

Analyse de séquences vidéo : le projet ANR KIVAOU

Paul WELTI¹

¹Sagem Sécurité, (coordonnateur du projet) Direction R&T et Business Development, 14-18 rue de la Vanne, 92120 Montrouge

paul.welti@sagem.com

Résumé – Le projet ANR CSOSG2007 KIVAOU vise à développer des outils innovants d'analyse vidéo dédiés à deux problématiques:

1/ Outil mobile (valise) d'identification et indexation biométrique faciale portable par analyse temps réel vidéo
2/ Plate-forme d'analyse de vidéos multiples enregistrées lors d'un événement, utilisant la synchronisation de vidéos, l'extraction de signatures pour les personnes, et la constitution de trajectoires. Le but est de permettre ou de faciliter une analyse a posteriori des données enregistrées en un lieu et pendant une même période à des fins d'investigations
Le projet KIVAOU réunit Sagem Sécurité, EVITECH, FACING-IT, le Ministère de l'Intérieur, l'Institut Telecom , et ARMINES et propose des approches innovantes basées sur la combinaison de briques de base maîtrisées par chacun des partenaires: biométrie, analyse d'image, analyse vidéo, tracking, synchronisation, besoins IHM utilisateurs etc. Il comporte une phase de test auprès des utilisateurs finaux pour valider ou améliorer la pertinence des outils.

Abstract – The KIVAOU project aims at developing innovative tools dedicated to two main goals:

1/ Mobile facial identification and biometric indexation tool (suitcase), with real-time video analysis
2/ Platform for offline multiple video analysis of an event (one place, one period).. These video are recorded during a particular event, and then collected for investigation purpose. After synchronisation between different sources, an analysis is performed in order to extract individual signatures, and to compute person's trajectories. The goal is to enable or to facilitate an offline analysis, for investigation purpose.

The KIVAOU project brings together Sagem Sécurité, EVITECH, FACING-IT, French Ministry of Interior, The Institut Telecom, and ARMINES, and proposes innovative approaches based on the combination of components already mastered by each partner: biometrics, image analysis, video analysis, tracking, synchronization, users HMI needs, etc. It includes a test phase with end-users for assessing and improving the relevance of the developed tools.

1. La multiplication de sources vidéo à analyser

Depuis plusieurs décennies, la vidéosurveillance se développe, aussi bien dans les lieux publics, les zones sensibles, ou dans les entreprises et chez les particuliers. Historiquement, l'usage de la vidéosurveillance a d'abord consisté en l'observation des données par un agent, et en l'enregistrement des vidéos pour les consulter en cas de besoin. Aujourd'hui, la quantité de données enregistrées est telle que l'examen des vidéos, que ce soit en temps réel par un opérateur, ou en différé via des enregistrements est très fastidieux. De fait, l'utilisation de ces vidéos comporte un coût caché qui peut être très important. Par exemple, lors des investigations conséquentes aux attentats du métro de Londres, une équipe importante a été mobilisée pendant un temps relativement long pour dépouiller « manuellement » les enregistrements relatifs à cette enquête. A ce coût important s'ajoute le manque d'efficacité d'une analyse visuelle longue, répétitive et fastidieuse, qui rend possible la non-exploitation de renseignements qui auraient pu être précieux.

Parallèlement à cela, il est apparu que l'évolution récente des technologies de traitement d'images, jointe aux

gains de puissance de traitement des ordinateurs, rendaient aujourd'hui possibles des applications intelligentes de vidéosurveillance mêlant des fonctionnalités de surveillance générale (analyse de scène, détection des individus, trajectoires...) et des fonctionnalités plus ciblées (reconnaissance faciale, amélioration de la résolution).

2. L'état de l'art et les verrous de l'analyse automatique

2.1 Reconnaissance faciale

Plusieurs expériences ont montré la difficulté d'utiliser de manière opérationnelle des techniques de détection automatique d'évènements complexes dans des infrastructures équipées de vidéosurveillance. On peut citer notamment les expériences menées par la police de Tampa en 2001 [1], ayant pour but d'identifier automatiquement le visage de personnes recherchées dans des lieux publics. Cette initiative n'a pas donné suite en raison de performances inadéquates à une utilisation opérationnelle.

La police fédérale allemande a lancé en 2007 une campagne d'évaluation expérimentale de la reconnaissance faciale automatique de personnes recherchées [2]. Il s'est avéré que pour une liste de 200 personnes recherchées, la probabilité de retrouver l'une d'entre elles passant devant la caméra était de 60%. Ce chiffre est atteint pour un taux de fausse alarme de 0,1% dans des conditions d'éclairage de jour. Ces taux de détection et de fausse alarme sont à la fois très encourageants et insuffisants pour une utilisation de reconnaissance automatique.

Les tests FRVT (Face Recognition Vendor Test) organisés aux USA par le NIST (National Institute for Standards and Technology) en 2002 et 2007 montrent que les taux d'erreur de la reconnaissance faciale ont été divisés par 10 entre 2002 et 2007 [3]. Cependant, ces tests sont réalisés en laboratoire avec des images d'une qualité ne correspondant pas aux données issues de systèmes de vidéosurveillance déployés sur le terrain. L'un des verrous important pour l'utilisation de la reconnaissance faciale interactive est donc la qualité de l'enregistrement, qui est fonction de la caméra utilisée, du format des données utilisé (compression...), du contrôle de l'environnement de capture, et de la coopération de l'individu (illumination, pose, expression, etc.). Les très bons résultats annoncés par le NIST correspondent à des conditions proches de la prise de photographie d'identité, conformes aux standards recommandés par l'OACI pour les documents de voyage.

2.2 Suivi d'individus et signatures d'objets mobiles

Outre l'identification faciale des personnes, le suivi (ou « tracking ») et l'extraction de signatures d'individus sont des fonctionnalités souvent recherchées. Il existe depuis longtemps des fonctions de tracking, qui ont été utilisées dans d'autres domaines (poursuite de cibles par exemple). Cependant, la robustesse par rapport à la complexité des scènes et à la qualité des données est toujours un point difficile.

Le projet européen ISCAPS (Integrated Surveillance of Crowded Areas for Public Security) du PASR a montré que la détection de comportements anormaux dans des vidéos, ne fonctionne actuellement que pour des scénarios simples dans des environnements à la complexité réduite. On peut notamment citer la détection d'intrusion dans un périmètre donné, ou la détection de bagage abandonné dans un local avec peu de mouvements. Ces fonctions utilisent le tracking et l'extraction de signatures.

Un verrou important dans ce domaine est également la capacité à traiter des sources hétérogènes et simultanées, c'est-à-dire la calibration et la synchronisation d'un ensemble de caméras afin d'extraire des signatures stables, et de pouvoir mettre en correspondance temporelle plusieurs enregistrements.

2.3 Amélioration de la qualité

Un problème délicat relatif à un enregistrement donné, est de post-traiter les données afin d'en améliorer la qualité. En particulier, la définition d'un critère mesurable de qualité est un point clé permettant de comparer les mérites de différentes méthodes.

La vidéo permet l'observation d'objets du monde réel à travers une grille pixelique (celle du capteur). Lorsque ces objets se déplacent légèrement par rapport à cette grille, on relève de petites modifications des valeurs des pixels. Ces variations permettent, lorsqu'un nombre suffisant d'observations sont disponibles, et lorsqu'on détermine précisément la position réelle de l'objet, d'établir, sur la base de ces observations, une image à la résolution plus fine que celle de la grille pixelique (c'est la sur-résolution). Les problèmes de sur-résolution entrent dans la classe des problèmes inverses.

3. Le projet KIVAOU

Le projet KIVAOU est un projet de recherche coopératif supporté par l'Agence Nationale pour la Recherche, dans le cadre de l'appel CSOSG 2007. Il commence début 2008 pour une durée totale de deux ans.

Le consortium KIVAOU est composé d'un industriel, d'une PME, d'une « jeune pousse », de deux laboratoires et d'un utilisateur public :

- Sagem Sécurité,
- EVITECH,
- FACING-IT,
- INSTITUT TELECOM (anciennement GET),
- ARMINES,
- Le Service des Technologies de la Sécurité Intérieure, Ministère de l'Intérieur.

Le projet KIVAOU vise à étudier et démontrer la faisabilité de nouveaux systèmes et outils pour la surveillance, avec identification et authentification de personnes, appartenant à une liste d'intérêt de taille réduite, de manière non coopérative, sans contact, et à distance (thématique 1). Au cours de ces traitements, des personnes non connues via une liste de personnes connues seront détectées. Ces passages seront indexés, c'est-à-dire que la signature biométrique permettra de reconnaître cette personne lorsqu'elle repassera par la suite. La thématique 2 désigne l'analyse a posteriori de scènes enregistrées par un ensemble de caméras, à des fins d'investigation ou d'enquête. Ces deux thématiques ne sont pas disjointes : des éléments de solutions seront communs aux deux.

Le but du projet est de rechercher, d'inventer, de mettre au point, et d'intégrer différentes technologies de traitement d'images et reconnaissance des formes pour réaliser et mettre en œuvre les deux plates-formes démonstratrices suivantes, utilisables pour les contextes décrits ci-dessus :

- **Thématique 1 :** Outil (valise) de surveillance embarquée permettant de surveiller un lieu (rue, entrée d'immeuble), d'indexer au fil de l'eau tous les passants et d'enregistrer leur biométrie faciale, de la comparer à une base de données biométrique embarquée (à la volée, ou après la surveillance, en complétant la base de données biométriques). Cette valise pourra être utilisée en différé pour analyser des enregistrements et produire un bilan global de l'enregistrement analysé.



FIG 1 : SCHÉMA GÉNÉRAL DE L'OUTIL

- **Thématique 2 :** Plate-forme de dépouillement d'un ensemble de vidéos associées à un même lieu et un même temps pour reconstituer les événements qui s'y sont produits et en retrouver les acteurs. Cette plate-forme doit intégrer une ou plusieurs bases de données biométriques de visages d'individus recherchés ou connus, et récupérer et traiter des vidéos hétérogènes, fixes ou mobiles. L'opérateur doit également pouvoir localiser l'emplacement des caméras dont ces vidéos sont issues sur un plan associé au lieu. Les outils associés à cette plate-forme doivent alors permettre de re-synchroniser ces vidéos (à la main, ou automatiquement à l'aide d'événements observables simultanément depuis plusieurs vidéos, comme une explosion, le passage d'un bus), de constituer une base avec l'identification des passants et enfin de reconstituer les trajectoires de ces derniers.

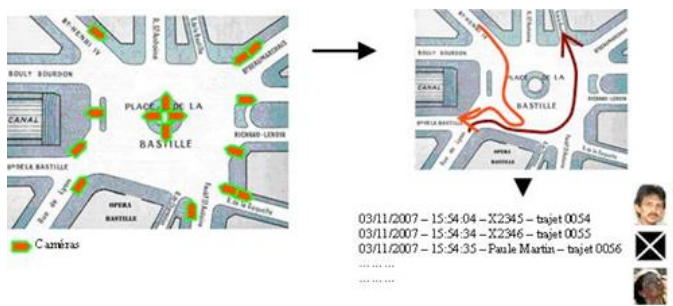


FIG 2 : SCHÉMA GÉNÉRAL DE LA PLATE-FORME

Lorsqu'une même personne est reconnue sur une caméra et détectée sans être reconnue sur une autre caméra, le suivi de trajectoire doit incorporer à la trajectoire (et associer au visage qui a été vu) les déplacements ultérieurs de cette personne (la technologie mise en œuvre se doit d'être robuste aux occlusions et ainsi de proposer un suivi même lorsque le visage n'est plus visible et/ou reconnu).

Le projet KIVAOU vise à poursuivre les travaux de recherche et de développement menés au sein des entités de R&D des participants dans les domaines concernés de l'analyse intelligente d'images pour faire progresser en qualité et en fonctionnalités ces technologies à travers deux démonstrateurs.

Les améliorations de qualité sont notamment attendues de la mise en œuvre de nouveaux algorithmes sur des images améliorées, tandis que les progrès de fonctionnalités sont attendus des outils qui seront destinés à des missions précises d'analyse de scène, soit pour surveiller à la volée un point de passage à l'aide d'un outil mobile, soit pour dépouiller a posteriori les images issues de différents capteurs filmant une même scène, et aider l'opérateur à reconstruire les trajectoires et retrouver les identités des acteurs présents (et associer les trajectoires aux identités).

Les travaux menés dans le cadre du projet sont les suivants :

- Elaboration avec la Police des besoins, des scénarios représentatifs, et des spécifications ;
- Recherches en amont sur les sujets difficiles : interopérabilité entre bases biométriques, biométrie faciale avec images partiellement masquées, sur-résolution d'images, constitution de signatures globales multi-facettes de silhouettes ;
- Etude sur l'acceptation juridique et sociale de tels outils ;
- Développement de logiciels illustrant les fonctionnalités des outils et systèmes visés ;
- Intégration sur des matériels de démonstrateurs ;
- Evaluation des utilisateurs, et retours d'expérience ;

Références

- [1] http://www.biometrie-online.net/dossiers/technique/visage/drawing_blank.pdf
- [2] http://www.bka.de/kriminalwissenschaften/fotofahndung/pdf/fotofahndung_final_report.pdf
- [3] www.frvt.org
- [4] F. Humblot, Thèse de doctorat menée au GIP (DGA).