

De la reconnaissance de visages à l'interprétation de scènes complexes

Françoise Breton, Cordelia Schmid

► **To cite this version:**

Françoise Breton, Cordelia Schmid. De la reconnaissance de visages à l'interprétation de scènes complexes. Collection "20 ans d'avancées et de perspectives en sciences du numérique", INRIA, 2012, 2 p. hal-00812717

HAL Id: hal-00812717

<https://hal.inria.fr/hal-00812717>

Submitted on 12 Apr 2013

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.

De la reconnaissance de visages à l'interprétation de scènes complexes



© Inria / Photo Kaksonen

Les recherches en reconnaissance automatique d'images ont connu un grand essor au début des années 2000. On peut citer par exemple le contrôle d'accès (visage, iris) ou la mise au point sur les visages détectés automatiquement au sein des appareils photos numériques. Dix ans plus tard, émergent des services permettant de rechercher des informations sur le web à partir d'images...

Témoignage de **Cordelia Schmid**, responsable de l'équipe Lear

Au début des années 1990, on commençait à appliquer des techniques de reconnaissance à des objets très simples, comme par exemple des outils noirs sur fond blanc, au sein d'ensembles très petits qui, même en 1996, ne comptaient pas plus de quelques milliers d'images.

Aujourd'hui, nous travaillons sur la recherche d'images beaucoup plus complexes dans des bases pouvant contenir cent millions d'images ! Cette progression résulte d'un ensemble de travaux permettant d'identifier les caractéristiques pertinentes pour représenter l'image, d'organiser ces caractéristiques et de les stocker de façon à faciliter leur recherche ultérieure. Nous avons contribué à ces avancées en développant des descripteurs discriminants, robustes aux changements de points de vue, et en élaborant des outils d'indexation d'images.



© Inria / Photo Kaksonen

“ L'apprentissage pour mieux identifier les contenus ”

Ces approches sont cependant loin de pouvoir répondre à des requêtes du type « trouver des images d'enfants jouant dans un jardin ». C'est en effet un problème très difficile à résoudre car cela suppose une machine capable d'interpréter la scène, c'est-à-dire de savoir reconnaître un enfant, une action dans l'image (jouer) et savoir ce qu'est un jardin. Nous avons progressé dans ce sens en nous appuyant, depuis une dizaine d'années, sur des techniques d'apprentissage : on fournit à l'ordinateur de très nombreuses images de jardins en le chargeant d'extraire des éléments pertinents pour la construction d'un modèle de jardin. Plus récemment, dans le cadre de l'identification du contenu des vidéos, nous avons complété cette approche par une exploitation des informations de mouvement. L'analyse du mouvement permet en effet l'obtention d'indications sur l'action en train de se faire. Ces techniques nous permettent aujourd'hui d'identifier rapidement des vidéos représentant par exemple une fête d'anniversaires ou une personne buvant du café.

Comment en savoir plus sur l'objet que vous venez de prendre en photo ?



Si des applications visant la reconnaissance de visages ont pu voir le jour au tournant des années 2000, il faudra en revanche attendre une dizaine d'années supplémentaires pour être en mesure d'identifier des objets plus complexes. Par exemple, la start-up Inria Milpix, créée en 2007, exploite des travaux précurseurs de l'équipe Lear dans le domaine de la reconnaissance d'objets. Cette jeune société réalise des applications permettant, en prenant la photo d'un objet, livre ou affiche de cinéma, d'avoir accès sur le web à des informations complémentaires concernant l'objet, comme les boutiques où l'on peut acheter ce livre ou les cinémas jouant le film présenté sur l'affiche. Une application similaire est proposée depuis

peu par Google Goggles.

ET DANS 20 ANS ?

« Il y a encore beaucoup à faire ! C'est en effet un réel défi que de faire apprendre aux ordinateurs des modèles visuels pour l'ensemble des classes et catégories existantes à partir d'énormes quantités d'images et de vidéos disponibles aujourd'hui. En particulier, les techniques d'apprentissage doivent évoluer afin d'obtenir des outils capables de traiter de très grandes bases de données et ceci de façon faiblement supervisée. Il faut aussi développer des méthodes actives de recherche dans ces bases, afin de dégager des informations pertinentes pour la construction des modèles par apprentissage. Je suis convaincue que l'approche « apprentissage » doit progresser conjointement avec la modélisation visuelle des données. »

Numérique & société

Reconnaissance de visage sur Facebook, appareils photos ou smartphones

- **Facebook** a accepté de suspendre son outil de reconnaissance faciale dans l'Union Européenne et d'effacer les données collectées sur les profils de ses usagers, ont annoncé vendredi 21 septembre le réseau social et l'autorité irlandaise chargée de la protection des données privées (DPC). *Source : nouvelobs.com*
- **Depuis 2005**, les appareils photos disposent d'une possibilité de reconnaissance faciale. L'appareil fait un focus sur les visages et réalise la photo en moins de 0,05 seconde. L'appareil peut cibler un visage et y appliquer la bonne mise au point même lorsque celui-ci est en mouvement.
- **Depuis 2011**, La reconnaissance faciale est aussi utilisée sur Smartphone pour déverrouiller le téléphone, mais la technique est à améliorer (des tests ont montré qu'une simple photo pouvait tromper le logiciel). Selon le cabinet d'étude ABI Research, un smartphone sur cinq vendu aux consommateurs en 2012 sera équipé de la reconnaissance faciale et d'ici cinq ans, ce sont 665 millions de téléphones et tablettes qui disposeront nativement de cette technologie. Pour l'instant, seuls quelques rares privilégiés en bénéficient. On compte parmi eux les Samsung Galaxy S3, Google Galaxy Nexus ou encore HTC One X. *Source : Tomsguide.fr*

1992 - 2012



- Collection "20 ans d'avancées et de perspectives en sciences du numérique" par les chercheurs d'équipes Inria de Grenoble et Lyon.
- www.inria.fr/20ansgrenoble

© Inria - Editions
Victoria