

défense nationale

études politiques - stratégiques - militaires
économiques - scientifiques

EUROSATORY

Europe et armement terrestre

Serge VINÇON, François LUREAU, Luc VIGNERON,
Christian SCHMIDT, Gérard BEZACIER,
Jean-Paul PERRUICHE, Pierre-Jean LASSALLE,
Antonin TISSERON, Darko RIBNIKAR, François VELTZ

60^e ANNIVERSAIRE DU DÉBARQUEMENT

René de NAUROIS, Alexandre ORLOV,
Louis DURTESTE, Alain DUMONTET

L'ONU à la veille de nouveaux défis

Pascal CHAIGNEAU



revue mensuelle
juin 2004
10 €

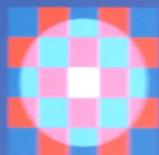


Image mining : **voir plus vite**

Recherche, filtrage et catégorisation d'images

François VELTZ

L'analyse des images satellites et la biométrie sont à l'origine de nombreuses technologies innovantes. Depuis une décennie, l'explosion du multimédia et la prolifération sur *Internet* des données visuelles, licites ou illicites, ont largement contribué à renforcer la demande en matière d'investigation et de surveillance des images et des vidéos, faisant émerger de nouvelles manières de naviguer dans les données.

Concepteur de l'« ADN de l'image », la société LTU Technologies développe et commercialise un moteur de recherche d'images, ainsi que des outils d'investigation et de filtrage évolués. A l'instar du *data mining*, cette technologie française exclusive et brevetée applique des techniques particulières aux grandes bases de données d'images.

De nombreuses agences de renseignement et institutions de forces de l'ordre emploient aujourd'hui cette technologie pour lutter contre la pédopornographie, le trafic d'objets culturels ou la contrefaçon. D'importantes sociétés privées protègent, valorisent leurs actifs visuels et mènent des actions d'intelligence économique à l'aide de ces outils.

La technologie n'a pas la prétention de se substituer à l'opérateur humain mais au contraire de l'assister et d'augmenter sa capacité d'action. Dans le contexte d'une production mondiale d'images qui s'accélère, les perspectives technologiques sont saisissantes.

En quelques années seulement, fait assez rare pour être souligné, une technologie française s'est imposée - par ses résultats – auprès de grandes agences d'investigation américaines. L'interopérabilité requise par la coopération des institutions des forces de l'ordre au plan international en a fait un standard de l'investigation liée à l'image.

UNE PRODUCTION D'IMAGES EXPONENTIELLE

La quantité d'images produites dans le monde croît beaucoup plus vite que la capacité humaine – collective – de traitement de ces images. Et le mouvement ne cesse de s'accélérer.

Tous les secteurs de l'activité humaine sont concernés. Plus de 100 milliards de photos sont prises chaque année dans le monde, dont une proportion toujours croissante au format numérique. Les téléphones mobiles équipés d'un appareil photo battent des records de ventes et se miniaturisent. Aujourd'hui, on constate que plus d'images sont générées par ces appareils que par aucun autre moyen, numérique ou argentique. Chacune des dix milliards de pages du *web* contient en moyenne quatre images et les réseaux *Peer-to-Peer* démultiplient les échanges. Le milieu industriel et scientifique compte aussi nombre de secteurs qui produisent de grandes quantités d'images. Par exemple, la recherche pharmaceutique génère d'innombrables photos lors de tests combinatoires et robotisés sur des cultures de cellules. Il est humainement impossible d'examiner et de classer tous les résultats. Les résultats de simulation, les schémas techniques, les bases de données de brevets sont aussi concernés.

C'est donc une grande partie de la connaissance humaine qui est encodée de manière graphique et dont il importe de maîtriser les outils de lecture et d'accès. Le recours aux techniques de vision par ordinateur, dite également vision artificielle, est inéluctable si l'on souhaite contrôler, filtrer ou exploiter les informations de ces images.

DE L'ADN DES IMAGES A LEUR SEMANTIQUE

Doter la machine de la faculté de trouver des corrélations entre images sensibles ou de repérer des éléments graphiques subtils constitue un défi ancien pour les scientifiques. De la recherche pure des années 1970, la décennie suivante a consacré la vision industrielle en 3D. Les années 1990 ont marqué le retour du 2D, avec pour objectif la reconnaissance automatique du contenu des images. Aujourd'hui, le résultat de ces recherches débouche sur la commercialisation de technologies.

* Echanges de pair à pair, sans passer par un serveur. Voir Naptser, Kazaa, eMule et BitTorrent.

L' « ADN de l'image »

Les images sont complexes à analyser, dans la mesure où il n'existe ni alphabet, ni relation d'ordre, ni taxonomie, ni même de propriété de transitivité. La constitution d'un dictionnaire ou d'un thésaurus est donc impossible. Les technologies actuelles de « vision par ordinateur » s'inspirent parfois du fonctionnement de la vision humaine. Les quelques deux mille canaux du nerf optique envoient en effet une information réduite au cerveau qui les intègre et « voit », c'est-à-dire interprète les signaux en des concepts.

Emprunt élégant à la biologie, la technologie française de LTU Technologies définit un « ADN de l'image qui définit les propriétés photométriques (couleurs, lumières) et géométriques (formes des objets) de l'image.

Pour chaque image, le système extrait en effet un identifiant unique, également appelé signature numérique, d'où l'expression « ADN de l'image ». Cet ADN représente l'image sous la forme d'un ensemble compact de valeurs numériques qui codent efficacement toutes les informations relatives au contenu visuel. Plus de deux mille facteurs sont pris en compte. L'ADN est un fichier très léger (environ 1Ko), robuste aux variations classiques des images (résolution, contrastes, rotation, distorsions, découpe, changement de format...) et optimisé pour les recherches dans de très grandes bases de données.

Par la suite, des métriques de similarité calculent des distances entre les différents ADN. Les images sont donc recherchées, filtrées et catégorisées par comparaison relative de leurs ADN. L'ordinateur peut véritablement interpréter l'image par son contenu visuel et non exclusivement numérique. Cette spécificité offre la possibilité de retrouver des fichiers *clones* ou *similaires* ce qui s'avère impossible avec les algorithmes MD5 ou SHA-1 (fonction de hachage) pour lesquels la moindre altération du fichier d'origine engendre deux empreintes distinctes.

Images dupliquées, clones et similaires

	Dupliquées	Les images sont exactement les mêmes (copies binaires)
	Clones	Images proches, modulo des transformations (étirement, coupe, recompression, ...)
	Similaires	Images visuellement similaires. Niveau le plus sémantique.

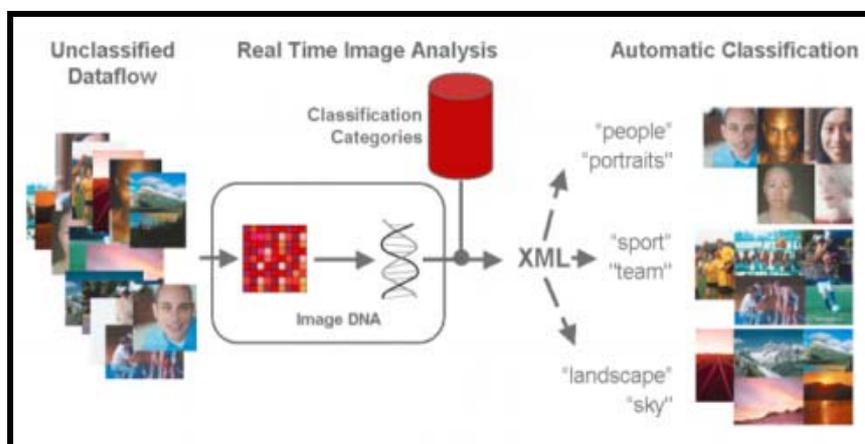
Sémantique et apprentissage

La technologie permet aussi, dans un cadre bien défini, de passer de l'ADN des images à leur description sémantique. En d'autres termes, il devient possible d'associer automatiquement des mots-clefs aux images.

Le contenu de l'ADN est analysé au cours du processus d'indexation par différents modules experts afin d'être reconnu relativement à des bases de connaissances prédéfinies. La description peut être *absolue* (« qu'est-ce ? » , « ce sont des personnes ») ou *relative* (« qu'est-ce qui s'apparente à ça ? » , « ceci ressemble à cela »)

Les bases de connaissances peuvent être enrichies : le moteur d'analyse peut apprendre de nouvelles catégories d'objets.

Classification automatique



Indexation semi-automatique

L'indexation des images conditionne les traitements ultérieurs de recherche et de filtrage, puisque indexer revient à numéroter. D'ordinaire, cette indexation se fait manuellement, par l'ajout d'un ensemble d'annotations descriptives de l'image.

Dans le cas de bases de données d'images très conséquentes, le défi de l'indexation devient critique. En termes de délais, mais aussi et surtout en termes de ressources humaines. Pour cette raison, il est économiquement intéressant de « propager » les annotations descriptives d'une base d'images renseignée à l'ensemble des images considérées comme similaires par l'ordinateur. Appelé « Suggestion Automatique de Mots-clefs », le procédé se présente alors comme un système semi-automatique : proposition par la machine de mots-clefs, puis confirmation - ou infirmation - par l'humain. On peut ainsi capitaliser sur les méta-données existantes d'une organisation et réduire considérablement les coûts d'indexation.

Navigation dans les données

Les bases d'images gigantesques qui naissent aujourd'hui font émerger de nouvelles façons de naviguer dans les données.

La navigation « classique » dans les bases de données d'images opère sur les annotations associées aux images. Celles-ci peuvent être générées *directement* par les appareils de prise de vue (date, voire lieu d'acquisition par géolocalisation) ou retrouvées *indirectement* (avec une analyse des mots de la page *web* où la photo est publiée, par exemple) Cette navigation ne prend pas en compte le contenu des images en elles-mêmes.

Ces annotations n'existent pas toujours, loin s'en faut. Lorsqu'elles sont absentes, la navigation doit être purement visuelle : la navigation par « similarité visuelle » consiste à soumettre au moteur une « image requête », qui répondra par un ensemble d'images « visuellement similaires » Appelée *bouncing*, cette recherche permet de converger vers un résultat, de façon naturelle et interactive. Elle s'affine au fur et à mesure par la modification des pondérations de couleur et de forme. Juxtaposer des images considérées comme similaires par l'ordinateur permet aussi de faire émerger du sens (corrélations, tendances, recoupements)

Il reste que le texte et l'image sont complémentaires : la combinaison des deux étant meilleure que le meilleur des deux, il est avantageux de coupler recherche visuelle et recherche par mot-clef classique. Cette complémentarité texte/image est d'ores et déjà en production et s'avère très performante.

Le processus de navigation, d'investigation le cas échéant, se trouve remanié en profondeur, mais surtout accéléré.

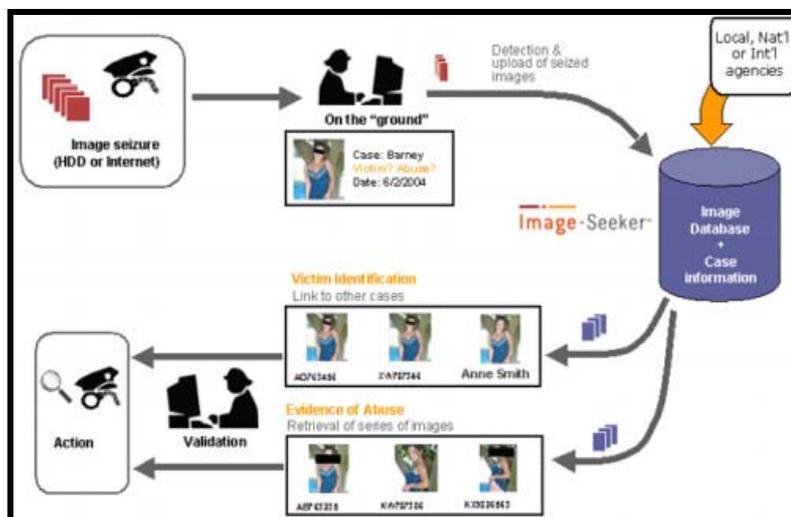
APPLICATIONS DE LA VISION PAR ORDINATEUR

Lutte contre la cyber pédo-criminalité

Déployée par le *FBI*, les douanes américaines ainsi que par *Interpol*, la vision par ordinateur apporte sa contribution dans le démantèlement de réseaux d'échanges d'images pédo-pornographiques. La technologie est également en usage dans plusieurs polices nationales en Europe et en voie de l'être dans de nombreuses organisations européennes ou internationales qui recensent des bases de données d'images pédophiles, dans un objectif d'investigation. Certaines organisations souhaitent rester confidentielles.

Le processus de comparaison entre images saisies et images référencées dans les bases de données est largement automatisé, accélérant d'autant les investigations. Des rapports détaillés différencient les images *dupliquées*, *clones*, *similaires* ainsi que les *nouvelles* images (c'est-à-dire inconnues jusqu'alors) Cela permet d'effectuer des recoupements concernant les circuits de distribution des images, l'identification de victimes et de criminels.

Investigation par l'image



Les contenus échangés par les pédophiles sont des images et des vidéos, pour la plupart anciennes et connues des services de police. On dénombrerait aujourd'hui plusieurs centaines de milliers d'images différentes dont l'utilisation d'*Internet* multiplie considérablement les échanges.

Les acteurs technologiques travaillent activement à développer les technologies de reconnaissance existantes pour traiter les vidéos, celles-ci pouvant être considérées comme une succession d'images fixes. On constate également des progrès quant à l'interception de flux *Peer-to-Peer* et l'identification des diffuseurs de fichiers.

Agréger les ADN des images pédophiles connues des services de police du monde entier permettrait de constituer un filtre incontournable pour stopper la prolifération de ces images sur *Internet*. Les trois millions de pages personnelles hébergées sur des serveurs français et les ADN des images qu'elles contiennent pourraient aisément être comparés aux ADN des images pédophiles connues. Cette comparaison est exclusivement automatique : elle ne nécessite aucune intervention humaine. La robustesse des ADN fait que la publication d'une image criminelle connue, même déformée, serait impossible. Le fonctionnement est par analogie celui d'un anti-virus informatique, qui détecte aussi les empreintes de virus connus.

Le filtrage d'images *dupliquées* et *clones* peut donc être entièrement automatique. La question légitime de la modération, et donc de la censure, ne se poserait que dans le cas d'une extension de ce type de filtrage aux images considérées comme *similaires* par l'ordinateur. En effet, certaines parmi celles-ci peuvent ne *pas* être pédophiles dans la réalité. *In fine*, seul le jugement humain a le dernier mot. La détection de contenus négationnistes, incitant à la haine raciale ou autres, rencontre les mêmes limites et impose la mise en place d'équipes de modérateurs.

Lutte contre la contrefaçon et le trafic des objets culturels

Selon l'OCDE, la contrefaçon représente 5 à 7% du commerce économique mondial. Le préjudice atteint les Etats, les entreprises et les consommateurs. Il a été rappelé que la contrefaçon était depuis quelques années aux mains d'organisations criminelles, voire de groupes terroristes. La contrefaçon touche tous les secteurs d'activité et plus particulièrement l'informatique, l'audiovisuel, l'industrie du jouet, la parfumerie, les produits pharmaceutiques et le secteur automobile.

La technologie LTU est employée aux Etats-Unis dans la lutte contre les faussaires et les trafiquants qui fabriquent fausses cartes d'identité, passeports, monnaies, chèques et cartes bleues (sans puce électronique outre-atlantique) Pour leurs desseins, les faussaires du monde entier utilisent des images en haute résolution des éléments qu'ils veulent reproduire. A la suite d'arrestations et de perquisitions, des disques durs d'ordinateurs sont saisis et examinés. Une revue manuelle de tous les fichiers serait strictement impossible, aussi l'analyse automatique des images détermine-t-elle rapidement la présence ou non de fichiers images délictueux. Les images sont classées rapidement en catégories, permettant aux agents de se concentrer pleinement sur leurs investigations.

Selon *Interpol*, le vol et le trafic clandestin d'objets culturels sont devenus l'une des formes dominantes de la criminalité internationale. Les deux pays les plus particulièrement touchés par ce phénomène sont la France et l'Italie, mais les pays d'Europe Centrale ou d'Afrique noire sont aussi concernés. Un système d'investigation est à l'étude en Europe, impliquant les particuliers, les assurances et les services de police.

Intelligence économique et surveillance d'Internet

Les technologies de vision artificielle trouvent aussi des applications qui concernent la vie économique et juridique des entreprises. L'adjonction de *spiders* ou *crawlers* à la technologie de l'image décrite précédemment permet de parcourir *Internet* en fournissant des rapports automatisés de surveillance. Des robots informatiques parcourent le web, « regardent » les images et les classent. Ils déjouent les camouflages de fichiers (modification d'extension, entre autres) et analysent les fichiers compressés ou les pièces jointes des courriers électroniques. La liste des visuels à surveiller sur le web « visible » ou « invisible » est longue et correspond à autant d'objectifs différents : visuels délictueux, diffamatoires, suspects (plans de bombes artisanales), etc.

De grandes entreprises surveillent l'utilisation de leurs marques sur *Internet*, de façon automatisée. Principalement pour détecter et contre-attaquer le marché dit « gris », constitué par l'ensemble des canaux de distribution non officiels des produits griffés.

Certaines sociétés, dont les marques emblématiques représentent une part importante de leur capital immatériel, surveillent le détournement de leurs logos et anticipent certaines campagnes de dénigrement ou tentatives de déstabilisation. D'autres surveillent la publication d'images sans

acquiescement de droits, voire étudient l'extension de la technologie pour contrevenir au piratage des contenus vidéo numériques. Ce dernier point représente 8 milliards de dollars de pertes pour les entreprises de ce secteur.

La politique de communication n'est pas seule concernée. Certaines entreprises travaillant sur des sujets sensibles surveillent et interceptent la fuite accidentelle ou malveillante de documents (plans ou prototypes) par le filtrage des flux sortants des messageries. Si l'offre de solutions de filtrage des flux entrants (*spam*) est pléthorique, celle concernant l'analyse des flux sortants reste peu développée.

Aspects biométriques

Depuis les attentats du 11 septembre 2001 et le renforcement des menaces terroristes, l'attention s'est fortement portée sur les réponses que la technologie pouvait apporter aux problématiques d'automatisation totale ou partielle des contrôles des personnes. Cela concerne la circulation, la détection et la mémorisation des personnes, en lien avec le concept de la « reconnaissance »

En l'état actuel de la technique, il faut clairement établir que toute tentative visant à essayer de repérer un criminel avec des caméras installées dans la rue et couplées à un logiciel de reconnaissance des photos relève de la science-fiction.

Des trois composantes relatives à la reconnaissance des personnes (détection, vérification, identification), les deux premières sont aujourd'hui relativement bien maîtrisées. L'identification pose réellement problème. La précision est intimement liée à la participation active ou passive du sujet, qui détermine les conditions d'acquisition de l'image.

A défaut d'une participation active, la performance de reconnaissance se détériore rapidement. Pour que l'identification des visages soit fiable et atteigne des performances industrielles, les sujets doivent se prêter au jeu et respecter une procédure prédéfinie de prise de vue. Ce qui remet bien sûr en question les technologies de visages visant à surveiller automatiquement des suspects passifs de façon fiable (stade ou aéroport)

Une reconnaissance efficace des visages n'est donc pas envisageable à ce jour, et il n'existe pas non plus de technologies en cours de développement dans les centres de recherche qui puissent laisser espérer une analyse pertinente des photos, dans un proche avenir. Une recherche entièrement

automatisée où le système indiquerait en première réponse le bon suspect n'est pas crédible. En revanche une approche semi-automatique permettant de gagner un temps précieux dans les enquêtes peut être envisagée, au travers d'une agrégation de différents systèmes mis en concurrence (technologies biométriques disponibles, recherche d'images par similarité et autres)

Le volet de la reconnaissance des comportements (surveillance dans les transports en communs) est moins répandu que la reconnaissance des personnes. Il est plus récent et à ce titre plus confidentiel, peu de solutions ou de travaux de recherche étant passés à la phase de commercialisation

PERSPECTIVES

Les technologies de *text mining* intéressent industriels du *knowledge management* et services de renseignement depuis de nombreuses années. Pour l'image, il est frappant de constater qu'il existe moins d'une dizaine de fournisseurs innovants dans le monde. La technologie d'*image mining* la plus avancée est française.

Les performances concernant la recherche, le filtrage et la catégorisation d'images dépendent de nombreux paramètres. Mais deux facteurs ressortent plus particulièrement : la puissance du processeur et l'importance de la mémoire vive. Il est parfaitement clair que le sens de l'histoire informatique, en décuplant régulièrement les capacités précitées va permettre à cette technologie d'analyse d'image d'être couronnée de succès. Les ressources supplémentaires en calcul vont permettre d'obtenir des résultats plus rapidement. La qualité, c'est à dire la pertinence des résultats va également s'améliorer.

La conjugaison des technologies de *data mining* (texte, audio, image) démultipliera les performances d'analyse des flux de données.

Enfin, les progrès relatifs aux interfaces homme-machine (commande vocale et gestuelle, techniques de représentation et de visualisation des données, espaces de travail immersifs, etc) enrichiront certainement les possibilités et les stratégies de recherche.

CONCLUSION

La prolifération sur le réseau *Internet* des données visuelles a largement contribué à renforcer la demande en matière de gestion de contenus. En particulier, la volumétrie des données images connaît une croissance si importante que le recours à des techniques de vision artificielle est devenu incontournable. La technologie n'a pas la prétention de se substituer à l'opérateur humain mais au contraire de l'assister et d'augmenter sa capacité d'action.

Si la collecte massive d'informations ne pose pas réellement problème, il en va tout autrement pour leur analyse et leur traitement logiciel. L'articulation *Sigint (Signal Intelligence) – Humint (Human Intelligence)* reste plus que jamais d'actualité. *In fine*, seul le jugement humain est apte à décider du caractère illicite d'une image ou capable d'extraire les éléments « pertinents » pour une problématique donnée. Néanmoins, l'ordinateur apporte une contribution majeure par l'analyse automatique et rapide de bases de données images très importantes. La maîtrise de la technologie liée à la reconnaissance d'images est une des clefs dans la maîtrise globale de l'information.

François VELTZ

François Veltz est ingénieur et diplômé de l'IAE de Paris. Il a travaillé chez Silicon Graphics (SGI) avant de rejoindre LTU Technologies.

Abstract

Law Enforcement and intelligence agencies have always been keen on using image processing and computer vision techniques (e.g. remote sensing, biometrics). For a decade, the exponential growth of digital imaging and the *Internet* have led to a new series of issues: image-based investigations, image-guided computer forensics, surveillance of sensitive images.

LTU Technologies is a company that seems to have tackled this problem successfully. Several Law Enforcement and Intelligence agencies in the US and worldwide are using its "image mining" technology, in applications ranging from child exploitation to counterfeiting and tracking of stolen art.

This is not a technology aiming at replacing the human being. To the contrary, it is a powerful tool at the service of the investigators, who can let the software find image correlations and focus themselves on making decisions and undertaking action. Given the importance of images and videos in today's insecure world, the perspectives of such a technology for homeland security are stunning.