relates to legacy cosmic ray recordings, produced with Carnegie Type C Ionization Chambers in 1942, the images themselves are the result of a complex acquisition pipeline. As described in

the paper “The Carnegie Type C Ionization Chambers [...] were designed and built for the purpose of the continuous recording of cosmic rays. Essentially the ionization chamber was a steel sphere containing purified argon at a pressure of 50. The argon was ionized as cosmic rays passed through the chamber. A Lindemann electrometer was used to measure the ionization current. To record the ionization level due to cosmic rays passing through the chamber, the shadow of the electrometer needle was projected onto a continuously moving strip of photographic paper. Furthermore, the barometric pressure and the temperature of the cosmic-ray meter could also be recorded on the same strip of photographic paper. Every hour the ionization chamber was grounded for 3 minutes, zeroing the ionization current and therefore bringing the electrometer needle back to the zero position. At the same time the lamp of the recorder was dimmed resulting in hourly vertical lines.” [Drevin, 2011] furthermore, “The photographic paper recordings [...] were scanned at a resolution of 600 dpi and 8 bits per pixel (256 gray levels) using an Epson Perfection 4490 Photo scanner.” (*ibid.*)

This whole acquisition context is essential for the analysis process to succeed and achieve usable interpretations. This is not only true for the automated processes we’re mainly concerned with, but it is equally important for the human interpretations, since the acquisition process introduces constraints on shapes and creates artifacts that could be interpreted as informational, while they are not, *etc.*

As a summary, we can consider these contexts as being very much related to the broader domain of Signal Processing and filtering in which the notion of interpretation becomes the one of interpreting a noisy, raw signal, as an instance of a specific ideal one, having undergone a series of perturbations. Binarization (taken in its very broadest sense: separating signal from background) is one emblematic case of the interpretations falling in this class.

2. The second category of contexts consists of considering graphical documents as supports for a message designed to be interpreted. These contexts suppose the graphical representations are part of a visual language that has a reasonably formal grammar associated to it, and as such, imply that the meaning of the message is fully embedded in the document and that the knowledge of the language in which it was expressed is sufficient to recover its meaning.

This is usually considered as the core domain of graphical document analysis. It is important, however not to restrict interpretations solely to high level interpretations as those related to architectural floor plans, mechanical blueprints or electrical wiring diagrams. Lower level segmentation interpretations also fall into this category: determining whether an alignment of pixels should be considered as a line segment, a circular arc, or whether disconnected components actually form a higher level dashed line, should also be considered as interpretations. The fact that segmentation issues are actually related to interpretation can be easily illustrated by the example of straight line detection. In function of the context: on-line hand-

written stroke analysis, off-line printed material, flatbed scanned or hand held high distortion capture devices, the interpretation of “identical” sets of pixels will be completely different.

The common ground between all these interpretations is that their context can be captured (although not necessarily fully formalized) and can often be accurately approximated by an average human interpreter, at the view of a limited set of documents. In that sense, the context and documents are self-contained. With respect to the context developed in item 1, just before, the image acquisition process and its associated perturbations are not an issue here. While it is important to stress that image acquisition artifacts can have a significant impact on the difficulty of treating the kinds of documents under consideration (those expressing a “language”), the analysis process and associated interpretation still contain a serious number of hard problems, even when handling perfectly noiseless documents (in the sense of their acquisition process).

Segmentation, for instance, can be seen in the light of this class of contexts, as well as Symbol recognition. However, the largest and most representative class of problems falling into this context are those related to the interpretation of technical drawings (mechanical blueprints, electrical wiring diagrams, architectural floor plans . . .). What characterizes all of these examples is that they call upon the following conceptual process: the intention of an author to represent a concept from a commonly agreed upon collection of concepts (semantics), followed by the expression of its representation, following a commonly agreed upon convention (syntax), results in a physical representation of this expression (document). The transcription of the representation onto the document may or may not be subject to noise and perturbations. The essence of the analysis process and the resulting interpretation is twofold:

- (a) Given a document, and given the appropriate context (syntax, semantics and the conventions relating them to each other), what was the concept that the author initially intended to convey?
- (b) If the appropriate context is lost or incomplete, but if sufficient data are available, are there means to approach the context by computational discovery or learning? A dual problem to this is when different intentions can give rise to similar or identical syntactical expressions.

Some possible illustrations, in the context of the examples given before (segmentation, symbol recognition, technical drawing interpretation) are developed here.

In the case of line segmentation for instance, the analysis process will consist in trying to recover the author’s intention to produce a drawing with a set of geometric primitives (say, straight lines and circular arcs) connected to each other or related by topological or geometrical relations. Regardless of issues related to the transcription (noise and deformations, for instance) that may interfere with this

process, the interpretation still requires knowledge about the context in which the author considers a line as straight or curved, how precise geometric constraints (incidence, perpendicularity . . . ) are represented or even when lines or strokes are significant for interpretation, and when they are just clutter without any meaning.

Symbol recognition and its relation to interpretation is another interesting study case. Let us first consider symbol recognition in its most common form: symbols are considered as isolated visual elements in a complex scene. It doesn't matter, from an interpretative point of view, whether the shapes are segmented out of their visual environment or still embedded within the scene. The questions that need to be answered are of a lexicographical kind: "What symbol did the author intend to represent" and related to the interpretation context: "What deformations of the symbol are deemed acceptable".

As a side note, it should be obvious to the reader that the above issue does not involve any semantics or meaning, notwithstanding many references in the state-of-the-art that pretend it does. Semantics and meaning start having an influence on interpretation in the case where there is a level of ambiguity on how to interpret isolated visual information. Interpretation of technical drawings are an example of that. Elementary visual information (say, a foreground pixel) should be interpreted differently when part of a textual element, a logo, a connection line or a symbol element, since the previously mentioned segmentation questions make sense for large, linear, visual structures, but not for textual parts or logos. Similarly doorways make sense when they are within a wall section, but their corresponding symbols may also appear in isolated form, within the context of a thesaurus of used parts, or a legend, for instance. In a same document, perfectly identical visual items may therefore lead to completely different interpretations given their context.

Still, given these different interpretation contexts (and completely open and unsolved problems they raise for automated interpretation), the common factor between them is that they remain a message, expressed in a particular language, produced by an author having a specific intention. Whether the language can be explicitly formalized, or if the author's intention can be fully recovered is still open to debate. It is, however, commonly admitted, that these documents have an "obvious" ("natural" ?) interpretation context for a reasonably knowledgeable human interpreter. Something that is not necessarily true for the next category of interpretation contexts.

3. The last category of contexts aims for interpretations that lie beyond the document image and for which the latter is just an element. In these contexts, the interpretation requires knowledge that is not contained in the document under consideration.

Examples are, for instance, the interpretation of drop caps and illuminations in medieval or Renaissance documents, that, given the correct interpretation, can provide valuable historical information on the origin, author or intended public of the document. Interpretation of certain kinds of forensic data also falls into

this category. In this case, the question whether the “producer”<sup>24</sup> of the document intended to deliberately convey a message or not is only secondary. What is of importance in this context of interpretation are the signs and indications related to how the document was produced and what visual information relative to the way the document was shaped can be interpreted with respect to the context in which the producer was making the document.

This requires specific knowledge that is not naturally available from the document. Signature forgery detection, for instance is one of these cases. More graphics related problems are those having been addressed in some historical document analysis challenges, in which historians are trying to track uses of drop cap print blocks over a series of documents, in order to study commercial and technological exchange patterns in Renaissance Western Europe<sup>25</sup>. Indeed, in that time, wooden print block were used by one print editor, and then exchanged with other editors ... over time the print blocks accumulate wear and tear that can be observed through the printing quality of the corresponding drop caps. Carefully analyzing the printed documents, combined with information about their origin of can lead to interpretations of possible connexions and exchanges between printers, *etc.*

The key factor here is that the contextual information needed to obtain a plausible interpretation calls upon knowledge that is unavailable within the document itself and needs to be contributed and combined with other sources. From a computational or knowledge-representational view, solving these issues clearly relates to Artificial Intelligence themes, as much as it requires Pattern Recognition and Image Analysis knowledge.

### 5.3.3 Analysis

Analysis is the process aiming for the interpretation. It needs to capture the interpretation context and combine knowledge sources and processing in an appropriate way to achieve that goal. From that viewpoint, analysis undeniably has a strong engineering component to it. In a very strict sense, analysis can be defined as a sequence of decisions and operations, based on raw input and contextual knowledge that eventually results in the interpretation of the raw data. For human analysis, this would consist in perception and reasoning through brain activity. In our case, which is more generally the one of machine perception, analysis consists in the execution of an algorithm, which is far more constrained.

The execution of an algorithm depends on three, and only three parameters:

---

<sup>24</sup>We insist on making a distinction between “author” and “producer”. The author is producing a document with the explicit and conscious intent of making it the way it is. The “producer” just happened to produce a document (not necessarily consciously, or trying to control the way it was produced). The term “document” refers to whatever physical support is under consideration for interpretation.

<sup>25</sup>This was actually one of the driving end-user goals of the ANR Navidomass (<http://navidomass.univ-lr.fr>) project we were involved in during 2007-2010.

1. its instruction sequence or program,
2. its input data,
3. the limits of its execution environment.

The current technological state-of-the-art considers that the first and last item of this list should be considered as rigidly fixed during the execution of an analysis process. The only parameters that can vary are the input data. With respect to what was mentioned previously concerning the context of interpretation, this has quite a significant influence on how analysis software performs with respect to human analysis. This means that either the interpretation context is fixed (when it is embedded in the program) or that it needs to be explicitly formalized as input data, or a combination of both. The various existing learning and classification techniques mentioned previously are not an exception to this, since they consist, in some sense, in a preliminary stage through which part of the context is learned, and then fed as input data to the analysis process. Even when considering the analysis process as integrating this learning phase, the initial assertion remains true, since in that case, one considers that the interpretation context is formalized as the set of learning samples provided as input.

This means that, in the light of all previous sections that constructing a generic analysis algorithm leads to an impasse. On the one hand, we have shown in § 5.2 that ground truth, and therefore the definition of a perception problem is inherently ambiguous, and dependent on the interpretation context. On the other hand, we have come to the conclusion here, that, to achieve flexible and generic analysis algorithms, this interpretation context needs to be formalized to some extent.

The following sections will draw some plans and possible directions to how to approach and model interpretation contexts.

## 5.4 Modeling Contexts and Interpretations

In the previous sections we have been addressing the relationship of context with interpretation. We have tried to establish a relation with the post-modernist philosophical theses of infinity of possible interpretations. While, on the one hand, it is comforting to (re)discover that many of the current apparent limitations or hurdles to what is commonly called the *Semantic Gap* [Smeulders et al., 2000] have very profound origins (as we have tried further investigate in the proposal p. 193), the constant progress of Machine Perception and the need for computationally efficient solutions to interpretation problems are strong incentives to both understand the limiting mechanisms and to investigate means to leverage solutions.

The fact that Machine Perception entirely relies on computational resources (compared to the more philosophical and anthropo-linguistic considerations of interpretation [Eco, 2007]) is going to be a great help in these investigations. This chapter will explore pathways

to combine existing tools that have yet not been used or investigated in the context of artificial perception and interpretation.

### 5.4.1 Interpretation is Undecidable

Up to now, we have been passing under silence, the fact that formal interpretation and semantics have been the focus of thorough investigation since A. TURING, and have given rise to extensive research domains investigating the limitations and properties of semantics of formal languages.

We are not going to develop a complete survey of existing work on syntax and semantics on a mathematical level. However, it is important to relate all previous sections to the fact that they are, in fact, instances of RICE's theorem [Rice, 1953]. RICE's theorem is described in [Jones, 1997] as follows:

Rice's theorem shows that the unsolvability of the halting problem is far from a unique phenomenon; in fact, all *nontrivial extensional* program properties are undecidable.

#### Definition 5.4.1

1. A program property  $A$  is a subset of WHILE-programs.
2. A program property  $A$  is *non-trivial* if  $\{\} \neq A \neq \text{WHILE-programs}$ .
3. A program property  $A$  is *extensional* if for all  $p, q \in \text{WHILE-programs}$  such that  $\llbracket p \rrbracket = \llbracket q \rrbracket$  it holds that  $p \in A$  if and only if  $q \in A$ .

In other words, a program property is specified by dividing the world of all programs into two parts: those which have the property, and those which do not. A non-trivial program property is one that is satisfied by at least one, but not all, programs. An extensional program property depends exclusively on the program's input-output behaviour, and so is independent of its appearance, size, running time or other so-called *intensional* characteristics.

An example property of program  $p$  is the following: is  $\llbracket p \rrbracket(\text{nil}) = \text{nil}$ ? This is *extensional*, since  $\llbracket p \rrbracket = \llbracket q \rrbracket$  implies that  $\llbracket p \rrbracket(\text{nil}) = \text{nil}$  if and only if  $\llbracket q \rrbracket(\text{nil}) = \text{nil}$  [...]

**Theorem 5.4.2** If  $A$  is an extensional and nontrivial program property, then  $A$  is undecidable.

This is exactly the context we have been describing throughout the rest of this chapter, and RICE's definition of *extensional* and *intensional* coincide with WITTGENSTEIN's, given p. 150. The bottom line is that if we make the assumption that the human output

of an interpretation problem can be represented by an algorithm, than it is undecidable, in the general case, to determine whether another algorithm falls in the same equivalence class defined by property of sharing the same output.

Hence, one can conclude that trying to measure analysis and interpretation in an absolute and univocal way is vain at best, and wrong in any case, and that pursuing in trying to obtain a formal and comprehensive description of an interpretation problem makes no sense.

What can we do about this? First of all, not panic [Adams, 1979] ... RICE's theorem applies, but to the general case, for any Turing machine and for infinite countable input sets. There are a wide variety of cases in Machine Perception where it is possible to reduce this set of constraints. The following sections will provide an overview of possible ways around the theoretical limitations set by RICE.

### 5.4.2 Interpretation as a Computational Problem

The first step in the direction of the proposal described previously is to continue considering interpretations as the result of an algorithm conducting the analysis process<sup>26</sup>.

Let us define an interpretation task  $\mathcal{A}$  depending on  $\delta \in \Delta$ , the input data to be interpreted,  $\gamma \in \Gamma$  the interpretation context (including a set of possible interpretations  $\mathcal{I}$ ), and producing  $i \in \mathcal{I}$ , interpretation of  $\delta$ . This gives us a family of functions  $\mathcal{A}_{\mathcal{I}}$  parametrized by  $\mathcal{I}$

$$\begin{aligned} \mathcal{A}_{\mathcal{I}} : \Delta \times \Gamma_{\mathcal{I}} &\rightarrow \mathcal{I} \\ (\delta, \gamma) &\mapsto i \end{aligned} \tag{5.6}$$

The first remark to be made here is that this is a family of parametrized functions over the set of interpretations  $\mathcal{I}$  and that, consequently the interpretation context  $\gamma$  itself is parametrized by the set of allowed interpretation concepts.

#### Consequences:

- This means that, if an analysis process is modeled as a function, and, subsequently, as an algorithm, the set of allowable interpretations  $\mathcal{I}$  is an external parameter of the problem at hand.
- The sets  $\Delta$ ,  $\mathcal{I}$  and  $\Gamma_{\mathcal{I}}$  need to be clearly defined.
- The chosen interpretation context, being an argument to  $\mathcal{A}$ , can influence on the final interpretation  $i$ , but is constrained by the fact that the latter should be within  $\mathcal{I}$ .

---

<sup>26</sup>[Jones, 1997] insists on the difference between a *function* and a *program* when considering decidability issues. We shall make this distinction here, for ease of reading

**Definition of  $\mathcal{I}$** 

$\mathcal{I}$  is the *a priori* given set of allowable interpretations: it can range from a simple discrete list of concepts to a more complex grammar of concepts and relations between them.

it does not seem to make much sense to consider the possibility to have interpretations in a continuously valuated domain. Interpretation seems somehow related to a discrete decision process, rather than a numeric evaluation. Even in the context of, for instance, recognition and computation of mathematical expressions,  $\pi$  or  $\sqrt{2}$  are, taken as concepts, discrete entities, despite the fact they are a transcendental real numbers.

However, one of the main points to consider, is that, given the model we expressed in (5.6), the analysis process is only required to express its result in terms of  $\mathcal{I}$ , the way it actually does, and how it does is conditioned by the actual interpretation context  $\gamma$ . This means that nothing in the model is opposing the following situation:

- $\mathcal{I}$  defines the set of allowable concepts. *e.g.*  $\mathcal{I} = \{circle, triangle, square\}$
- The interpretation context  $\gamma$  provides “rules” to associate these concepts to the input data and validate their existence, without necessarily fitting to conventional semantics<sup>27</sup>. For instance

$$\begin{aligned} \bigcirc & \models triangle \\ \triangle & \models square \\ \square & \models circle \end{aligned}$$

We purposefully took a non-conventional example to make a point that will allow us to introduce the notions of ambiguity later. In the example, the interpretation terms clearly do not fit the usual labels we would attach to the depicted shapes. It therefore allows us to focus on three levels of information that will allow us to better define the concepts we are interested in: on one side of the spectrum are the *semiotics*, defined by the set of tokens bearing a signification, in the particular context  $\gamma$ ; on the other end of the spectrum are the concepts  $\mathcal{I}$ . In between is a fuzzy area describing the semantics related to the interpretation of tokens<sup>28</sup> into concepts.

In our particular case, the concepts could as well have been labeled  $A, B, C$ . As long as  $\gamma$  expresses the fact that “ $A$  is the locus of points equidistant to a given point” or “ $B$  is a geometric figure composed by three connected points”, for instance, or any other property useful to the considered interpretation context.

<sup>27</sup>*Semantics* is one of those terms the semantics of which are either ill defined, since taken for granted, or defined in a very specific mathematical context. This self-reference is an attempt to stress the difficulty to express what semantics actually are in the case of Machine Perception. The interested reader can refer to the final chapters of [Eco, 2007] to get an interesting, generic classification of “layers” of semantic interpretation.

<sup>28</sup>The term *token* is to be taken as *distinct perceived object* in a very broad sense and not as a human created lexical item as is usually the case in *Symbol Recognition*.

**Corollary 1** *the role of  $\mathcal{I}$  can be nothing more than syntactic labeling and cannot be assumed to be carrying intrinsic semantics. This is due to the fact that the interpretation context  $\gamma \in \Gamma_{\mathcal{I}}$  is responsible for defining the semantics attached to interpreting  $\delta$  as  $i$ .*

As a consequence, there is no use in making further assumptions concerning  $\mathcal{I}$  other than being a subset of  $\mathcal{N}$  (since we're operating in a computational context, any data set can be reduced to a set binary strings). In many cases,  $\mathcal{I}$  is finite.

### Definition of $\Gamma_{\mathcal{I}}$

The main difficulty thus becomes to correctly capture the interpretation context, and to find a way to describe what it means to decide that an image depicts a *circle*, for instance [12,28,51].

Instead of trying to formalize the context and essentially reformulating what we have enunciated before, we are proposing another approach, related to the use of *consensus* and *differentiation* developed in Proposal 2.  $\Gamma_{\mathcal{I}}$  thus gets characterized with respect to other contexts with which it agrees or differentiates on specific data.

The ideas we are developing to achieve this will be described in the next section. The main driving idea is that, given the fact there is no complete and coherent way to express interpretation contexts in the general case, we shall try to express how interpretation contexts agree and disagree on specific data sets, in the hope that we can uncover cases where agreement occurs on “important” data, and disagreement on less important data, as to provide efficient tools that are usable in real-world situations.

## 5.4.3 Interpretations as Data Mining

### Context and Semantics as Algorithms

Algorithms have very precise semantics, as defined by their execution interpretation context, and in the light of the analysis of the previous sections most of their interpretation context  $\Gamma_{\mathcal{I}}$  is embedded within their code and structure, and generally speaking, as far as most Machine Perception algorithms are concerned, they only rely on a very limited set of input parameters that allow to parametrize  $\Gamma_{\mathcal{I}}$  at the margin.

Recent<sup>29</sup> trends towards the use of learning, semi- or unsupervised classification approaches, *etc.* do not fundamentally change this paradigm, in the sense that the learning phase is done *before* the actual use of the program. Since their parametrization is very often beyond human intuitive or perceptual understanding, it is safe to consider them as being part of the program itself.

---

<sup>29</sup> a decade

In the previous sections, we have implicitly admitted that a written program, taking  $\delta \in \Delta$  as input and providing  $i \in \mathcal{I}$  as output fully characterizes the interpretation context  $\Gamma_{\mathcal{I}}$  of the interpretation problem it is addressing, then, as per Turing-Church thesis, it is equivalent to any other formalization of the same interpretation context. However, given two formalizations of presumably the same algorithm, it is undecidable, in general, to formally prove and establish their equivalence. We have therefore concluded that, for Machine Perception and interpretation problems, it is impossible to totally characterize (and therefore compare) the interpretation contexts, and thus establish how their performance can be ranked with respect to some hypothetical benchmark.

**Research Proposal 3:** The question that we intend to answer is: if it is computationally too difficult, intractable or undecidable to characterize the interpretation context of either a problem, an algorithm or a set of programs in a formal way, is it possible to at least find a description (possibly incomplete) of the context that they share, as well as a description of the context where they differ from each other ?

This brings us back to the points made in Proposal 2, where we suggested ways to identify concepts in terms of interpretation categories. In the next section we are going to look into how these concept categories can actually be used to guide users to the data they may be interested in, given the particular interpretation context they are in.

### Characterization of Interpretation Data

While we have insisted in the previous sections on the ambiguity of interpretation contexts, and the fact that unique ground truth is an illusion, trying to characterize both data and algorithms with respect to their shared interpretation context and their differences in interpretation context is not sufficient. We initially made the assumption, in Proposal 2, that the considered algorithms were operating on the same set of possible interpretations  $\mathcal{I}$ , the example in Fig. 5.4 shows that this need not necessarily be the case. It shows three algorithms, classifying the same data into either striped *vs.* non-striped animals, zebras *vs.* others or horselike *vs.* other animals. Contrarily to what was described in Proposal 2, their presumed classification domains, which can be assimilated to  $\mathcal{I}$ , are not identical. Arguably, before they were not either, but at least they shared the same lexical expression through  $\mathcal{I}$ . A more extreme (and funnier) example is given in Fig. 5.5.

The question therefore becomes in how far it is possible to represent differences in context, without being confronted to tractability and decidability problems, and to detect what the most appropriate or most likely context should be applied to a given situation. This seems to be a very complex problem. As hinted in parts of this document, it is confronted to some very fundamental theoretical limitations on the one hand, and has no solid existing experimental ground on the other.

Indeed, if the focus of this document has mainly been on performance analysis and re-

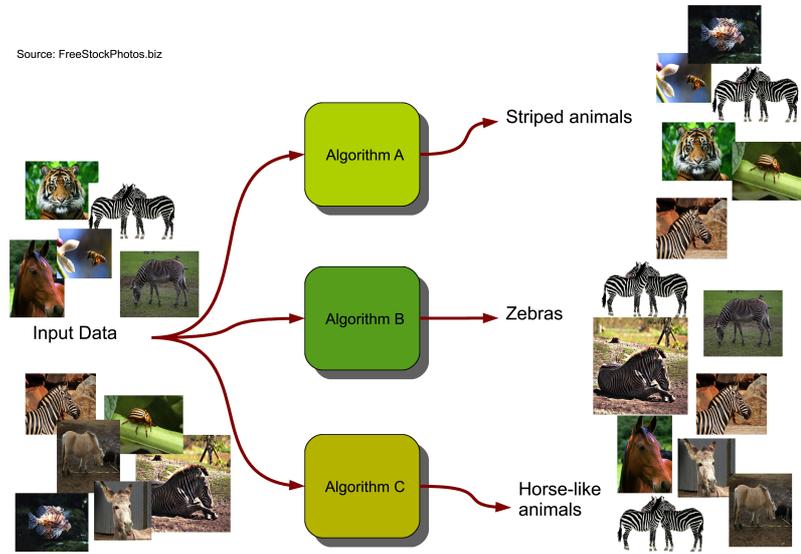


Figure 5.4: Situation where FCA can infer semantic hierarchies from perception algorithms

search evaluation, the main objective of Machine Perception and interpretation is to transform the perceptual data in information that is relevant to specific needs in specific contexts. The following section will therefore explore some possible extensions of preliminary existing tools we have developed, and which, eventually could open challenging new directions for Machine Perception applications.

### Tools and New Ways of Using them

The rest of this document will look back to the contributions we have recently made to the domain of performance evaluation for document analysis systems, and how they fit in the larger picture drawn in the preceding sections of this thesis. Their pursuit into creative new directions will provide new ways of conceiving interpretations and their interactions, and will largely call upon approaches that, while having never been applied to document analysis or machine perception in general, have been developed in other research areas.

#### The DAE Platform

We have started developing a paradigm for reproducible and traceable experimental research in collaboration with Lehigh University<sup>30</sup> [21,23,24,40,41,42]. While referring the interested reader to the corresponding articles, there are a number of

<sup>30</sup><http://dae.cse.lehigh.edu>



Figure 5.5: Joke about star rating (source: <http://imgs.xkcd.com/comics/tornadoguard.png>) or how interpretations on identical data, but using different contexts can make a huge difference.

features and properties related to it that have particular importance to the developments made before.

- It hosts algorithms (or, more precisely, implementations of algorithms) and records all their interactions with data, storing input parameters, results, who has invoked them and when.
- It hosts reference data (mainly targeted to document image analysis).
- It hosts interpretations and annotations of reference data (similar to *ground-truth*) that can be typed by users, and do not need to be unique.

This makes this platform a nice experimental environment to explore all ideas expressed before. Since it stores full provenance information and can retrace every bit of information to its origins (if produced by an algorithm it can retrace its full creation pedigree). Furthermore SQL querying can extract this information and combine it with other data on request.

However, one of its features: free and open annotation of data; can both be consid-

ered an advantage as a disadvantage. While it allows all kinds of different annotations of data (either by human annotators as by considering results of algorithms as annotations) it is a great tool for capturing different interpretation contexts. However, it also creates the possibility to have a high degree of polysemy within the annotations, exactly as we described earlier.

This platform is an essential experimental toolbox for many of the described research proposals and forms the backbone for the validation of most of the ideas expressed in this document. We are going to continue to investigate the following topics, related to this platform:

1. Its use as mutualized resource for results and annotations on shared data allows to test the FCA-based algorithms described in § 5.2.4. This work has started in 2012 with N. DROT, and is currently ongoing, notably with the proposals in annex A.
2. To achieve a critical mass of data for attaining the previous point, however, the platform needs to be sufficiently adopted by multiple users, and have proven its robustness and scalability. We have been studying adequate data and storage models using NoSQL (work in 2013, done with V. KELLER and A. TRY), on the one side, and have been promoting its use for international contests on the other side (ICDAR 2011 [16], GREC 2011 [Al-Khaffaf et al., 2010],[10], GREC 2013, as well as work in 2013 with L. DELADIENNÉE and M. WAJNBERG on the OpenHart 2013 contest).

#### Inferring Semantic Relatedness

One of the strengths of the aforementioned platform is its capacity to address a wide variety of annotations and interpretations of experimental data. The downside of this approach is that interpretation data can become very unstructured. This is exactly the point made, although in slightly other terms, in § 5.4.3.0.

We have therefore started investigating how the observation of all provenance data stored in the platform can lead to knowledge discovery, and more specifically whether either algorithms or data items are semantically related. To achieve this, we have started modeling the whole information acquisition process and the provenance data with a specific ontology in OWL [Knublauch et al., 2004].

Description logic reasoners like Fact++ [Dmitry and Ian, 2006] or HermiT [Motik et al., 2009] are capable of using the knowledge that specific data items have served as input or output parameters of a given algorithm, and therefore, that they (per semantics of those algorithms) share common properties. Combining this with related ideas in [Arévalo et al., 2010] may help us characterize common interpretation contexts or semantic relatedness.

This work is currently under investigation in collaboration with J. PRUVOST (2012) and S ZHENG (2013), and also directly relates to the FCA-based algorithms de-

scribed in § 5.2.4 and the proposals in annex A.

We also intend to pursue a very interesting idea, used in linguistics [Victorri, 1994b, Victorri, 1994a, Venant and Victorri, 2007] consisting in considering interpretations in a “metric” space (or in a partially ordered sense). In the light of the techniques and proposals suggested further in this chapter, it might be interesting, given the partial order constructed by FCA lattices to try and express “distances” between concepts or interpretation contexts.

#### Crowd-Sourced Ground Truth Evaluation and Social Networks

Similarly, since the platform is completely open to any kind of annotation, we have started to look into the analysis of social data to see if it is possible to extract context information that would be useful for either data categorization, quality assessment or interpretation context description. The driving reason behind this is that data annotations or interpretations, and especially their interpretation context, can benefit from weighing their value based on confidence and trust, extracted from who produced what data. This is the central goal of the proposal described p. 197.

Indeed, it is possible to extract (in our case, by setting up a Vivo<sup>31</sup> instance and populating it with bibliographic data coming from Google Scholar or DBLP) research domain knowledge from users’ scientific track record. This information can then be used to evaluate or classify annotations made on the DAE platform.

This work has been initiated in collaboration with a second year Master student, X. CAO in 2012 and will continue in collaboration with Lehigh University.

## 5.5 Related Initiatives

The problems and interrogations that were approached and sketched in this chapter are far from being solely related to machine perception interpretation evaluation, and many of the topics find their equivalent in other domains. Trying to encompass them all would probably require an extremely extensive review effort that would span across a large number of scientific areas. This section will be more of an unordered, seemingly random list of initiatives that were encountered during our intellectual errands... a part of them have already been enumerated in § 3.3.1, p. 45.

## 5.6 Conclusions

We have outlined and defined a number of new looks to Machine Perception and related interpretation problems. We have shown that automated interpretation of perception

---

<sup>31</sup>[www.vivoweb.org](http://www.vivoweb.org)

data is a very ill posed problem in the current state-of-the art. We have also shown that this is very likely because of the intrinsic intractability of interpretation itself.

We have given a number of formal representations of the interpretation problem, and have shown that only partial solutions can be found. These solutions are further developed and show that there exist a significant possible convergence with existing data mining and information discovery domains, in a way that has yet to be explored and formalized. This requires a significant amount of cross-domain interactions, and combinations of techniques and research problems coming from different communities: pattern analysis, knowledge discovery, formal semantics, semantic web and databases.

Once done, the resulting work will have a very positive impact on the measurement of performance comparisons, benchmarking and evaluation of Machine Perception algorithms, on the one hand, and may open new directions in the domains of semantic information retrieval on the other.

# Bibliography

- [DBL, 2003] (2003). *9th IEEE International Conference on Computer Vision (ICCV 2003), 14-17 October 2003, Nice, France*. IEEE Computer Society.
- [DBL, 2011] (2011). *2011 International Conference on Document Analysis and Recognition, ICDAR 2011, Beijing, China, September 18-21, 2011*. IEEE.
- [Adams, 1979] Adams, D. (1979). *The hitch hiker's guide to the galaxy*. Pan Books, London.
- [Al-Khaffaf et al., 2010] Al-Khaffaf, H., Talib, A., Osman, M., and Wong, P. (2010). Grec'09 arc segmentation contest: Performance evaluation on old documents. In Ogier, J.-M., Liu, W., and Lladós, J., editors, *Graphics Recognition. Achievements, Challenges, and Evolution*, volume 6020 of *Lecture Notes in Computer Science*, pages 251–259. Springer Berlin / Heidelberg. 10.1007/978-3-642-13728-0\_23.
- [Al-Khaffaf et al., 2012] Al-Khaffaf, H. S. M., Talib, A. Z., and Osman, M. A. (2012). Grec'11 arc segmentation contest: Performance evaluation on multi-resolution scanned documents. In Ogier, J.-M. and Kwon, Y. B., editors, *Graphics Recognition: Achievements, Challenges, and Evolution: 9th International Workshop, Grec 2011*, volume 7423. Springer Verlag.
- [Amsaleg et al., 2004] Amsaleg, L., Gros, P., and Berrani, S.-A. (2004). Robust Object Recognition in Images and the Related Database Problems. *Multimedia Tools and Applications*, 23(3):221–235.
- [Amsaleg et al., 2000] Amsaleg, L., Gros, P., and Mezhoud, R. (2000). Mise en base d'images indexées par des descripteurs locaux : problèmes et perspectives. Rapport de recherche RR-3903, INRIA.
- [Arévalo et al., 2010] Arévalo, G., Ducasse, S., Gordillo, S., and Nierstrasz, O. (2010). Generating a catalog of unanticipated schemas in class hierarchies using formal concept analysis. *Information and Software Technology*, 52(11):1167 – 1187. Special Section on Best Papers PROMISE 2009.
- [Armstrong et al., 2009] Armstrong, T. G., Moffat, A., Webber, W., and Zobel, J. (2009). Evaluatir: an online tool for evaluating and comparing ir systems. In Al-

- lan, J., Aslam, J. A., Sanderson, M., Zhai, C., and Zobel, J., editors, *SIGIR*, page 833. ACM.
- [Ayala-Ramirez et al., 2006] Ayala-Ramirez, V., Garcia-Capulin, C. H., Perez-Garcia, A., and Sanchez-Yanez, R. E. (2006). Circle detection on images using genetic algorithms. *Pattern Recognition Letters*, 27(6):652–657.
- [Belongie et al., 2002] Belongie, S., Malik, J., and Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(4):509–522.
- [Biederman, 1981] Biederman, I. (1981). On the semantics of a glance at a scene. In Kubovy, M. and Pomerantz, J., editors, *Perceptual Organization*, chapter 8, pages 213–255. Erlbaum, Hillsdale, N.J.
- [Biederman, 1985] Biederman, I. (1985). Human Image Understanding: Recent Research and a Theory. *Computer Vision, Graphics and Image Processing*, 32:29–73.
- [Biederman, 1987] Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, 94:115–147.
- [Breuel, 2008] Breuel, T. M. (2008). The ocropus open source ocr system. In Yanikoglu, B. A. and Berkner, K., editors, *DRR*, volume 6815 of *SPIE Proceedings*, page 68150. SPIE.
- [Brown, 1992] Brown, L. G. (1992). A Survey of Image Registration Techniques. *ACM Computing Surveys*, 24(4):325–376.
- [Cervera et al., 2003] Cervera, E., Pobil, A. P. D., Berry, F., and Martinet, P. (2003). Improving image-based visual servoing with three-dimensional features. *I. J. Robotic Res.*, 22(10-11):821–840.
- [Chang et al., 1997] Chang, W., Hespanha, J., Morse, A., and Hager, G. (1997). Task re-encoding in vision-based control systems.
- [Chaumette, 1997] Chaumette, F. (1997). Potential problems of stability and convergence in image-based and position-based visual servoing. In *Workshop on Vision and Control, Block Island, Rhode Island*.
- [Chen and Chung, 2001] Chen, T.-C. and Chung, K.-L. (2001). An Efficient Randomized Algorithm for Detecting Circles. *Computer Vision and Image Understanding*, 83(2):172–191.
- [Cheng and Liu, 2003] Cheng, Y. C. and Liu, Y.-S. (2003). Polling an Image for Circles by Random Lines. *IEEE Transactions on PAMI*, 25(1):125–130.
- [Chiu and Liaw, 2005] Chiu, S.-H. and Liaw, J.-J. (2005). An effective voting method for circle detection. *Pattern Recognition Letters*, 26:121–133.

- [Clavelli et al., 2010] Clavelli, A., Karatzas, D., and Lladós, J. (2010). A framework for the assessment of text extraction algorithms on complex colour images. In *Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*, pages 19–26, Boston, MA, USA. ACM.
- [Coustaty et al., 2011] Coustaty, M., Bertet, K., Visani, M., and Ogier, J.-M. (2011). A new adaptive structural signature for symbol recognition by using a galois lattice as a classifier. *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, 41(4):1136–1148.
- [Dengel, 2009] Dengel, A. (2009). The semantic desktop as a means for personal information management. In *KDIR 2009 - Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, page 5.
- [Dmitry and Ian, 2006] Dmitry, T. and Ian, H. (2006). Fact++ description logic reasoner : System description.
- [Dosch and Valveny, 2005] Dosch, P. and Valveny, E. (2005). Report on the Second Symbol Recognition Contest. In Liu, W. and Lladós, J., editors, *Sixth IAPR International Workshop on Graphics Recognition (GREC'05)*, volume 3926 of *Lecture Notes in Computer Science*, pages 381–397, Hong Kong SAR, China. City University of Hong Kong, Springer. <http://www.springer.com/lncs> Spanish project CICYT TIC2003-09291 French project Technivision EPEIRES.
- [Dosch et al., 2008] Dosch, P., Valveny, E., Fornes, A., and Escalera, S. (2008). Report on the Third Contest on Symbol Recognition. In Wenyin Liu, J. L. and Ogier, J.-M., editors, *Graphics Recognition. Recent Advances and New Opportunities*, volume 5046 of *Lecture Notes in Computer Science*, pages 321–328. Springer. French Techno-Vision program (Ministry of Research) Spanish project TIN2006-15694-C02-02 Spanish research program Consolider Ingenio 2010:MIPRCV (CSD2007-00018).
- [Drevin, 2011] Drevin, G. R. (2011). Extracting the ground level events of march 1942 from legacy cosmic ray recordings. In *Ninth IAPR Graphics Recognition Workshop - GREC*, Chung-Ang University, Seoul, Korea., IAPR.
- [Drummond and Cipolla, 1999a] Drummond, T. and Cipolla, R. (1999a). Real-time tracking of complex structures for visual servoing. In *Vision Algorithms, Theory and Practice*, pages 69–83. Springer Verlag, LNCS.
- [Drummond and Cipolla, 1999b] Drummond, T. and Cipolla, R. (1999b). Real-time tracking of complex structures with on-line camera calibration. In *Proceedings of British Machine Vision Conference*, volume 2, pages 574–583, Nottingham, UK.
- [Eco, 1990] Eco, U. (1990). *The limits of interpretation*. Indiana University Press, Bloomington.
- [Eco, 2007] Eco, U. (2007). *Dall'albero al labirinto: studi storici sul segno e l'interpretazione*. Bompiani.

- [Eco et al., 1992] Eco, U., Collini, S., Culler, J., Rorty, R., and Brooke-Rose, C. (1992). *Interpretation and Overinterpretation*. Tanner Lectures in Human Values. Cambridge University Press.
- [Espiau et al., 1992] Espiau, B., Chaumette, F., and Rives, P. (1992). A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8(3):313–326.
- [Everingham et al., 2010] Everingham, M., Van Gool, L., Williams, C. K. I., Winn, J., and Zisserman, A. (2010). The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338.
- [Ganter et al., 2005] Ganter, B., Stumme, G., and Wille, R., editors (2005). *Formal Concept Analysis, Foundations and Applications*, volume 3626 of *Lecture Notes in Computer Science*. Springer.
- [Ganter and Wille, 1999] Ganter, B. and Wille, R. (1999). *Formal concept analysis - mathematical foundations*. Springer.
- [Garris and Klein, 1998] Garris, M. D. and Klein, W. W. (1998). Creating and validating a large image database for METTREC. Technical Report NISTIR 6090, National Institute of Standards and Technology.
- [Gent and Kotthoff, 2011] Gent, I. P. and Kotthoff, L. (2011). Reliability of computational experiments on virtualised hardware. In *AI for Data Center Management and Cloud Computing*.
- [Geoffrois, 2009] Geoffrois, P. B. J. B.-T. E. (2009). Techno-vision special session: performance assessment of vision systems. In *Advanced Concepts for Intelligent Vision Systems*, Bordeaux, France.
- [Goecks et al., 2010] Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol*, 11(8):R86.
- [Gros, 1993] Gros, P. (1993). *Outils géométriques pour la modélisation et la reconnaissance d’objets polyédriques*. Thèse de doctorat, Institut National Polytechnique de Grenoble.
- [Grosicki et al., 2009] Grosicki, E., Carree, M., Brodin, J.-M., and Geoffrois, E. (2009). Results of the rimes evaluation campaign for handwritten mail processing. In *Document Analysis and Recognition, 2009. ICDAR '09. 10th International Conference on*, pages 941–945.
- [Hager et al., 1995] Hager, G., Chang, W., and Morse, A. (1995). Robot hand-eye coordination based on stereo vision. *IEEE Control Systems*, page 30.
- [Hager, 1997] Hager, G. D. (1997). A modular system for robust positioning using feed-

- back from stereo vision. *IEEE Transactions on Robotics and Automation*, 13(4):582–595.
- [Hanahara and Hiyane, 1991] Hanahara, K. and Hiyane, M. (1991). A circle-detection algorithm simulating wave propagation. *Machine Vision and Applications*, 4:97–111.
- [Hankerson et al., 2003] Hankerson, D., Menezes, A. J., and Vanstone, S. (2003). *Guide to Elliptic Curve Cryptography*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.
- [Heath et al., 2010] Heath, K., Gelfand, N., Ovsjanikov, M., Aanjaneya, M., and Guibas, L. (2010). Image webs: Computing and exploiting connectivity in image collections. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3432–3439.
- [Heidegger, 1967] Heidegger, M. (1967). *Being and Time*. Library of philosophy and theology. Blackwell.
- [Hilaire and Tombre, 2006] Hilaire, X. and Tombre, K. (2006). Robust and Accurate Vectorization of Line Drawings. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(6):890–904.
- [Horaud et al., 1998] Horaud, R., Dornaika, F., and Espiau, B. (1998). Visually guided object grasping. *IEEE Transactions on Robotics and Automation*, 14(4):525–532.
- [Hu et al., 2001a] Hu, J., Kashi, R., Lopresti, D., Nagy, G., and Wilfong, G. (2001a). Why table ground-truthing is hard. In *ICDAR01*, pages 129–133.
- [Hu et al., 2001b] Hu, J., Kashi, R., Lopresti, D., Nagy, G., and Wilfong, G. (2001b). Why table ground-truthing is hard. In *Proceedings of the Sixth International Conference on Document Analysis and Recognition*, pages 129–133, Seattle, WA.
- [Hynes et al., 2006] Hynes, P., Dodds, G., and Wilkinson, A. (2006). Uncalibrated visual-servoing of a dual-arm robot for mis suturing. In *Biomedical Robotics and Biomechatronics, 2006. BioRob 2006. The First IEEE/RAS-EMBS International Conference on*, pages 420–425.
- [IPOL, ] IPOL. Image Processing On Line. <http://www.ipol.im/>. ISSN:2105-1232, <http://dx.doi.org/10.5201/ipol>.
- [Jegou et al., 2008] Jegou, H., Douze, M., and Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In Forsyth, D. A., Torr, P. H. S., and Zisserman, A., editors, *ECCV (1)*, volume 5302 of *Lecture Notes in Computer Science*, pages 304–317. Springer.
- [Jones, 1997] Jones, N. D. (1997). *Computability and complexity - from a programming perspective*. Foundations of computing series. MIT Press.
- [Kim and Kim, 2000] Kim, W.-Y. and Kim, Y.-S. (2000). A region-based shape descriptor using zernike moments. *Signal Processing: Image Communication*, 16(1-2):95–102.

- [Knublauch et al., 2004] Knublauch, H., Musen, M. A., and Rector, A. L. (2004). Editing description logic ontologies with the protégé owl plugin. In Haarslev, V. and Möller, R., editors, *Description Logics*, volume 104 of *CEUR Workshop Proceedings*. CEUR-WS.org.
- [Lamdan and Wolfson, 1988] Lamdan, Y. and Wolfson, H. J. (1988). Geometric Hashing: A General and Efficient Model-Based Recognition Scheme. *Proceedings of 2nd International Conference on Computer Vision, Tampa, FL (USA)*, pages 238–249.
- [Lazebnik et al., 2003] Lazebnik, S., Schmid, C., and Ponce, J. (2003). Affine-invariant local descriptors and neighborhood statistics for texture recognition. In [DBL, 2003], pages 649–655.
- [Lejsek et al., 2006] Lejsek, H., Ásmundsson, Heiðar, F., Jónsson, B. P., and Amsaleg, L. (2006). Scalability of Local Image Descriptors: A Comparative Study. In *Proceedings of the 14th annual ACM international conference on Multimedia*, Santa Barbara, États-Unis.
- [Liang et al., 1997] Liang, J., Rogers, R., and Haralick, R. (1997). UW-ISL document image analysis toolbox: An experimental environment. In *In Proc. of the 4th International Conference on Document Analysis and Recognition*, pages 984–988.
- [Longuet-Higgins, 1981] Longuet-Higgins, H. C. (1981). A computer program for reconstructing a scene from two projections. *Nature*, 293:133–135.
- [Lopresti and Nagy, 2001] Lopresti, D. and Nagy, G. (2001). Issues in ground-truthing graphic documents. In *Proceedings of the Fourth IAPR International Workshop on Graphics Recognition*, pages 59–72, Kingston, Ontario, Canada.
- [Lopresti et al., 2010] Lopresti, D., Nagy, G., and Smith, E. B. (2010). Document analysis issues in reading optical scan ballots. In *Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*, pages 105–112, Boston, MA, USA. ACM.
- [Lopresti and Nagy, 2011] Lopresti, D. P. and Nagy, G. (2011). When is a problem solved? In [DBL, 2011], pages 32–36.
- [Lopresti and Nagy, 2012] Lopresti, D. P. and Nagy, G. (2012). Adapting the turing test for declaring document analysis problems solved. In Blumenstein, M., Pal, U., and Uchida, S., editors, *Document Analysis Systems*, pages 1–5. IEEE.
- [Lowe, 2004] Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.
- [Luong and Faugeras, 1996] Luong, Q. and Faugeras, O. (1996). The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76.
- [MADCAT, ] MADCAT. Multilingual automatic document classification analysis and

- translation (MADCAT). <http://www.darpa.mil/ipto/programs/madcat/madcat.asp>.
- [Marr, 1982a] Marr, D. (1982a). *Vision*. W. H. Freeman, San Francisco.
- [Marr, 1982b] Marr, D. (1982b). *Vision*. W. H. Freeman, San Francisco, CA.
- [Maru et al., 1993] Maru, N., Kase, H., Yamada, S., and Nishikawa, A. (1993). Manipulator control by visual servoing with the stereo vision. In *International Conference on Intelligent Robots and Systems, Yokohama, Japan*, pages 1866–1870. IEEE.
- [Moret and Shapiro, 2001] Moret, B. M. E. and Shapiro, H. D. (2001). Algorithms and experiments: The new (and old) methodology. *Journal of Universal Computer Science*, 7(5):434–446.
- [Motik et al., 2009] Motik, B., Shearer, R., and Horrocks, I. (2009). Hypertableau reasoning for description logics. *J. Artif. Int. Res.*, 36(1):165–228.
- [Müller et al., 2010] Müller, H., Clough, P., Deselaers, T., and Caputo, B., editors (2010). *ImageCLEF: Experimental Evaluation in Visual Information Retrieval*, volume 32 of *The Information Retrieval Series*. Springer, Berlin.
- [Murase and Nayar, 1995] Murase, H. and Nayar, S. (1995). Visual learning and recognition of 3D objects from appearance. *International Journal of Computer Vision*, 14:5–24.
- [Nagy, 2010] Nagy, G. (2010). Document systems analysis: Testing, testing, testing. In Doerman, D., Govindaraju, V., Lopresti, D., and Natarajan, P., editors, *DAS 2010, Proceedings of the Ninth IAPR International Workshop on Document Analysis Systems*, page 1.
- [Nelson, 1996] Nelson, R. (1996). Memory-based recognition of parts for curved and polyedral objects. In *Proc. of the ARPA Image Understanding workshop, Palm Springs, CA*.
- [Nietzsche, 1873] Nietzsche, F. (1873). On truth and lies in a nonmoral sense. Über Wahrheit und Lüge im außermoralischen Sinn.
- [Ockham, 1323] Ockham, W. o. (1323). *Summa logicae*.
- [Oliveira et al., 2011] Oliveira, D. M., Lins, R., Silva, G. P., Fan, J., and Thielo, M. (2011). Deblurring textual document images. In *Ninth IAPR Graphics Recognition Workshop - GREC*, Chung-Ang University, Seoul, Korea,. IAPR.
- [Pari et al., 2008] Pari, L., Sebastián, J. M., Traslosheros, A., and Ángel, L. (2008). Image based visual servoing: Estimated image jacobian by using fundamental matrix vs analytic jacobian. In Campilho, A. C. and Kamel, M. S., editors, *ICIAR*, volume 5112 of *Lecture Notes in Computer Science*, pages 706–717. Springer.

- [Peirce, 1998] Peirce, C. S. (1998). *Collected Papers of Charles Sanders Peirce*. Thoemmes Continuum.
- [Perd'och et al., 2009] Perd'och, M., Chum, O., and Matas, J. (2009). Efficient representation of local geometry for large scale object retrieval. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 9–16.
- [Popper, 1992] Popper, K. R. (1992). *The Logic of Scientific Discovery*. Routledge, reprint edition. Original edition, 1934 “Logik der Forschung”.
- [Rendek et al., 2004] Rendek, J., Masini, G., Dosch, P., and Tombre, K. (2004). The search for genericity in graphics recognition applications: Design issues of the qgar software system. In Marinai, S. and Dengel, A., editors, *Document Analysis Systems*, volume 3163 of *Lecture Notes in Computer Science*, pages 366–377. Springer.
- [Rice, 1953] Rice, H. G. (1953). Classes of recursively enumerable sets and their decision problems. *Trans. Amer. Math. Soc.*, 74:358–366.
- [Rohkohl et al., 2009] Rohkohl, C., Keck, B., Hofmann, H. G., and Hornegger, J. (2009). Technical note: Rabbitct—an open platform for benchmarking 3d cone-beam reconstruction algorithms. *Medical Physics*, 36(9):3940–3944.
- [Rothwell, 1995] Rothwell, C. (1995). The Importance of Reasoning about Occlusions during Hypothesis Verification in Object Recognition. Rapport de recherche 2673, INRIA.
- [Samson et al., 1990] Samson, C., Borgne, M. L., and Espiau, B. (1990). *Robot Control: the Task Function Approach*. Clarendon Press, Oxford University Press, Oxford, UK.
- [Schiele and Crowley, 1996] Schiele, B. and Crowley, J. (1996). Object recognition using multidimensional receptive field histograms. In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pages 610–619.
- [Schmid and Mohr, 1996] Schmid, C. and Mohr, R. (1996). Combining greyvalue invariants with local constraints for object recognition. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Francisco, California, USA*.<sup>32</sup>
- [Schmid et al., 2000] Schmid, C., Mohr, R., and Bauckhage, C. (2000). Evaluation of interest point detectors. *International Journal of Computer Vision*, 37(2):151–172.
- [Schulz et al., 2009] Schulz, F., Ebbecke, M., Gillmann, M., Adrian, B., Agne, S., and Dengel, A. (2009). Seizing the treasure: Transferring knowledge in invoice analysis. In *10th International Conference on Document Analysis and Recognition*, pages 848–852, Barcelona, Spain.
- [Schwab et al., 2000] Schwab, M., Karrenbach, M., and Claerbout, J. (2000). Making scientific computations reproducible. *Computing in Science and Engg.*, 2:61–67.

<sup>32</sup>[ftp://ftp.imag.fr/pub/labo-GRAVIR/MOVI/publications/Schmid\\_cvpr96.ps.gz](ftp://ftp.imag.fr/pub/labo-GRAVIR/MOVI/publications/Schmid_cvpr96.ps.gz).

- [Sebastián et al., 2009] Sebastián, J., Pari, L., Angel, L., and Traslosheros, A. (2009). Uncalibrated visual servoing using the fundamental matrix. *Robotics and Autonomous Systems*, 57(1):1 – 10.
- [Shafait et al., 2008] Shafait, F., Keysers, D., and Breuel, T. (2008). Grec 2007 arc segmentation contest: Evaluation of four participating algorithms. In Liu, W., Lladós, J., and Ogier, J.-M., editors, *Graphics Recognition. Recent Advances and New Opportunities*, volume 5046 of *Lecture Notes in Computer Science*, pages 310–320. Springer Berlin / Heidelberg. 10.1007/978-3-540-88188-9\_29.
- [Sigurdardottir et al., 2005] Sigurdardottir, R., Hauksson, H., Jónsson, B. Þ., and Amaleg, L. (2005). The Quality vs. Time Trade-off for Approximate Image Descriptor Search. In *21st International Conference on Data Engineering Workshops (ICDEW'05), EMMA - International Workshop on Managing Data for Emerging Multimedia Applications*, Tokyo, Japon.
- [Simmons, 2010] Simmons, R. J. (2010). Profile luis von ahn: Recaptcha, games with a purpose. *ACM Crossroads*, 17(2):49.
- [Sirovitch and Kirby, 1987] Sirovitch, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America*, 2:586–591.
- [Sivic and Zisserman, 2003] Sivic, J. and Zisserman, A. (2003). Video google: A text retrieval approach to object matching in videos. In [DBL, 2003], pages 1470–1477.
- [Smeaton et al., 2006] Smeaton, A. F., Over, P., and Kraaij, W. (2006). Evaluation campaigns and TRECVID. In *MIR '06: Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval*, pages 321–330, New York, NY, USA. ACM Press.
- [Smeulders et al., 2000] Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380.
- [Smith, 2010] Smith, E. H. B. (2010). An analysis of binarization ground truthing. In *Proceedings of the 8th IAPR International Workshop on Document Analysis Systems*, pages 27–34, Boston, MA, USA. ACM.
- [Stefan Jaeger et al., 2006a] Stefan Jaeger, Guangyu Zhu, David Doermann, Kevin Chen, and Summit Sampat (2006a). DOCLIB: a Software Library for Document Processing. In *International Conference on Document Recognition and Retrieval XIII*, pages 1–9. San Jose, CA.
- [Stefan Jaeger et al., 2006b] Stefan Jaeger, Guangyu Zhu, David Doermann, Kevin Chen, and Summit Sampat (2006b). DOCLIB: a Software Library for Document Processing. In *International Conference on Document Recognition and Retrieval XIII*, pages 1–9. San Jose, CA.

- [Tabbone et al., 2006] Tabbone, S., Wendling, L., and Salmon, J.-P. (2006). A new shape descriptor defined on the radon transform. *Computer Vision and Image Understanding*, 102(1):42–51.
- [Tarantola, 2005] Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation*. Society for Industrial Mathematics.
- [Tobacco, ] Tobacco. Tobacco800 litigation data set. <http://www.umiacs.umd.edu/~zhugy/Tobacco800.htmlzhugy/Tobacco800.html>.
- [Tombre et al., 2005] Tombre, K., Tabbone, S., and Dosch, P. (2005). Musings on Symbol Recognition. In *Graphics Recognition - Ten Years Review and Future Perspectives - Revised Selected Papers from GREC'2005*, volume 3926 of *Lecture Notes in Computer Science*, pages 23–34, Hong Kong/China. Springer Verlag.
- [Turing, 1950] Turing, A. M. (1950). Computing machinery and intelligence. *Mind*, LIX(236):433–460.
- [Turk and Pentland, 1991] Turk, M. and Pentland, A. (1991). Face recognition using eigenfaces. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Maui, Hawaii, USA*, pages 586–591.
- [UNLV, ] UNLV. UNLV data set. <http://www.isri.unlv.edu/ISRI/OCRtk>.
- [UW1, ] UW1. UW english document image database I: A database of document images for OCR research. <http://www.science.uva.nl/research/dlia/datasets/uwash1.html>.
- [UW2, ] UW2. UW-II english/japanese document image database: A database of document images for OCR research. <http://www.science.uva.nl/research/dlia/datasets/uwash2.html>.
- [UW3, ] UW3. UW-III english/technical document image database. <http://www.science.uva.nl/research/dlia/datasets/uwash3.html>.
- [Valveny and Dosch, 2004] Valveny, E. and Dosch, P. (2004). Symbol Recognition Contest: A Synthesis. In Lladós, J. and Kwon, Y.-B., editors, *Graphics Recognition: Recent Advances and Perspectives, Fifth International Workshop - GREC 2003*, volume 3088 of *Lecture Notes in Computer Science*, pages 368–385, Barcelone, Espagne. Josep Lladós, Springer-Verlag. Colloque avec actes et comité de lecture. internationale. A04-R-219 || valveny04b A04-R-219 || valveny04b.
- [van Rijsbergen, 1979] van Rijsbergen, C. J. (1979). *Information Retrieval*. Butterworth.
- [Venant and Victorri, 2007] Venant, F. and Victorri, B. (2007). Représentation géométrique de la synonymie. *Le Français Moderne*, (1):81–96.
- [Victorri, 1994a] Victorri, B. (1994a). La construction dynamique du sens. In Porte, M., editor, *Passions des formes - à René Thom*, pages 733–747. ENS Éditions Fontenay St Cloud.

- [Victorri, 1994b] Victorri, B. (1994b). The use of continuity in modeling semantic phenomena. In Victorri, C. F. . B., editor, *Continuity in linguistic semantics*, pages 241–251. Benjamins.
- [Wenyin, 2006] Wenyin, L. (2006). The Third Report of the Arc Segmentation Contest. In Liu, W. and Lladós, J., editors, *Graphics Recognition—Ten Years Review and Future Perspectives*, volume 3926 of *Lecture Notes in Computer Science*, pages 358–361. Springer-Verlag.
- [Winn and Everingham, 2007] Winn, J. and Everingham, M. (2007). The pascal visual object classes challenge 2007 (voc2007) annotation guidelines. <http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2007/guidelines.html>.
- [Wittgenstein, 2001] Wittgenstein, L. (2001). *Philosophical investigations : the German text, with a revised English translation*. Blackwell, Oxford Malden, Mass.
- [Wojnarski et al., 2010] Wojnarski, M., Stawicki, S., and Wojnarowski, P. (2010). TunedIT.org: System for automated evaluation of algorithms in repeatable experiments. In *Rough Sets and Current Trends in Computing (RSCTC)*, volume 6086 of *Lecture Notes in Artificial Intelligence (LNAI)*, pages 20–29. Springer.
- [Yang, 2005] Yang, S. (2005). Symbol recognition via statistical integration of pixel-level constraint histograms: A new descriptor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(2):278–281.
- [Zhang and Lu, 2002] Zhang, D. and Lu, G. (2002). Shape-based image retrieval using generic fourier descriptor. *Signal Processing: Image Communication*, 17:825–848.
- [Zhang et al., 2006] Zhang, W., Wenyin, L., and Zhang, K. (2006). Symbol recognition with kernel density matching. *Pattern Recognition*, 28(12):2020–2024.
- [Zobel et al., 2011] Zobel, J., Webber, W., Sanderson, M., and Moffat, A. (2011). Principles for robust evaluation infrastructure. In *Proceedings of the 2011 workshop on Data infrastructurEs for supporting information retrieval evaluation*, DESIRE '11, pages 3–6, New York, NY, USA. ACM.

