

Analyse de séquences vidéo : le projet ANR KIVAOU

Dijana Petrovska-Delacrétaz, Mohamed-Anouar Mellakh, Dianle Zhou (Institut TELECOM - SudParis, dijana.petrovska@it-sudparis.eu); Gérard Chollet, Christophe Riedinger, Nadhir Khemakhem (Institut TELECOM - ParisTech, gerard.chollet@telecom-paristech.fr); Serge Beucher, Louise Naud (Armines - CMM, serge.beucher@ensmp.fr); Bogdan Stanciulescu (Armines-CAOR, bogdan.stanciulescu@mines-paristech.fr), Pierre BERNAS (Evitech, pbernas@e-vitech.com), Bouchra ABBOUD, (Facing-It, bouchra.abboud@facing-it.net),

Samuel VINSON¹

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

samuel.vinson@sagem.com

Résumé : Le projet ANR CSOSG2007 KIVAOU vise à développer un démonstrateur comprenant des outils innovants d'analyse vidéo dédiés à deux problématiques : **1**) Un dispositif mobile (valise) d'identification et indexation biométrique faciale portable pour une analyse temps réel vidéo et **2**) Une 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. Il propose des approches innovantes basées sur la combinaison de briques de base maîtrisées par chacun des partenaires: biométrie, analyse d'image, analyse vidéo, suivi, synchronisation, besoins IHM utilisateurs, etc. Il comporte une phase de test auprès des utilisateurs finaux pour valider ou améliorer la pertinence des outils proposés.

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, and **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 purposes. 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.

The KIVAOU project brings together Sagem Sécurité, EVITECH, FACING-IT, French Ministry of Interior, Institut TELECOM, and ARMINES. It 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.

La multiplication de sources vidéo à analyser

Depuis plusieurs décennies, la vidéosurveillance se développe, aussi bien dans les lieux publics, les zones sensibles, 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 consécutives 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).

Dans le projet KIVAOU nous proposons de faire un pas en avant en associant les technologies diverses liées à la vidéosurveillance et à la biométrie, qui vont nous permettre de traiter un flux vidéo de la manière la plus appropriée. Le reste de cet article est organisé comme suit. Dans la Section 2 se trouve un aperçu des verrous à lever. La Section 3 présente les objectifs et les partenaires du projet

KIVAOU. La Section 4 donne plus de détails sur des briques de base implémentés dans le démonstrateur, tandis que la Section 5 introduit le démonstrateur. Les conclusions et perspectives se trouvent dans la Section 6.

L'état de l'art et les verrous de l'analyse automatique des vidéos

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. En présence d'un flux vidéo à analyser, on se trouve devant une multitude de tâches à résoudre : quel prétraitement utiliser, comment segmenter puis suivre des zones en mouvement, de quelle manière extraire des informations pertinentes pour faire de la classification des objets détectés, décider jusqu'à quel niveau de détail peut-t-on pousser notre analyse, etc. etc. Chacune de ces techniques a ses propres verrous à lever.

Par exemple, une étape cruciale pour une bonne analyse des vidéos est la bonne détection des régions d'intérêt. La difficulté consiste dans l'augmentation du taux de bonnes détections en même temps que la diminution du nombre de fausses alarmes. Un nombre trop important de fausses alarmes, même accompagné d'un taux de détection élevé, posera des difficultés de compréhension pour l'utilisateur. Un des objectifs de ce projet est l'amélioration de ces paramètres.

Segmentation de zones en mouvement

La segmentation de zones en mouvement dans des champs de caméras fixe peut être effectuée suivant trois principes :

- Flot optique
- Soustraction de fond
- Soustraction trame à trame

C'est cette dernière approche qui a été choisie pour KIVAOU.

Suivi d'individus et signatures d'objets mobiles

Outre la segmentation et la détection, 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. Lorsque dans les régions d'intérêts détectés une résolution suffisante est disponible, le traitement peut être approfondi en s'intéressant aux visages.

Reconnaissance faciale

Plusieurs expériences ont montré la difficulté d'utiliser de manière opérationnelle des techniques de reconnaissance faciale dans des infrastructures équipées de vidéosurveillance. On peut citer notamment les expériences menées par la police de Tampa en 2001 [www-biometrie], 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 [www-bka]. 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 par le NIST (National Institute for Standards and Technology) montrent que les taux d'erreur de la reconnaissance faciale ont été divisés par 10 entre 2002 et 2007 [www-frvt]. Cependant, ces tests sont réalisés sur des bases de données enregistrées avec des scénarios (et la qualité des images) qui ne correspondent 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 (Organisation de l'Aviation Civile Internationale) pour les documents de voyage.

Depuis mai 2008, le NIST propose une nouvelle campagne de tests ouverts MBGC (Multiple Biometric Grand Challenge) [www-mbgc] qui a cette fois-ci aussi pour objectif d'évaluer les algorithmes biométriques dans des conditions non contrôlées (pose non frontale, vidéo de personnes qui marchent dans la rue ou traversent un portique métallique...). Ces conditions étant proches de celles du projet KIVAOU, deux partenaires (Institut TELECOM et Sagem Sécurité) ont décidé à titre privé de participer à ces tests pour évaluer leurs algorithmes de reconnaissance biométrique. Une première campagne d'évaluation (MBGCv1) a eu lieu en décembre 2008 et une seconde (MBGCv2) vient d'avoir lieu en décembre 2009. Ces évaluations montrent que dans le cas du scénario «Portique» ou «Portal challenge» les taux de vérification excellents sont atteints. Par contre, les algorithmes soumis pour le « Video challenge » en 2008, présentent dans le cas le plus difficile - qui ressemble le plus à des données de type vidéosurveillance- des performances qui sont proches d'un tirage aléatoire. Ces exemples illustrent bien le fait que dans de telles conditions, d'autres informations deviennent nécessaires pour une exploitation à des fins de reconnaissance de personnes. Pour pouvoir traiter la variété présente dans ces séquences, il faut passer par une détection de la silhouette, et aussi posséder un détecteur de visage pour des poses autres que de face. De plus, lorsque les comparaisons se font dans des conditions de comparaison face-profil, les méthodes «classiques» 2D sont insuffisantes, et *l'utilisation des informations 3D* est nécessaire. Un des objectifs de KIVAOU consiste à déterminer des conditions qui permettent de faire de la reconnaissance faciale à partir des séquences vidéo.

Amélioration de la qualité avec la super-résolution

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 pixellique (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 pixellique (techniques dites de sur- ou super-résolution). Ainsi dans le cas où l'on possède des séquences d'images basse résolution dégradées et/ou affectées de repliement (espacement entre pixels supérieurs à la finesse de l'image), on pourrait appliquer les *techniques de super-résolution* ([capel-cvpr-1998], [gunturk-ip-2003], [park-sp-2003], [van-eeken-eurasip-2007]) qui ont pour but de restaurer une image haute résolution à partir de plusieurs images basse résolution.

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 a débuté en février 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.

Il 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. Elle est illustrée dans la Figure 3.1.
- 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



Figure 3.1 : Schéma général de l'outil valise

- 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 reconstruire les trajectoires de ces derniers. Elle est illustrée dans la Figure 3.2.

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 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 ces deux outils.

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

destinées 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 de tels outils ;
 - Développement de logiciels illustrant les fonctionnalités des outils et systèmes visés ;
 - Intégration sur des matériels de démonstration ;
 - Evaluation des utilisateurs, et retours d'expérience ;
- Enfin, pour faciliter l'échange de documents entre les partenaires durant le projet, une plate-forme de travail coopératif a été mise en place.

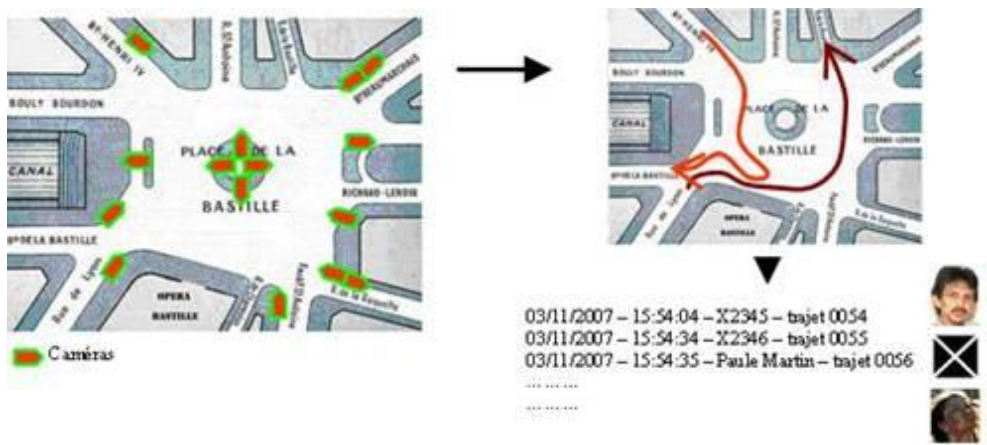


Figure 3.2 : Schéma général de l'outil valise

Briques de base implémentés dans le démonstrateur

Segmentation de zones en mouvement

Le Centre de Morphologie Mathématique (CMM-ARMINES) développe un module de segmentation automatique des zones en mouvement dans les champs de caméras. Ce module, basé sur un détecteur de contours en mouvement dénommé MCM « Motion Contour Measurement [biancardini-orasis-2005] est défini de la manière suivante :

$$MCM(t) = \inf(|g(t) - g(t-i)|, |g(t) - g(t+i)|)$$

Où g représente le module du gradient de l'image, calculé sur trois images successives

Cet opérateur possède une bonne localisation des contours en mouvement, et ce de manière relativement indépendante à la vitesse des objets. Il est également robuste au bruit aléatoire et possède une certaine indépendance à la texture (Figure 4.1).



Figure 4.1 : Trame au temps t et MCM associé (Images : PETS 2006).

Cependant, cet opérateur présente le défaut de ne pas fermer les contours en mouvement. C'est pourquoi il est complété par une opération de fermeture géodésique consistant à «réparer» les contours du MCM en reconstruisant les portions de contours absentes à l'aide des contours de toutes les régions de la scène (Figure 4.2)

L'information qui nous intéresse est disponible dans l'ensemble des contours de la scène. Afin de déterminer les régions en mouvement à partir des contours trouvés précédemment, on calcule la ligne de partage des eaux sur ces contours en mouvements obtenus après fermeture géodésique (Figure 4.3).



Figure 4.2 : MCM complété par les contours de la scène toute entière.



Figure 4.3 : Ligne de Partage des Eaux ([beucher-phd-1990]) du MCM fermé géodésiquement.

Module de détection, classification et reconnaissance de personnes (CAOR)

Le module de détection de personnes proposé par CAOR-ARMINES rassemble une palette de fonctionnalités temps-réel qui sont intégrés dans la plateforme de dépouillement ainsi dans la valise mobile.

Le module comporte des fonctions relatives à la détection-classification et à la reconnaissance des silhouettes (deux liées directement à la détection/classification et deux autres à la reconnaissance). La fonction '*detectionSilhouettes*' explore l'image d'entrée et localise les silhouettes des personnes en les mettant dans une boîte englobante rectangulaire. Une deuxième fonction, '*classificationMobile*', sert à la classification d'un mobile (personne ou non-personne, véhicule ou non-véhicule) présent dans une région de l'image. Une fonction d'initialisation des structures de détection/classification est appelée avant tout processus,

ainsi qu'une fonction de fin de cycle qui est jouée afin de libérer la mémoire.

Pour la reconnaissance de personnes à partir de leur silhouette, la fonction *'processFrameKPSignature'* sert à extraire et à construire une signature, à partir de l'image d'une personne donnée. Cette image pourrait être fournie automatiquement par un des algorithmes de détection ou par tout autre algorithme. Par la suite, cette signature est utilisée par la fonction de reconnaissance *'getSectionBestMatch'* dont la tâche est de trouver les occurrences de cette personne dans une section vidéo. Comme dans le cas de la détection/classification une fonction d'initialisation du processus de reconnaissance est requise ainsi qu'une phase finale de suppression des structures.

Trajectoires et référentiel

Une des spécificités du projet KIVAOU consiste à retrouver les éléments propres à chaque individu observé dans la scène (un lieu, un intervalle de temps) en construisant des associations entre visage, trajectoire, et portions de séquences vidéo comportant la personne. A ce titre, EVITECH intervient en tant qu'acteur principal sur l'élaboration de la trajectoire et les portions de séquences vidéo pertinentes.

A partir de son socle technologique ([bernas-avirs-2008], [schwerdt-avsbs-2005]) basé sur Linux, EVITECH a produit un moteur de traitement d'images afin de détecter, de segmenter, et de suivre les silhouettes des personnes présentes dans la scène KIVAOU à partir des séquences vidéo, de les localiser dans l'espace (position sur le plan calculée d'après la mesure de la distance à la caméra et l'angle avec l'axe central de la caméra) afin de pouvoir les positionner sur le plan de la scène. Ce moteur dispose également d'une primitive permettant de déterminer des ruptures de séquence (cas d'une caméra orientable qui aurait été manipulée par un opérateur au cours de l'enregistrement la concernant, par exemple un mouvement entre deux positions fixes), ce qui est nécessaire pour configurer une calibration adaptée de chaque scène observée, entre les déplacements. Il est mis à disposition sous forme de DLL (« Dynamic Linkable Library ») Windows y donnant accès, la DLL étant elle-même intégrée dans le logiciel principal KIVAOU.

Par ailleurs, une recherche a été entreprise sur la séparation des groupes de personnes, la super-résolution des visages, dont les résultats seront intégrés dans une prochaine itération du projet.

Reconnaissance de visages

Un des objectifs du projet KIVAOU est de pouvoir identifier des visages des personnes dans des séquences vidéo. Plusieurs partenaires travaillent sur cette problématique et développent des modules de reconnaissance faciale. Une première étape consiste à détecter les visages. Ensuite il faut traiter ces images de visages (par exemple par une normalisation géométrique)

pour ensuite en extraire les caractéristiques pertinentes qui vont nous permettre de faire des comparaisons. Dans la majorité des méthodes actuelles on effectue une comparaison entre deux visages : des visages préalablement enrôlés dans la base et des visages dont il faut vérifier leur identité ou les identifier parmi une liste de personnes existantes. Lorsqu'en phase de test on a à disposition une séquence vidéo, il faut définir une stratégie pour traiter cette séquence. Les partenaires FACING-IT et TELECOM-SudParis ont fourni des DLL de reconnaissance faciale.

Le système « open-source » de reconnaissance de visages de TELECOM-SudParis

Les DLL de TELECOM-SudParis relatives à la reconnaissance faciale (intégrés dans le démonstrateur KIVAOU) sont dérivées du système SUD-FROG (« *SudParis Face Recognition with Gabor features* ») mis à disposition de la communauté scientifique en tant que logiciel libre [www-sudfrog]. Ce système intègre différents modules utiles pour de la reconnaissance faciale, qui ont été testés aux évaluations NIST MBGCv1 et MBGCv2 [www-mbgc]. Parallèlement aux code « open-source », sont aussi fournis des paramètres utilisés pour ces deux évaluations. Ainsi ces résultats sont reproductibles et pourront servir de point de comparaison.

Le système SUD-FROG est composé de modules suivants : détection de visage et de points caractéristiques ; prétraitements ; extraction de caractéristique pertinentes des visages et création des modèles ; puis d'un module dédié à la comparaison. Le module *'detection'* utilise soit l'implémentation AdaBoost de OpenCV [www-opencv], soit la méthode décrite dans [zhou-btas-2009] basée sur les ASM « Active Shape Model » de [cootes-tech-report-2004]. Dans le module *'prétraitement'* sont implémentées plusieurs fonctions telles qu'une normalisation géométrique (typiquement par rapport aux positions des yeux et de la bouche), ou conversion des images en niveaux de gris (selon divers critères), ou réduction de l'influence de la variation d'illumination (avec par exemple une égalisation d'histogramme ou un lissage anisotropique [gross-avbpa-2003]). Dans la partie *'extraction caractéristiques et modélisation'*, les images de visages sont représentés par des coefficients des familles de filtres de Gabor (avec des orientations et résolutions différentes). Ces coefficients peuvent être ensuite traités par plusieurs algorithmes tels que la LDA « Linear Discriminant Analysis », ou KFA « Kernel Factor Analysis ». C'est l'algorithme de DLDA « Direct Linear Discriminant Analysis » [yu-pr-2001] qui a été choisi pour les DLL de KIVAOU, car il permet un compromis entre performance de reconnaissance et complexité. L'étape finale de comparaison se fait avec des mesures de distance classiques, combinées à des stratégies de

fusion qui intègrent l'information des trames qui composent une séquence vidéo.

La DLL de reconnaissance de visages FACING-IT

La DLL de reconnaissance faciale fournie par FACING-IT est basée sur un algorithme de modélisation globale de l'apparence des visages couplé à un détecteur des points caractéristiques (53 points) servant notamment à l'alignement et à la normalisation de la taille, de l'orientation et de la position des visages. Par la suite un algorithme d'optimisation par descente de gradient est utilisé en vue de déterminer les paramètres de position, de taille, d'orientation et d'apparence optimaux par rapport à une texture faciale déterminée. Ces paramètres optimaux sont finalement mis en correspondance avec les visages préalablement enregistrés dans la base de données et la similitude est calculée à l'aide d'une distance de Mahalanobis. Le modèle d'apparence codant un visage est enrichi avec de nouvelles informations de texture tant que le visage relatif à une personne donnée reste visible à l'écran. Cette étape confère plus de robustesse à la reconnaissance faciale en présence d'occultations partielles mais elle n'est pas indispensable pour faire une comparaison [abboud-wisg-2009].

Amélioration de la qualité

Il y a plusieurs manières d'améliorer la qualité d'une image : utiliser des prétraitements tels que l'égalisation d'histogramme ou le lissage anisotropique (voir paragraphe précédent), exploiter la vidéo ou faire de la super-résolution.

Dans le contexte de KIVAOU, TELECOM-ParisTech a développé un module de super-résolution. Ce module a pour but d'améliorer les images de basse résolution, à partir d'une séquence vidéo. Ainsi par exemple si on définit une région de 50*50 pixels dans l'image totale, on peut l'agrandir avec un facteur de super-résolution égal à 2, en utilisant 16 images de basse résolution. Avant la reconstruction, il est nécessaire d'effectuer un recalage des images. La méthode de recalage choisie est celle de la correspondance de blocs (« block matching »). L'algorithme implémenté pour la super-résolution effectue une estimation MAP « Maximum A Posteriori » [farsiu-ip-2004]: on cherche l'image haute résolution la plus proche de toutes les images basse résolution (au sens des moindres carrés) et on régularise la solution obtenue en imposant que la somme sur tous les pixels du gradient de l'image soit minimale. L'implémentation s'est faite sous forme de DLL, une seule fonction effectuant à la fois le recalage des images et l'estimation MAP.

Il est nécessaire que le suivi associé à la détection donne le même identifiant pour la même personne tant qu'elle est présente dans le flux. Les stratégies de traitement du flux vidéo sont dépendantes du mode de

fonctionnement. La manière de traiter de manière optimale les séquences vidéo reste encore un verrou à lever.

Démonstrateur

Sagem Sécurité est responsable de l'intégration des différents modules algorithmiques développés par les partenaires du projet. Sagem Sécurité a développé une plateforme logicielle IHM qui cadence le lancement des algorithmes. Ce démonstrateur permet de lancer l'acquisition d'une source vidéo live type caméra (mode valise surveillance), ou l'importation de plusieurs fichiers vidéos (mode plateforme dépouillement). Dans le cadre d'une acquisition live, le flux est enregistré sur disque afin de pouvoir revenir sur des parties de la vidéo. Pour chaque vidéo en entrée, l'utilisateur pourra définir un profil d'analyse (liste des traitements à appliquer). L'analyse peut se faire en mode temps réel (pour du live) ou offline (fichiers importés). L'ensemble des résultats issus des algorithmes sont stockés dans une base de données dédiée. Un module de visualisation permet de présenter à l'utilisateur les résultats des algorithmes de façon générique. Ce module comporte un player vidéo permettant d'afficher et jouer de façon synchrone les différentes vidéos utilisées.

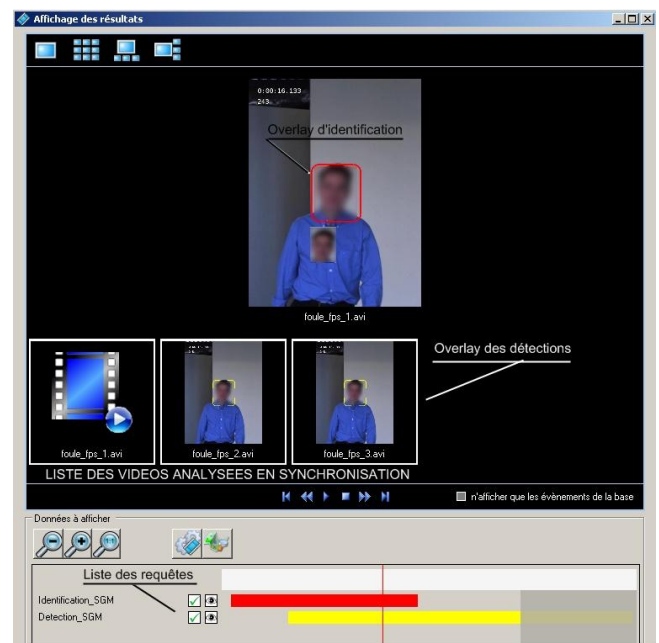


Figure 5.1 : Illustration du démonstrateur.

Les résultats d'un même type d'algorithme (détection de visages par exemple) sont représentés sous la forme d'une « timeline » graphique regroupant les résultats pour toutes les vidéos. L'utilisateur peut se déplacer le long des « timelines » (comme au sein d'un logiciel de montage vidéo) et voir les résultats associés s'afficher simultanément sur l'ensemble des vidéos concernées.

Conclusions et perspectives

L'intégration des DLL fournis par les partenaires est en cours. Des améliorations sont en cours pour permettre soit un traitement le plus proche du temps réel, ou pour choisir des paramètres adaptés pour les démonstrateurs.

La synthèse du « Workshop juridique » relative à l'acceptation juridique des outils proposés est aussi en cours.

Il reste aussi à réaliser la phase de test auprès des utilisateurs finaux, représentés par le Service des Technologies de la Sécurité Intérieure, du Ministère de l'Intérieur, pour valider ou améliorer la pertinence des outils proposés.

Références

- [abboud-wisg-2009] B. Abboud. *Face recognition using appearance models*. Interdisciplinary Workshop on Global Security, WISG, Troyes, France, 2009.
- [bernas-avirs-2008] P. Bernas. *Intérêts et limites de la vidéo-surveillance intelligente pour la Sécurité Globale*. Conférence AVIRS'08, Paris. Avril 2008.
- [beucher-phd-1990] S. Beucher. *Segmentation d'images et morphologie mathématique*. Thèse de Doctorat, Ecole des Mines de Paris, 1990.
- [biancardini-orasis-2005] L. Biancardini, S. Beucher, L. Letellier. *Extraction des objets en mouvement : une approche mixte contours - régions*. ORASIS, 2005.
- [capel-cvpr-1998] D. Capel et A. Zisserman. *Automated mosaicing with super resolution zoom*. Computer vision and pattern recognition, Proceedings. 1998.
- [cootes-tech-report-2004] T. Cootes and C. Taylor. *Statistical Models of Appearance for Computer Vision*. Technical Report, University of Manchester, March 2004.
- [farsiu-ip-2004] S. Farsiu, M. Elad, D. Robinson et P. Milanfar. *Fast and robust super resolution*. IEEE transactions on Image Processing. 2004.
- [gross-avbpa-2003] R. Gross and V. Brajovic. *An Image Preprocessing Algorithm for Illumination Invariant Face Recognition*. 4th International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA). Springer, June 2003.
- [gunturk-ip-2003] B. K. Gunturk, A.U. Batur, Y. Altunbasak, M. H. Hayes, R. M. Mersereau. *Eigenface-domain super-resolution for face recognition*. IEEE Transactions on Image Processing, May 2003, Volume: 12, Issue 5.
- [park-sp-2003] S. Park et al. *Super resolution image reconstruction : A technical overview*. IEEE Signal Processing Magazine. May 2003.
- [schwerdt-avsbs-2005] K. Schwerdt, D. Maman, P. Bernas, E. Paul. *Target segmentation and event detection at video-rate: the EAGLE project*. Proceedings. IEEE Conference on Advanced Video and Signal Based Surveillance. Come, pp. 183- 188, 15-16, Sept. 2005.
- [van-eeken-eurasip-2007] A. Van Eekeren, K. Schutte, O. Oudegeest, L. Van Vliet. *Performance evaluation of super resolution reconstruction methods on real world data*. EURASIP journal on advances in signal processing. 2007.
- [www-biometrie] http://www.biometrie-online.net/dossiers/technique/visage/drawing_blank.pdf
- [www-bka] http://www.bka.de/kriminalwissenschaften/fotofahndung/pdf/fotofahndung_final_report.pdf
- [www-frvt] <http://www.frvt.org>
- [www-mbgc] Multi Biometrics Grand Challenge, <http://face.nist.gov/mbgc/>
- [www-sudfrog] <http://it-sudparis/svn>
- [yu-pr-2001] H. Yu and J. Yang. *A Direct LDA Algorithm for High-Dimensional Data - with Application to Face Recognition*. Pattern Recognition, 34(10):2067-2070. 2001
- [zhou-btas-2009] D. Zhou, D. Petrovska-Delacrétaz and B. Dorizzi. *Automatic Landmark Location with a Combined Active Shape Model*. Proc. IEEE Third International Conference on Biometrics: Theory, Applications and Systems (BTAS 09), Washington DC, September 2009.