

Intérêts et limites de la vidéo-surveillance intelligente pour la Sécurité Globale

Pierre BERNAS¹

¹EVITECH – +33.820.2008.39 – pbernas at evitech dot com

1. Introduction

Depuis la fin des années 90, la numérisation des contenus et la progression de puissance des ordinateurs ont rendu possible le traitement en temps réel des images de la vidéo **pour en extraire des interprétations** (que voit-on à l'image, que se passe-t-il, qui va où, etc). D'abord effectués en noir et blanc, puis en couleur, ces traitements ont commencé à sortir des laboratoires de recherche dans cette période, et à constituer des solutions exploitables, d'abord pour la surveillance routière, puis pour la surveillance de personnes et d'objets, et plus récemment pour la biométrie faciale.

Le domaine progresse désormais dans de multiples applications pour la Sécurité Globale dont nous allons essayer de tracer les perspectives ici. L'enjeu en est essentiel : **c'est l'interprétation automatique de l'information vidéo qui permet l'exploitation des caméras de vidéo-surveillance à une grande échelle** : elle rend possible le déploiement utile de multiples caméras de surveillance dont les logiciels intelligents garantissent que seule une information élaborée et filtrée parviendra, en temps réel, aux responsables de la sécurité intéressés. Les murs d'écrans des années 80 sont aujourd'hui démontés.

Les thèmes de la numérisation et de la compression d'images, ceux du traitement d'images fixes (empreintes digitales, photos satellite), et les applications militaires embarquées de détection de missiles ou d'approches aériennes (radar) ne sont pas abordés ici.

2. Traitements d'images vidéo

2.1. Qu'est ce qu'une séquence vidéo ?

Une séquence vidéo numérique peut être lue et manipulée par un programme sur un ordinateur en tant que flot d'images annotées (date, numéro d'image, ...). Chaque image de ce flot est constituée de pixels (terme issu de la contraction des mots anglais « *picture elements* ») qui constituent autant de points caractérisant la taille/résolution de l'image.

Les capteurs d'images à l'origine de la vidéo caractérisent le nombre de pixels de large et de haut des images qu'ils engendrent ; ce sont des paramètres importants lors du choix d'un capteur puisqu'ils vont conditionner la résolution à laquelle on voit les détails de l'image, une fois numérisée. On trouve sur le marché des capteurs produisant des images vidéo de tailles très variées entre la centaine et un peu plus du millier de pixels de côté. Plus il y a de pixels, plus riche est l'information produite. Chaque pixel est en général représenté par une ou quelques valeurs entières qui codent son intensité (en noir et blanc et en caméra thermique/infrarouge) et, dans le cas de la couleur, sa chromaticité et sa saturation.



300x400 pixels

100x133 pixels

33x44 pixels

Figure 1 : résolution et capacité de reconnaissance

2.2. Comment comparer deux images d'une vidéo ?

Pour comparer deux images d'une vidéo, on s'appuie simplement sur le fait que les images successives d'une vidéo sont la plupart du temps toutes de la même taille, et donc on compare un à un les pixels dans l'ordre « ligne x colonne ». Tous les pixels comportant une valeur différente d'une image à l'autre appartiennent à la différence entre les deux images.

Comme les capteurs sont des équipements physiques imparfaits, les images successives d'un même plan vidéo (même éclairage, caméra immobile, ...) présentent en général de l'une à l'autre de très petites différences dues à des incertitudes ou des erreurs de mesure effectuées par le capteur : on regroupe tous ces aléas, quelle que soit leur cause, dans un terme qu'on appelle « le bruit ». Le bruit d'un capteur dépend de ce capteur mais aussi de l'optique, des composants électroniques de la caméra, de la chaîne de numérisation utilisée, jusqu'à disposer de l'image numérique, qui une fois élaborée est transportée avec des protocoles assurant son intégrité par des mécanismes de vérification (checksum), et de répétition en cas d'erreur à la transmission.



Figure 2 : à gauche, 2 images successives du balayage d'une façade, à droite leur différence, rapportée à une image moyenne « grise » afin de rendre visible les assombrissements et les éclaircissements.

Pour en revenir à la différence entre deux images, on « seuille » en général les écarts constatés par une constante placée très au dessus de la valeur moyenne du bruit, ce qui permet de comparer des différences plus certaines entre les deux images, c'est-à-dire des différences qui sont valides dans le monde réel, au-delà du bruit du capteur.

En pratique ce seuillage établit une limite à la capacité de détection du système qui n'est pas pénalisante lorsqu'on recherche de gros objets bien visibles, mais qui peut s'avérer très pénalisante lorsqu'on recherche un petit objet peu différent de son environnement. Lorsqu'on connaît l'effet recherché (par exemple l'éclaircissement furtif d'un point brillant affectant successivement deux ou trois pixels), on parvient toutefois à contourner cette difficulté pour détecter correctement (dans ce cas) un départ de missile.

2.3. Comment détecter un mouvement ?

Pour détecter un mouvement dans une séquence vidéo, on utilise en général des techniques issues de la recherche dans les disciplines du traitement d'images, du traitement du signal, et de l'intelligence artificielle.

La technique du **flot optique** permet la mise en œuvre d'une analyse globale du mouvement à l'aide d'une équation [3] reliant la variation d'intensité lumineuse en un point avec la vitesse de déplacement de ce point. Cette technique permet d'analyser des scènes dont la totalité de l'image est en mouvement, et d'y distinguer des objets en mouvement relatif les uns par rapport aux autres. Cette technique est utilisée en météo (analyse et mesure des mouvements des nuages, des cyclones), et en aide à la conduite de véhicules (détection et analyse des objets mobiles devant le véhicule : autres voitures, piétons).

La technique de l'**image de fond** [8], de son côté, se limite aux caméras en position fixe et permet, grâce à la mise en place et à la mise à jour permanente d'une image du fond vide, de distinguer des objets mobiles par différence à ce fond.

La technique des **points caractéristiques** [6], se concentre enfin sur la recherche de points caractéristiques dans l'image (cf. [2] : points de l'image aux caractéristiques fortement marquées : coins, bords, puis par extension centres de régions) et recherche leur correspondants d'une image à l'autre pour en déduire un mouvement, et pour regrouper ensemble les points proches ayant un mouvement cohérent.

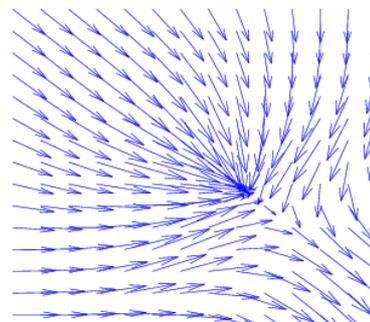


Figure 3 : le mouvement des pixels

Toutes ces techniques sont sensibles au bruit du capteur et aux mouvements de la scène (ex : vent sur la végétation, nuages au ciel, eau, ...) et nécessitent des traitements complémentaires, comme par exemple le filtrage, pour fonctionner de façon satisfaisante. Réciproquement, le filtrage est une limite à la capacité de détection d'un système, puisqu'on va filtrer aussi les petits objets, les mouvements très lents, ...

2.4. Comment détecter les visages ?

Une des applications récentes les plus fascinantes du domaine du traitement d'images concerne la détection et le suivi de visages à partir d'une vidéo.

Pour réaliser une telle application, on élabore des masques caractéristiques de visages, en travaillant à partir d'une base aussi importante que possible de visages, de façon à disposer d'une représentativité des échantillons. On programme alors une recherche rapide de zones de l'image dont les contrastes correspondent à ces masques. Une fois mis au point, le logiciel de traitement d'images balaie l'ensemble de l'image avec ces masques, et lève une détection de visage (carré rouge) lorsque la comparaison s'avère conforme sur la zone, après vérification de tous les masques appris [7].



Figure 3 : détection de visages à la volée : les visages partiellement occultés (le haut du visage au centre) et ceux qui sont trop petits (en haut de l'image) ou de profil (à gauche) ne sont pas encore détectés par ces algorithmes.

Les visages ainsi découverts peuvent alors faire l'objet d'un suivi d'une image à l'autre, suivi permettant de les retrouver plus rapidement et de pérenniser leurs caractéristiques à l'aide d'une moyenne calculée sur les observations effectuées de ce visage. La résolution du visage ainsi découvert et suivi peut aussi être améliorée par un traitement d'images, en exploitant et inversant plusieurs images successives pour en extraire une plus nette (surrésolution).

En pratique, les capacités de détection de ces algorithmes sont très limitées par la résolution de l'image : tant que les visages représentent plus de 10 voire 20 pixels de large, il est possible de les détecter, en deçà cela devient très difficile. La surrésolution mentionnée ici n'est possible que dans le cas d'objets quasi immobiles ou d'objets dont les contours sont très nettement estimés, elle ne peut donc être employée ici qu'après détection.

2.5. Comment reconnaître des objets ?

La reconnaissance d'objets est une application dans laquelle, avec un modèle, dont on dispose de une ou plusieurs prises de vues, et des données de recherche (des images d'une vidéo), on vise à automatiser la reconnaissance de l'objet par l'ordinateur lorsqu'il y est présent. L'objet peut y être vu à une distance différente, dans un autre contexte, un autre éclairage, tourné, ou renversé.

Pour reconnaître un objet assez plat (bâtiment) vu de très loin (satellite) ou un objet plat, on a utilisé une propriété géométrique des objets « plats » qui établit que dans une translation-homothécie-rotation (c'est la modélisation mathématique des transformations que peut subir un objet plat entre deux observations), les rapports des distances entre points caractéristiques (coins) d'un objet observé ne sont pas modifiés. Cette propriété est utilisée pour la reconnaissance d'objets par satellite depuis les débuts du traitement d'images, en vision industrielle pour faire reconnaître des pièces plates ou à face plate à un robot, ou plus simplement en reconnaissance de **plaques d'immatriculation** pour ensuite y appliquer un traitement de reconnaissance de caractères (OCR).

Là encore, l'écart de résolution entre les vues du modèle et les données de recherche ne doit pas être trop important : il doit permettre de retrouver des points qui peuvent être mis en correspondance : [5] démontre que la comparaison ne peut être effectuée qu'à l'échelle la moins résolue, comme pour l'œil humain : cf. les 3 photos de résolutions différentes en première page.

D'autres techniques ont vu le jour par la suite pour associer aux objets observés des propriétés d'un espace géométrique propre (eigenspace, eigenvectors, eigenfaces) telles que deux objets proches de cet espace caractérisent deux objets semblables du point de vue de la reconnaissance [4].

Ces dernières techniques sont utilisées dans la biométrie faciale pour reconnaître les individus dont les visages apparaissent nus et de face dans la vidéo.

Comme en pratique les visages ne sont pas parfaitement plats, pas nus (moustache, lunettes, écharpe, casquette, ...) et jamais parfaitement de face, des algorithmes supplétifs sont mis au point pour parvenir de façon performante à des résultats satisfaisants malgré ces difficultés.

Pour reconnaître de vrais objets non plats (3D), il faut manipuler des concepts en 3 dimensions, et les observer sur leurs différents côtés, sous différents angles de vue, et utiliser pour chacune toutes les caractéristiques de l'objet (couleurs, taches, points saillants) et leurs relations spatiales (proximité, distances, ...) de façon à reconstruire un modèle permettant de reconnaître effectivement l'objet sous un angle de vue et un grossissement quelconque [5].

Le traitement des modèles 3D est beaucoup plus complexe et coûteux en temps de calcul, si bien que pour le moment les applications disponibles sont très lentes, et souvent limitées à des classes d'objets homogènes (ex : les visages). Nous en sommes aux premiers résultats de la recherche dans le domaine, et le développement des applications industrielles est freiné par la puissance de calcul nécessaire à de telles opérations, notamment sur les objets complexes (tout objet du monde réel vu de près devient vite complexe !).

2.6. Un œil, deux yeux, ou plus ?

L'usage de 2 ou plusieurs caméras observant la même scène a également fait l'objet d'études et de développement de produits, car il permet soit de voir les objets sous plusieurs faces (caméras disposées autour d'un lieu), soit de voir les objets sous la même orientation, mais avec un certain décalage, d'autant plus important que l'objet est proche (effet stéréoscopique).

En démultipliant les techniques d'analyse vues précédemment sur les différentes caméras disponibles, quelques éléments d'analyse spatiale 3D sont en général possibles, mais ils sont néanmoins limités à l'observation de sujets isolés, et ne donnent pas d'applications particulièrement plus performantes, sauf sur le point de l'estimation de distance d'objets par la stéréoscopie, qui trouve des usages en conduite de véhicules (détection d'obstacles), ou pour des systèmes ayant besoin de connaître précisément la distance d'une cible (tir balistique).

2.7. Qu'est-ce qu'un ordinateur du commerce peut traiter ?



Figure 4 : le matériel reste la contrainte capitale pour l'analyse en temps réel

En une seconde de temps, un PC récent du commerce (Pentium Core 2 Duo X6800 à près de 3 GHz) sait traiter, pour un programme donné, 34×10^9 instructions entières, ou 13×10^9 instructions en réels (flottants). Ces mesures sont issues du test de performances dit de Dhrystone, et ne constituent qu'une des références utilisées parmi d'autres pour la mesure de performances, que nous choisissons ici pour illustrer notre propos.

Un flux d'images de 1000×1000 pixels en couleur à 25 images par seconde correspond à 75×10^6 d'entiers reçus par seconde. En rapprochant les deux chiffres, on constate que l'on dispose d'un temps de calcul maximal moyen de 200 à 500 opérations élémentaires par pixel si l'on veut pouvoir analyser tous ces pixels (tous les pixels de

toutes les images de la vidéo).

Cette limite est en fait très restrictive, et en pratique les algorithmes que nous avons cités ne peuvent être implémentés en si peu d'opérations (il en faut plutôt de l'ordre de 10.000, si on prend en compte tous les traitements qu'on applique à l'analyse d'une vidéo : décompression, filtrages, seuillages, différences, calculs de moyennes, variances...). Il convient donc, si on veut utiliser ce type de matériels, de procéder à des aménagements et de fortes optimisations pour pouvoir analyser la vidéo au mieux des possibilités de la machine.

En pratique on utilise des matériels de ce type, ou même 10 à 100 fois moins performants, avec des algorithmes très simplifiés, dans le cas où on souhaite traiter simultanément plusieurs voies vidéo,

ou intégrer le traitement dans un processeur dans la caméra. On se limite alors à ne traiter qu'une partie de l'image, et à sauter des images de la séquence, de façon à pouvoir donner une interprétation en temps réel.

3. Intégration dans des applications de sécurité et de défense

Pour intégrer ces technologies dans les applications de sécurité défense, on se fixe en général comme objectif et comme limite de pouvoir obtenir du traitement une interprétation aussi bonne que celle que produirait une personne très attentive observant la vidéo, c'est à dire la détection de 100% des événements recherchés, et l'absence totale de fausses détections.

Cet objectif implique deux limites :

- Les limites de la résolution de l'image et de la capacité de discernement de l'œil,
- Les informations non disponibles à la vidéo (objets cachés) ne sont pas visibles.

3.1. Intrusions, comportements

La détection de mouvements recouvre avec une même technologie un ensemble d'applications assez vaste :

- Détection d'intrusions (présence d'un mouvement dans une zone réputée vide ou interdite),
- Analyse de position et de déplacement à l'image (un objet entre, sort, possède une taille, une forme, une vitesse, ...),
- Détection d'objet immobile.



Figure 5 : à gauche, un suspect traverse une frontière en profitant du brouillard. A droite et en bas, le logiciel EAGLE (Evitech) signale l'incident.

Les applications les plus simples se cantonnent à signaler un rectangle entourant une silhouette dans des conditions de bonne visibilité et sous des changements de luminosité minimales, mais des algorithmes avancés sont aujourd'hui disponibles pour des applications au potentiel et à la fiabilité plus larges : nuit, brouillard, passages de nuages, neige, ...

Avec de tels outils, il est aujourd'hui possible de réaliser des applications de surveillance intelligente capables d'indiquer avec une précision du pixel la nature du mouvement observé, et ce même dans des conditions très difficiles.

L'analyse de forme permet la plupart du temps, sur une silhouette isolée, de la classifier automatiquement par sa forme (personne, véhicule, ...).



Figure 6 : détection de bateaux et analyse de vitesses (vecteur rouge). Image CEDIP. Le petit bateau numéroté I-3 a une vitesse élevée et se rapproche rapidement (trop ?) de nous.

Les applications visées en sécurité globale sont la protection de **frontières et de côtes**, la protection de **sites sensibles** (usines, centrales, ...), la protection de **véhicules** (notamment militaires), de **bateaux** (anti-collision, anti-piraterie), et **d'aéronefs**, la surveillance de **stocks** déposés en

extérieur, la détection d'objets déposés (objets mobiles devenus immobiles...), la surveillance de lieux isolés pouvant servir de cadre à des actions illicites, la surveillance des abords de prisons, l'annotation et le marquage de faits sur des séquences vidéos, pour l'aide au dépouillement ...

On insistera en particulier sur la **protection contre les attentats** perpétrés à l'aide d'engins rapides (véhicules, petits bateaux rapides, ...) que ces applications peuvent assurer en détectant la vitesse d'approche, et en **commandant automatiquement des mécanismes défensifs (herses, filets) ou offensifs** (tir de projectiles).



Figure 7 : le SGR-A1 de Samsung

D'autres asservissements sont possibles, comme le préfigure le fusil automatique Samsung SGRA1 qui devrait protéger automatiquement la frontière nord de la Corée du Sud contre des incursions à pied ou de véhicules isolés, dans l'année qui vient. Cette arme s'apparente à une mine, et même à une mine anti-personnel, dans la mesure où l'on confie une action de tir (12 mm) sur une cible personne ou véhicule à un engin, non contrôlé par un homme (ce dernier étant seul capable de rendre compte de ses décisions). Elle est donc interdite d'usage en zones civiles par les pays signataires des traités relatifs aux mines anti-personnels. Ce système présente toutefois l'avantage d'être activable et désactivable selon les crises : elle peut donc n'être déployée qu'en zone militaire sécurisée, ou en bande frontière, et en périodes de crise.

Les **menaces** auxquelles répondent ces outils sont le **terrorisme** (et notamment le repérage des lieux, qui devient détectable s'il semble anormal), le **vol** (matériels, propriété intellectuelle), les **trafics, l'immigration clandestine, ...**

La limite à ces applications concerne l'analyse des groupes (hommes, voitures, ...), sur lesquels la segmentation en individus est un problème complexe, et la détection et l'analyse des objets portés (sacoche, fusil, revolver, ...) pour lesquels il est très difficile de discerner un objet porté de son porteur, même si quelques travaux ont déjà été abordés dans ce sens.

Une fois l'interprétation établie, on utilise alors un simple système de gestion de règles (actions permises et/ou actions interdites) pour décider d'émettre l'information par réseau informatique, par contacteur sec, ou par réseau hertzien (y compris GSM : ainsi, la valise de détection Jaguar d'Evitech transmet des alarmes par mail, SMS, et MMS).

3.2. Surveillance publique, enquêtes

La **reconnaissance** des individus (biométrie faciale) et des véhicules (plaques minéralogiques) est un point central des enquêtes et de la surveillance publique (contrôle d'accès, contrôles volants dans les transports) et trouve des outils de traitement et d'analyse d'images qui sont aujourd'hui effectués « à la volée » et de façon « non coopérative » par les logiciels intégrant les techniques évoquées plus haut. Ces outils permettent de détecter **une personne ou un véhicule recherché**, et de disposer d'informations relatives à un auteur de faits (circulation, acte malveillant, ...).

Les **menaces** auxquelles répondent ces outils sont le **terrorisme**, la poursuite d'activités **d'individus recherchés**, la capacité de **retrouver les auteurs** de faits suspects ou délictueux.

Ces applications sont fortement réglementées et limitées par le droit, notamment en France (CNIL), et empiètent évidemment sur les libertés publiques. On peut tenter d'arguer de l'avantage qu'il y a à ne garder et ne faire sortir de l'analyse que les informations (images, interprétations) relatives à l'action ou la personne incriminée, mais l'erreur et l'abus sont toujours possibles.

Elles peuvent également être utilisées sur des bandes vidéo saisies et conservées après un attentat, pour lesquelles la loi prévoit un délai de conservation plus long et permet une analyse plus fouillée que pour les bandes enregistrées sans fait marquant.

3.3. Améliorations de l'image

Les services de sécurité opérationnels sont demandeurs de technologies capables d'améliorer les images, pour que les fonctions d'analyse aient de meilleurs résultats, comme par exemple ici :



Figure 8 : de gauche à droite : l'image brute, l'image rendue plus lisible, et enfin avec l'amélioration de netteté du fond (source : Eagle V2.1)

3.4. Capacités et limites des systèmes de vidéo-surveillance intelligente

Comme on l'a vu au fil du texte, trois limites essentielles s'opposent aux capacités de détection et à l'efficacité de ces systèmes : la **résolution** de l'image de référence ou de l'image à comparer, la **puissance de calcul** du calculateur support, et la **performance** de l'algorithme (orientation à détecter tel ou tel phénomène, ...).

Les capacités de ces logiciels en termes de taux de bonnes détections et de fausses détections doivent être établies et connues de leurs utilisateurs. Des outils ergonomiques d'élaboration de diagnostic circonstanciés doivent y être intégrés (annonce de la qualité de la mesure sur laquelle un diagnostic a été établi, explication du raisonnement tenu par le système, ...).

Les systèmes de vidéo-surveillance intelligents sont encore jeunes, et la forte attractivité qu'ils exercent sur les utilisateurs conduisent parfois leurs auteurs (et leurs utilisateurs) à négliger l'intégration en leur sein de contrôle du domaine de validité des faits interprétés. Ainsi en Juillet 2005, peu après les attentats du métro de Londres, un tel logiciel, directement transplanté de l'Université dans le métro, avait interprété comme suspect un brésilien voyageant avec un manteau de fourrure, ce qui a placé les services de police dans un état de stress tel que le voyageur, n'ayant pas compris les sommations, a été malencontreusement abattu. Un autre de ces logiciels, chargé d'empêcher la densité des pèlerins de dépasser une limite critique dans un lieu de culte, n'a pu empêcher quelques 300 morts en 2006.

Plus trivialement, on voit les opérateurs des centres de surveillance déconnecter les systèmes qui « sonnent » trop souvent et les perturbent à mauvais escient.

En tant qu'acteurs de cette discipline, nous travaillons avec acharnement à faire évoluer la maturité de nos systèmes et celles de leurs utilisateurs. Nous avons pu ainsi réduire à un taux de une fausse alarme par jour et par caméra ; en moyenne (0.5 en ville et 1.5 en campagne), les fausses détections d'intrusions sur des sites protégés par vidéo-surveillance intelligente. Les cas résiduels sont si divers (regroupements d'oiseaux, reflets de gros animaux dans des flaques d'eau, écharpes de brume, rouleau de plastique ou buisson arraché transporté par le vent,...) qu'il s'avère impossible de descendre de façon significative sous ces niveaux, sauf à combiner différentes voies de détection ensemble (ce qui est possible aujourd'hui dans le domaine de l'intrusion, mais pas dans celui de la lecture de plaques à distance, ni dans celui de la biométrie non coopérative).

4. Perspectives

Quels seront demain les évolutions des systèmes de traitement d'images qui pourront bénéficier aux applications de sécurité ? Il est difficile de répondre à cette question, même si on peut identifier quelques exemples d'axes de recherche et les demandes de ce marché :

- L'analyse d'objets portés et des gestes : quelques travaux de recherche ont déjà abordé ces thèmes [1] qui intéressent beaucoup la surveillance publique pour y détecter des armes.

- L'analyse de groupes d'individus et de foules : quelques applications d'analyse statistique du comportement global de la foule vue comme un fluide ont été déjà réalisées et commercialisées. Elles détectent l'arrivée d'un train (mouvements aux portes), les réactions de fuite (trou dans la foule), mais souffrent toutefois de ne pas traquer individuellement les individus dont on ne peut connaître le comportement individuel. La détection de bagarres, et d'objets portés dans la foule sont notamment des thèmes intéressants les forces de l'ordre.
- L'analyse et la reconnaissance d'objets 3D, par exemple pour classifier de façon plus fiable les véhicules.

5. A Propos de ...

Au sein du pôle **SYSTEM@TIC Paris Région**, Evitech (www.evitech.com) étudie et développe des systèmes de traitement d'images innovants pour la sécurité défense. Ses travaux de recherche et développement, menés en coopération avec l'IEF, l'ENS Ulm, l'INRIA, le GET ont été et sont financés notamment par la **DGA**, le **CG'92**, l'**ANR**.

P. Bernas est Ingénieur Civil des Mines et Docteur en Informatique. Après un début de carrière au sein des groupes aéronautiques français (Thalès avionics), et de sociétés spécialisées en logiciels complexes, il est depuis 2005 fondateur et dirigeant d'Evitech.

6. Bibliographie

- [1] Haritaoglu, Harwood, Davis : "W4, a real time system for detecting and tracking people", *3th conf. on face and gesture recognition, 1998*.
- [2] C. Harris, M. Stephens : "A combined edge and corner detector", *4th alvey conference, Manchester, 1998*, pp. 189-192.
- [3] Horn, Schunck : « Determining optical flow », *Artificial Intelligence 17, 1981*, 185-203.
- [4] Pentland, Moghaddam, Starner, « View based and modular eigen spaces for face recognition », *Conf. CVPR, 1994, Seattle*.
- [5] Rothganger, Lazebnik, Ponce, Schmid : « 3D Object modeling and recognition using local affine-invariant image descriptors and multi-view spatial constraints », *Conf. CVPR 2003*.
- [6] Schmid, Mohr : « Local greyvalue invariants for image retrieval », *conf. IEEE TPAMI 19, 1997*.
- [7] Viola Jones : "Rapid object detection using a boosted cascade of simple features", *Conf. CVPR 2001*.
- [8] Wren, Azarbayejani, Darell, Pentland : « Pfänder : real-time tracking of the human body », *Conf. IEEE TPAMI, 19, 1996*.