Relations explicites entre différentes représentations d'image dans un modèle de graphe visuel

Trong-Ton Pham, Philippe Mulhem, Loic Maisonnasse

Relations explicites entre différentes représentations d'image dans un modèle de graphe visuel

Trong-Ton Pham [1,2]**, Philippe Mulhem**[1]**, Loïc Maisonnasse**[3]

[1] *Laboratoire d'Informatique de Grenoble (LIG) - 38041 Grenoble Cedex 9, France*
[2] *Institut Polytechnique de Grenoble (Grenoble INP)*
[3] *TecKnowMetrix - Voiron, France*
{trong-ton.pham, philippe.mulhem}@imag.fr; lm@tkm.fr

RÉSUMÉ. *Nous présentons dans ce papier une nouvelle méthode pour exploiter la relation entre différents niveaux de représentation d'image afin de compléter le modèle de graphe visuel. Le modèle de graphe visuel est une extension du modèle de langue classique en recherche d'information. Nous utilisons des régions d'images et des points d'intérêts (associées automatiquement à des concepts visuels), ainsi que des relations entre ces concepts, lors de la construction de la représentation sous forme de graphe. Les résultats obtenus sur catégorisation de la collection RobotVision de la compétition d'ImageCLEF 2009 (contenant 5 pièces dans un environnement à l'intérieur du bâtiment) montrent que (a) la procédure de l'induction automatique des concepts d'une image est efficace, et (b) l'utilisation des inter-relations entre 2 niveaux de représentation, en plus de concepts, permet d'améliorer le taux de reconnaissance.*

ABSTRACT. *This paper presents a novel approach, the first to our knowledge, that exploits a complete extension of the language modeling approach from information retrieval to the problem of graph-based image retrieval and categorization. Since photographic images are 2D data, we first use image regions and local interest points (mapped to automatically induced concepts) and then relationships between these regions to build a complete graph representation of images. The results obtained on categorizing of RobotVision collection from ImageCLEF 2009 (containing of 5 rooms in an indoor environment) show that (a) the procedure to automatically induce concepts from an image is effective, and (b) the use of spatial relationships, in addition to concepts, for representing an image content helps improve the classifier accuracy.*

MOTS-CLÉS : *Représentation de graphes, recherche d'image, catégorisation d'image*

KEYWORDS: *Graph representation, image retrieval, image categorization*

1. Introduction

La recherche d'image, depuis ses balbutiements au début des années 90, est toujours un défi majeur pour les chercheurs. A la différence de problèmes comme la recherche d'information textuelle, qui reposent sur le fait que les symboles du langage traité véhicule du sens compréhensible par un humain, le cas des images nous confronte à ce qui est communément appelé le *fossé sémantique* (i.e. la difficulté de passer de pixels à du sens) ainsi qu'à la difficulté très grande de représenter le contenu des images de manière compacte et fidèle. Des travaux se sont portés sur l'utilisation de connaissances extérieures aux images, comme par exemple le fait que les images sont associées à des informations de date et d'heure dans (Platt *et al.,* 2003), permettant de les regrouper, ou bien des informations de géolocalisation (Kennedy *et al.,* 2007). Ces connaissances permettent de regrouper les images et de faire émerger des éléments utiles à la recherche et à l'interrogation de bases d'images fixes.

On peut par ailleurs faire un parallèle entre différents points de vue pour un humain et le fait que de nombreuses approches permettent de décrire le contenu des images. L'idée est alors de cumuler ces points de vue pour améliorer la qualité des résultats fournis. Plusieurs travaux ont par le passé proposé l'utilisation de relations spatiales entre régions d'image pour leur indexation et leur recherche, afin de cumuler caractérisations de régions et relations spatiales. Par exemple, les descriptions par chaînes 2D (2D strings) comme on les trouve dans le système Visualseek (Smith *et al.,* 1996) capturent les séquences d'apparition d'objets suivant une ou plusieurs directions de lecture.

D'autres travaux ont considéré l'utilisation de régions d'images dans des modèles probabilistes, en se basant par exemple sur des modèles de Markov cachés 1D (Iyengar *et al.,* 2005) ou 2D, comme dans (Smith *et al.,* 1996) et (Yuan *et al.,* 2007). Ces travaux s'intéressent à l'annotation d'images et n'utilisent pas les relations lors du traitement des requêtes. Les relations entre des éléments d'images peuvent également être exprimés par l'intermédiaire de conventions de nommage, comme dans (Papadopoulos *et al.,* 2007) où les relations sont utilisées pour l'indexation. Enfin des travaux tels que (Mulhem *et al.,* 2006) se sont focalisés sur des graphes conceptuels pour l'indexation et la recherche des images. Les représentations explicites de relations provoquent la génération de graphes complexes, ayant un impact négatif sur la correspondance de graphe qui est déjà coûteuse (Ounis *et al.,* 1998).

Des travaux ont permis de s'intéresser à des modèles de langue adaptés à représenter le contenu des images, comme par exemple (Pham *et al.,* 2009b). Les modèles de langue pour la recherche d'information existent depuis la fin des années 90 (Ponte *et al.,* 1998). Dans ce cadre, la valeur de pertinence d'un document pour une requête donnée est estimée par la probabilité que la requête soit générée par le document. Même si cette approche a été initialement proposée pour des unigrammes (c'est-à-dire des termes isolés), plusieurs extensions ont été proposées pour traiter des *n-grammes* (i.e. des séquences de termes) (Song *et al.,* 1999, Srikanth *et al.,* 2002), et plus récemment, des relations entre termes et également des graphes. Par exemple, (Gao *et al.,* 2004)

propose a) d'utiliser un analyseur de dépendance pour représenter les documents et les requêtes, et b) une extension de l'approche à base de modèle de langue pour manipuler ces arbres. Maisonnasse *et al.* (Maisonnasse *et al.*, 2008) ont étendu cette approche avec un modèle compatible avec des graphes plus généraux, comme ceux obtenus par une analyse conceptuelle des documents et des requêtes.

D'autres approches (comme (Fergus *et al.*, 2005, Gosselin *et al.*, 2007)) ont respectivement utilisé des réseaux probabilistes et des noyaux pour capturer des relations dans les images, ce qui est également notre intention ici. Dans le cas de (Fergus *et al.*, 2005), l'estimation des probabilités des régions repose sur l'algorithme EM, qui est sensible aux probabilités initiales. Dans le modèle que nous proposons, au contraire, la fonction de vraisemblance est convexe et possède un maximum global. Dans le cas de (Gosselin *et al.*, 2007), le noyau utilisé ne considère que les trois plus proches régions d'une région de référence. Nous intégrons dans notre modèle toutes les régions voisines d'une région. Enfin, contrairement à ces travaux, à ceux de (Iyengar *et al.*, 2005) et de (Barnard *et al.*, 2003) basés sur des modèles de langues, nous utilisons explictement des étiquettes de relations spatiales.

Afin de résoudre certains de ces problèmes, (Pham *et al.*, 2009b) a proposé de définir un modèle de langue sur des graphes représentant des images, en se basant sur des concepts caractérisant un point de vue de l'image et des relations spatiales. Aucun des travaux présentés ci-dessus ne considère différents points de vue de l'image provenant de différentes représentations des caractéristiques visuelles des images. Notre objectif ici est de présenter une approche qui permet d'intégrer des éléments de différentes natures dans une représentation à base de graphe, et d'utiliser un modèle de langue sur ces graphes pour permettre de rechercher des images. En particulier, nous nous intéressons ici à un niveau expérimental à la prise en compte d'éléments d'orientation et d'éléments provenant de régions d'intérêts.

La suite de cet article est organisée comme suit : la section 2 présente le modèle de langue visuel utilisé pour décrire le contenu des images, ainsi que la procédure de correspondance utilisée pour calculer la similarité entre images ; la section 3 décrit ensuite les résultats obtenus par notre approche pour un problème de catégorisation portant sur des données tirées de la compétition CLEF Robotvision, dont l'objectif est de retrouver pour un robot la pièce dans laquelle il se situe ; nous concluons en section 4.

2. Modélisation d'images avec des graphes visuels

Dans (Pham *et al.*, 2009b), nous avons représenté l'image comme un graphe probabiliste qui permet de capturer la complexité visuelle de l'image. Ces graphes sont représentés par un ensemble de concepts pondérés, reliés par un ensemble d'associations orientées. Les concepts visent à caractériser le contenu de l'image, tandis que les associations orientées expriment les relations entre les concepts (par exemple : relation spatiale ou relation symbolique). Par ailleurs, notre hypothèse est que chaque concept

est représenté par une caractéristique visuelle unique extraite à partir de l'image (par exemple : la couleur, la texture ou des caractéristiques locales SIFT). Dand cet article, nous étendons ce modèle de graphe afin de considérer plusieurs des ca racatéristiques à la fois.

2.1. *Défintion*

Nous supposons que chaque image i est représentée par l'ensemble des ensembles de concepts pondérés $S^i_{WC} = \{W^i_C\}$ et un ensemble des ensembles de relations pondérées $S^i_{WE} = \{W^i_E\}$. Chaque image est représentée par un graphe :

$$G_i =< S^i_{WC}, S^i_{WE} > \quad [1]$$

où chaque concept c de l'ensemble W^i_C correspond à un mot visuel utilisé pour représenter l'image. Le poids des concepts représente le nombre d'occurrences de ce concept dans l'image. $\mathcal{C}$ dénote un ensemble de concepts généré par une caractéristique visuelle sur toute la collection. W^i_C est un ensemble de couples $(c, \#(c, i))$, où c est un élément de $\mathcal{C}$ et $\#(c, i)$ est le nombre de fois c apparait dans l'image i :

$$W^i_C = \{(c, \#(c, i))|c \in \mathcal{C}\} \quad [2]$$

Chaque relation étiquetée entre une paire de concepts $(c, c') \in \mathcal{C} \times \mathcal{C}'$ est représentée par un triplet $((c, c'), l, \#(c, c', l, i))$, où l est un élément de $\mathcal{L}$, l'ensemble des étiquettes possibles pour une relation, et $\#(c, c', l, i)$ est le nombre de fois que c et c' sont liés avec l'étiquette l de l'image i. W^i_E est alors définie par :

$$W^i_E = \{((c, c'), l, \#(c, c', l, i))|(c, c') \in \mathcal{C} \times \mathcal{C}', l \in \mathcal{L}\} \quad [3]$$

La figure 1 décrit un exemple de génération de graphe en se basant sur un découpage en 3×3 blocs avec extraction de couleur uniquement, et utilisation de deux relations spatiales entre les concepts couleurs. Dans le cas d'une utilisation de plusieurs ensembles de concepts (couleurs et texture par exemple), nous aurions donc davantage de concepts et d'autres relations, le principe restant similaire.

2.2. *Le modèle de langue pour l'appariement des graphes visuels*

Après avoir défini la représentation pour les graphes visuels, nous passons maintenant au problème de l'appariement d'un graphe de requête G_q avec de graphe de document G_d. Basé sur le modèle de langue défini sur les graphes proposés dans (Maisonnasse *et al.,* 2009), nous présentons ici une extension de cette méthode qui gère des ensembles des concepts et des ensembles des relations.

La probabilité pour générer un graphe de requête $G_q =< S^q_{WC}, S^q_{WE} >$ à partir d'un graphe de document G_d est définie par :

$$P(G_q|G_d) \quad = \quad P(S^q_{WC}|G_d) \times P(S^q_{WE}|S^q_{WC}, G_d) \quad [4]$$

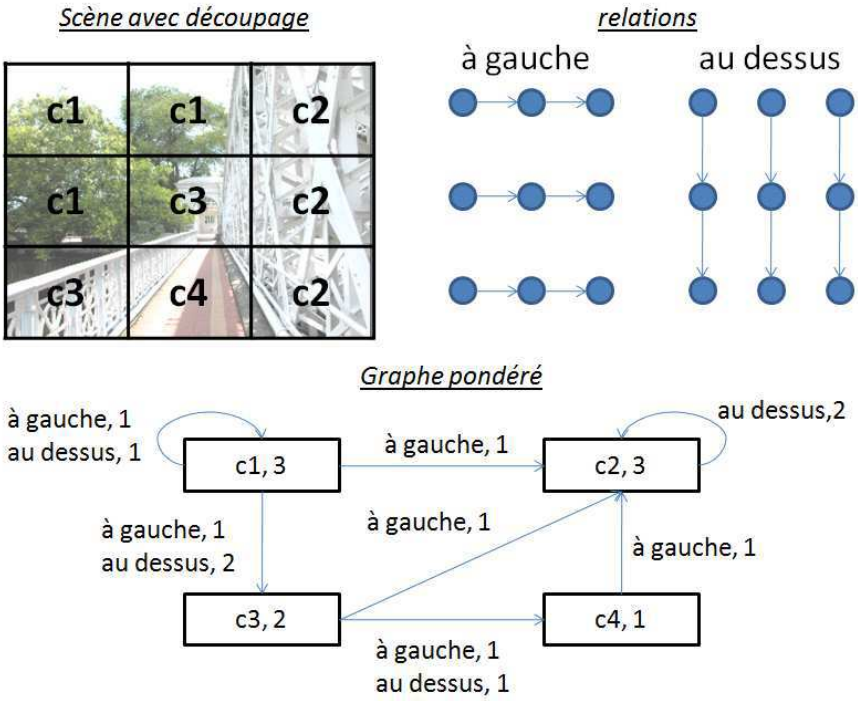

Figure 1. *Exemple d'une scène avec découpage en 3×3 blocs, extraction d'un ensemble de concepts avec deux relations, ainsi que le graphe correspondant.*

Nous nous reposons sur l'hypothèse d'indépendance conditionnelle entre ensemble de concepts, hypothèse classique en recherche d'information. La probabilité de générer l'ensemble de concepts de l'image requête peut donc être écrite comme suit :

$$P(S^q_{WC}|G_d) = \prod_{W^q_C \in S^q_{WC}} P(W^q_C|G_d) \quad [5]$$

En supposant l'hypothèse de l'indépendance entre les concepts et en posant que le nombre d'occurrences des concepts suit un modèle multinomial, nous calculons $P(W^q_C|G_d)$ comme suit :

$$P(W^q_C|G_d) \quad \propto \quad \prod_{n \in \mathcal{C}} P(c|G_d)^{\#(c,q)}$$

où $\#(c,q)$ désigne le nombre de fois que le concept c apparait dans le graphe requête. Cette contribution correspond à la probabilité du concept tel que proposé dans (Maisonnasse *et al.*, 2009). Similairement à ces travaux antérieurs, les paramètres du

modèle $P(c|G_d)$ sont estimés par maximum de vraisemblance, avec un lissage de Jelinek-Mercer :

$$P(c|G_d) \quad = \quad (1-\lambda_c)\frac{\#(c,d)}{\#(.,d)} + \lambda_u \frac{\#(c,D)}{\#(.,D)} \qquad [6]$$

où $\#(c,d)$ représente le nombre d'occurrences de c dans le graphe de document, et où $\#(.,d) = \sum_c \#(c,d)$. Les quantités $\#(c,D)$ sont calculées similairement, mais sont définies sur l'ensemble de la collection (i.e. union de toutes les graphes de documents dans la collection).

2.3. *Intégration des relations entre différents concepts*

Basé sur l'hypothèse indépendante entre ensembles de relations, nous suivons un processus analogue pour générer l'ensemble des relations de la requête :

$$P(S^q_{WE}|S^q_{WC}, G_d) = \prod_{W^q_E \in S^q_{WE}} P(W^q_E|S^q_{WC}, G_d) \qquad [7]$$

Pour la probabilité de générer les relations de requête à partir du document, nous supposons qu'une relation ne dépend que des deux ensembles de concepts liés. En supposant que les relations sont indépendantes et que leurs distributions suivent un modèle multinomial, nous calculons :

$$P(W^q_E|S^q_{WC}, G_d) \propto \prod_{(c,c',l)\in \mathcal{C}\times\mathcal{C}'\times\mathcal{L}} P(L(c,c') = l|W^q_C, W^{q\prime}_C, G_d)^{\#(c,c',l,q)}$$

où $c \in \mathcal{C}$, $c' \in \mathcal{C}'$ and $L(c,c')$ est une variable contenant dans $\mathcal{L}$, indiquant les étiquettes des relations possibles entre c et c', dans cet ensemble de relation. Comme précédemment, les paramètres du modèle $P(L(c,c') = l|W^q_C, W^{q\prime}_C, G_d)$ sont estimés par maximum de vraisemblance avec un lissage Jelinek-Mercer, ce qui donne :

$$P(L(c,c') = l|W^q_C, W^{q\prime}_C, G_d) = \qquad [8]$$
$$(1-\lambda_e)\frac{\#(c,c',l,d)}{\#(c,c',.,d)} + \lambda_e \frac{\#(c,c',l,D)}{\#(c,c',.,D)}$$

où $\#(c,c',l,d)$ représente le nombre de fois où les concepts c et c' sont liés avec l'étiquette l dans l'image, et $\#(c,c',.,d) = \sum_{l\in\mathcal{L}} \#(c,c',l,d)$. Par convention, dans le cas où l'un des deux concepts n'apparaît pas dans un graphe d, nous posons :

$$\frac{\#(c,c',l,d)}{\#(c,c',.,d)} = 0$$

Ici, les quantités $\#(c,c',l,D)$ sont similaires, mais sont définies sur toute la collection (i.e. comme vu précédemment, union de tous les graphes de l'ensemble des documents dans la collection).

Le modèle que nous venons présenter est une généralisation du modèle défini dans (Pham *et al.*, 2009b) qui correspondent au cas particulier où un seul ensemble de concepts et un seul ensemble de relations sont utilisés. Dans le cas particulier où les relations ne sont pas considérées, notre modèle correspond au modèle de (Pham *et al.*, 2009a).

A partir de ce modèle, nous calculons la valeur de pertinence (RSV) d'une image du document d pour la requête q à l'aide de la divergence de Kullback-Leibler entre le graphe du document G_d et le graphe de requête G_q.

3. Expérimentations

Dans cette partie, nous décrivons tout d'abord la collection utilisée pour notre expérimentation, puis nous présentons les résultats obtenus avec notre modèle sur cette collection. Notre objectif est de démontrer que le modèle de graphe visuel, présenté dans la section précédente, est bien adapté à la représentation du contenu de l'image. De plus, l'intégration des inter-relations entre différentes facettes, correspondant à différents niveaux de concepts, aide à améliorer la représentation classique d'images.

La collection utilisée pour nos expérimentations est celle de la tâche RobotVision[1], collection fournie dans le cadre de la campagne ImageCLEF 2009. Cette tâche aborde le problème de la localisation d'un robot mobile en utilisant uniquement l'information visuelle provenant d'une camera de qualité moyenne. La difficulté de cette tâche est que le robot doit reconnaître la pièce dans différentes conditions de luminosité et de s'adapter aux changements de l'environnement autour (telles que le déplacement de personnes, de nouveaux objets ajoutés au fil du temps, etc.) La collection de RobotVision contient un ensemble d'apprentissage de 1034 images et un ensemble de 909 images pour la validation (voir Fig. 1). L'ensemble d'apprentissage et l'ensemble de validation ont été enregistrés à l'intérieur d'un même étage de bâtiment composé de 5 chambres, et ces deux ensembles ont été filmés avec un intervalle de durée de 6 mois. Ensuite, les images de test (1690 images) ont été enregistrées 20 mois plus tard, il est a noté que ces images contiennent des images d'une pièce supplémentaire qui n'appartenaient pas à l'ensemble d'apprentissage.

Le système que nous avons utilisé pour participer à la compétition RobotVision est composé de deux processus : une étape de reconnaissance et une étape de post-traitement. Dans cet article, nous allons uniquement décrire l'étape de reconnaissance du système afin d'évaluer l'impact du modèle proposé. Le lecteur intéressé peut trouver davantage de détails sur l'étape de post-traitement dans (Pham *et al.*, 2009a).

Deux difficultés principales que notre système pris en compte pendant cette compétition sont :

1) l'adaptation avec le changement de condition d'éclairage. Comme on a vu précédemment, la variation de luminosité entre différents ensembles d'images (apprentis-

1. http ://imageclef.org/2009/robot

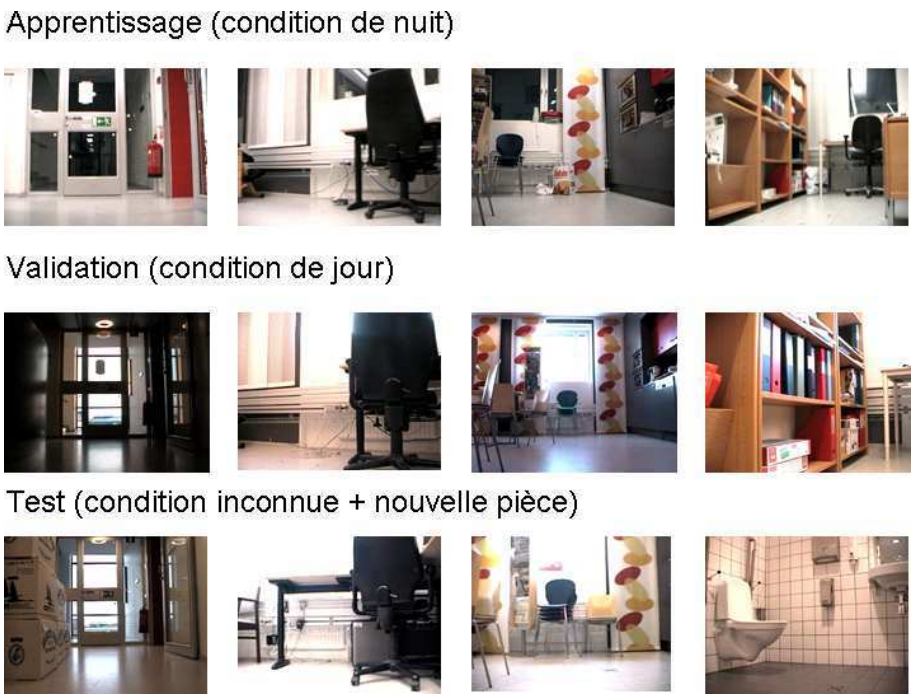

Figure 2. *Exemple des images de la collection RobotVision. L'ensemble d'apprentissage est pris dans la condition de nuit. Tandis que l'ensemble de validation est pris dans la condition de jour. Pour le test, la condition est inconnue en ajoutant une nouvelle pièce et des objets qui ne sont pas appris à priori.*

sage, validation et test) est très grande. Ca pourrait diminuer fortement la performance du système de reconnaissance. Pour cela, nous présentons la solution dans la partie qui suit.

2) l'évolution des objets à l'intérieur et l'ajout de nouvelle pièce. Face à ce challenge, nous utilisons des techniques comme le regroupement des images, ou l'application un filtre des pièces *inconnues*. Ces techniques sont présentées dans l'étape de post-traitement des résultats obtenus par l'étape de reconnaissance.

A partir de l'ensemble de validation, nous avons appris que l'histogramme de couleur ne s'adapte pas très bien au changement de luminosité (dans différent condition de jour, de nuit ou nuageux). Par exemple, dans les mêmes conditions d'éclairage, l'histogramme de couleur peut donner quelques bons résultats. Toutefois, dans le cas d'un changement brutal de l'état de la lumière (comme apprentissage en conditions de nuit et teste en condition de jour), le système ne parvient pas à porter un jugement satisfaisant. Nous avons donc décidé de n'utiliser que certaines caractéristiques visuelles qui sont moins sensibles à la condition de luminosité pour générer des concepts visuels.

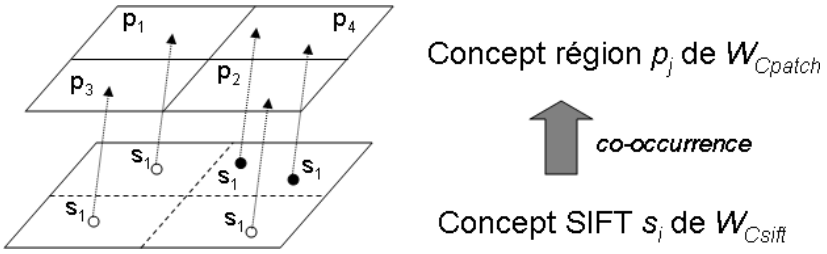

Figure 3. *Illustration de la relation co-occurrence entre le concept SIFT s_1 et le concept région p_j. Par exemple, s_1 (•) est lié 2 fois à p_4, avec la relation co-occurence. Et s_1 (○) est lié 3 fois à p_4, avec la relation non-occurrence.*

Pour cette raison, deux types de concepts visuels (région et SIFT) et une relation de co-occurrence sont extraits à partir des images :

1) Chaque image est divisée en 5x5 régions régulières. Nous avons extrait pour chaque région un histogramme d'orientation comme dans (Won *et al.*, 2002). Ensuite, un vocabulaire visuel de 500 concepts visuels est construit en utilisant l'algorithme de regroupement non supervisé (i.e. regroupement par k-moyenne). A partir de ce vocabulaire, nous construisons un ensemble des concepts régions pondérées W_{Cpatch}.

2) De manière similaire à ci-dessus, des caractéristiques visuelles locales SIFT sont extraites à partir des points d'intérêts. Plus précisément, des points d'intérêts sont détectés à l'aide du détecteur SIFT (Lowe, 2004) pour chaque image. Ensuite, les caractéristiques locales sont utilisées pour créer l'ensemble des 500 concepts SIFTs pondérés W_{Csift}.

3) En utilisant les deux ensembles de concepts précédents, nous définissons une relation co-occurrence entre les concepts régions et les concepts SIFT. Cette relation représente le lien entre les détails de l'objet (concepts SIFTs) et sa forme globale (concepts régions). Une relation de co-occurrence est définie si le concept SIFT s_i de W_{Csift} est situé dans la zone du concept region p_j de W_{Cpatch} (voir Fig. 2). Au contraire, une relation non-occurrence est définie lorsque concept SIFT s_i apparaît en dehors de la zone associée au concept région p_j. Nous notons cette relation pondérée W_{Rcooc}.

Basé sur les définitions précédentes, nous avons mis en place plusieurs modèles de graphes visuels pour évaluer la performance des modèles proposés.

– $G^P =< \{W_{Cpatch}\}, \emptyset >$, qui utilise des concepts regions.

– $G^S =< \{W_{Csift}\}, \emptyset >$, qui utilise des concepts SIFT.

– $G^{S.P} =< \{W_{Csift}, W_{Cpatch}\}, \emptyset >$, qui fusionne les deux concepts region et SIFT sans relation.

– $G^{S \rightarrow P} =< \{W_{Csift}, W_{Cpatch}\}, \{W_{Rcooc}\} >$, qui fusionne des concepts régions et des concepts SIFT en utilisant les relations co-occurrence entre deux niveaux.

Tableau 1. *Résultats de différents modèles de graphes visuels sur l'ensemble de validation et sur l'ensemble de test*

	G^P	G^S	$G^{S.P}$	$G^{S \to P}$
Validating set	345	285	334.5	**466.5** (+39.5%)
Test set	80.5	263	209.5	**293.5** (+40.1%)

Les trois premiers modèles ont été estimés suivant l'équation présentée dans la section 2.2. Le quatrième modèle est le graphe complet et la probabilité des relations a été calculée selon l'équation définie dans la section 2.3.

L'évaluation proposée sur ce corpus d'images est la suivante : on mesure la précision de prédiction entre l'annotation manuelle de la pièce avec celle classée par les systèmes. Le score officiel pour une séquence de test est calculé en assignant : *+1.0 point* pour chaque image bien classée et *-0,5 point* pour chaque image mal classée. Ce score dénote ainsi un taux de reconnaissance du system de classification.

Le Tableau 1 présente les résultats en termes des valeurs de score pour chaque modèle. Comme nous l'avions escompté, les deux modèles de base G^P et G^S donnent un bon score pour l'ensemble de validation. Cependant, le modèle G^P n'obtient pas un bon score sur l'ensemble de test à cause de l'introduction de la nouvelle pièce et de l'évolution des objets. La fusion simple de deux modèle $G^{S.P}$ fournit un résultat inferieur aux meilleurs résultats de G^P et G^S. Cependant, ce résultat est plus robuste car il diminue une partie de l'effet parasite de chaque élément visuel (ie. $G^{S.P}$ surpasse le résultat moyen de G^P et G^S dans les deux cas).

L'ajout des relations de co-occurrence dans le graphe de fusion $G^{S.P}$ améliore ces résultats de **39,5%** (pour l'ensemble de validation) et **40,1%** (pour l'ensemble de test) respectivement. Ces résultats confirment que l'intégration des relations joue un rôle très positif en considérant plusieurs points de vue du contenu des images qui sont complémentaires. En outre, nos résultats montrent que le lien entre la forme de l'objet et les détails de sa présentation donne une meilleure abstraction du contenu de l'image.

4. Conclusion

Nous avons introduit dans cet article un nouveau modèle de graphe visuel et une méthode d'appariement des graphes extraits à partir d'images. Ce modèle de graphe permet de représenter les relations entre les concepts associés aux régions d'une image et les concepts associés aux points d'intérêts. Notre modèle s'inscrit dans les approches à base de modèle de langue en recherche d'information, et étend un certain nombre de travaux sur ces modèles génératifs appliqués aux graphes. D'un côté plus pratique, l'examen des régions de concepts permet d'acquérir une généralité dans la description d'images, et dans le même temps, l'examen des concepts de points inté-

rêts permettent d'ajouter tous les détails de l'objet avant d'intégrer tous ces éléments à travers les relations de co-occurrence.

Les résultats expérimentaux ont confirmé non seulement la stabilité des graphes visuels construits uniquement par des concepts visuels, mais aussi les avantages de l'intégration de la relation établie entre les différents types de concepts. Un autre objectif de ce travail est de combler l'écart entre le niveau sémantique et le niveau signal du contenu de l'image ; ce travail sera donc complété à l'avenir par l'ajout de concepts visuels (global ou local) et les relations spatiales entre différents types de concepts, en intégrant des concepts symboliques comme des noms d'objets.

Dans le futur, nous allons nous baser sur ces travaux pour proposer des expérimentations prenant en compte des relations à l'intérieur des facettes et entre les facettes des graphes. Comme nous avons montré indépendamment dans cet article que les résultats sont améliorés, mais nous espérons obtenir des résultats enore meilleurs sur différents corpus de classification et de recherche d'image.

Remerciements

Ce travail a été financé par l'Agence Nationale de la Recherche (ANR-06-MDCA-002), projet AVEIR.

5. Bibliographie

Barnard K., Duygulu P., Forsyth D., de Freitas D., Blei D., Jordan M. J., « Matching Words and Pictures », *Journal of Machine Learning Research*, vol. 2003, n° 3, p. 1107-1135, 2003.

Fergus R., Perona P., Zisserman A., « A sparse object category model for efficient learning and exhaustive recognition », *Conference on Computer Vision and Pattern Recognition*, 2005.

Gao J., Nie J.-Y., Wu G., Cao G., « Dependence language model for information retrieval », *In SIGIR '04 : Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 170-177, 2004.

Gosselin P., Cord M., Philipp-Foliguet S., « Kernels on bags of fuzzy regions for fast object retrieval », *International conference on Image Processing*, 2007.

Iyengar G., Duygulu P., Feng S., Ircing P., Khudanpur S. P., Klakow D., Krouse M. R., Manmatha R., Nock H. J., Petkova D., Pytlik B., Virga P., « Joint Visual-Text Modeling for automatic Retrieval of Multimedia Documents », *In ACM Multimedia, Singapore*, p. 21-30, 2005.

Kennedy L., Naaman M., Ahern S., Nair R., Rattenbury T., « How flickr helps us make sense of the world : context and content in community-contributed media collections », *In Proceedings of the 15th international Conference on Multimedia*, p. 631-640, 2007.

Lowe D. G., « Distinctive image features from scale-invariant keypoints », *Journal of Computer Vision*, vol. 60, n° 2, p. 91-110, 2004.

Maisonnasse L., Gaussier E., Chevalet J., « Model Fusion in Conceptual Language Modeling », *In 31st European Conference on Information Retrieval (ECIR09)*, p. 240-251, 2009.

Maisonnasse L., Gaussier E., Chevallet J., « Multiplying Concept Sources for Graph Modeling », *In C. Peters, V. Jijkoun, T. Mandl, H. Muller, D.W. Oard, A. Peñas, V. Petras, D. Santos, (Eds.) : Advances in Multilingual and Multimodal Information Retrieval. LNCS #5152. Springer-Verlag.*, 2008.

Mulhem P., Debanne E., « A framework for Mixed Symbolic-based and Feature-based Query by Example Image Retrieval », *International Journal for Information Technology*, vol. 12, n° 1, p. 74-98, 2006.

Ounis I., Pasca M., « RELIEF : Combining Expressiveness and Rapidity into a Single System », *In SIGIR '98 : Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, p. 266-274, 1998.

Papadopoulos T., Mezaris V., Kompatsiaris I., Strintzis M. G., « Combining Global and Local Information for Knowledge-Assisted Image Analysis and Classification », *EURASIP Journal on Advances in Signal Processing*, 2007.

Pham T. T., Maisonnasse L., Mulhem P., « Visual Language Modeling for Mobile Localization », *CLEF working notes 2009, Corfu, Greece*, 2009a.

Pham T.-T., Maisonnasse L., Mulhem P., Gaussier E., « Modèle de langue visuel pour la reconnaissance de scènes », *CORIA*, p. 99-112, 2009b.

Platt J. C., Czerwinski M., Field B. A., « PhotoTOC : Automatic Clustering for Browsing Personal Photographs », *Proc. Fourth IEEE Pacific Rim Conference on Multimedia*, 2003.

Ponte J. M., Croft W. B., « A language modeling approach to information retrieval », *In SIGIR '98 : Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval*, p. 275-281, 1998.

Smith J. R., Chang S. F., « VisualSEEk : a fully automated content-based image query system », *In Proceedings of the Fourth ACM international Conference on Multimedia*, p. 87-98, 1996.

Song F., Croft W. B., « General language model for information retrieval », *CIKM'99*, p. 316-321, 1999.

Srikanth M., Srikanth R., « Biterm language models for document retrieval », *Research and Development in Information Retrieval*, p. 425-426, 2002.

Won C. S., Park D. K., Park S.-J., « Efficient Use of MPEG-7 Edge Histogram Descriptor », *ETRI Journal*, p. vol.24, no.1, 2002.

Yuan J., Li J., Zhang B., « Exploiting spatial context constraints for automatic image region annotation », *In Proceedings of the 15th international Conference on Multimedia*, p. 25-29, 2007.