
Colloque "Vidéo numérique: Traitement, analyse et reconnaissance"

79ème Congrès de l'Acfas, Sherbrooke, Québec, Canada

PROGRAMME

» 10h00-10h20: Accueil des participants



Café & assortiment de pâtisseries

» 10h20-10h30: Mot de bienvenue

10h20 Caroline Rougier, Université de Montréal, responsable du colloque

» 10h30-11h30: Conférence invitée

Présidente: Caroline Rougier, Université de Montréal

10h30 La vidéosurveillance intelligente : promesses et défis

Valérie Gouaillier, Centre de Recherche Informatique de Montréal (CRIM)

Résumé: Au cours de la dernière décennie, la sécurité des personnes et des biens s'est imposée comme un enjeu majeur à l'échelle planétaire. Confrontées à des problèmes tels que la lutte au terrorisme et le renforcement de la sécurité intérieure, nos sociétés investissent de plus en plus dans des solutions pour assurer leur protection. Parmi les solutions proposées, la vidéosurveillance constitue l'une des technologies de sécurité les plus répandues. Bien qu'encore principalement analogique, elle connaît une révolution numérique avec le passage à la vidéo sur réseaux IP. Intégrant parfois des centaines de caméras, ces nouveaux systèmes génèrent une quantité colossale d'informations vidéo qui dépassent les capacités de surveillance des agents de sécurité. Pour résoudre ce problème, la vidéosurveillance intelligente, par l'analytique vidéo, permet de traiter la vidéo captée pour n'en retenir que les données pertinentes pour la sécurité. Cette présentation exposera les grandes lignes d'un rapport de veille technologique et commerciale rédigé par le Centre de Recherche Informatique de Montréal et Technopôle Défense et Sécurité du Québec. Nous dresserons un portrait de la technologie de la vidéosurveillance intelligente destinée à la sécurité des personnes et des lieux. La présentation sera appuyée par des exemples de réalisations technologiques de divers centres de recherches et compagnies.

» 11h30-12h30: Session 1 - Analyse de visages

Présidente: Caroline Rougier, Université de Montréal

11h30 Analyse et traitement du comportement céphalo-oculaire des conducteurs dans un contexte de conduite simulée et réelle

Samy Metari, Université Laval, Denis Laurendeau, LVSN, Université Laval, Thierry Moszkowicz, LVSN, Université Laval, Normand Teasdale, Université Laval, Martin Lavallière, Université Laval, Martin Simoneau, Université Laval

Résumé: Dans ce travail de recherche nous présentons un nouveau système de vision par ordinateur conçu pour l'analyse et l'interprétation des comportements céphalo-oculaires des conducteurs automobiles. Ce système est muni de trois caméras, une centrale et deux latérales, à partir desquelles l'étude du comportement céphalo-oculaire des conducteurs est réalisée en deux parties. La première partie est dédiée à la détection et au suivi des points caractéristiques faciales, à savoir, les deux yeux et le nez. Le processus de détection fait appel aux classificateurs boostés basés sur les descripteurs rectangulaires de Haar. Le suivi des points caractéristiques est

réalisé par la méthode pyramidale de Lucas-Kanade. Les coordonnées de ces points caractéristiques sont utilisées pour modéliser les actions de recherche visuelle des conducteurs associées à des événements spécifiques tels que la vérification de l'angle mort et la vérification des rétroviseurs gauche et central lors d'un changement de voie. Le système proposé sera utilisé par un groupe de recherche en kinésiologie pour l'évaluation et l'amélioration des performances de conduite de personnes âgées dans un environnement de conduite sécuritaire et contrôlé (simulateur de conduite).

11h50 Reconnaissance d'émotions faciales par descripteurs HoG générés à partir de grilles adaptatives

Mohamed Dahmane, Université de Montréal, Jean Meunier, Université de Montréal

Résumé: L'analyse automatique des expressions faciales est l'aspect le plus souvent étudié en ce qui concerne la compréhension de comportements humains. La principale difficulté pour tout système de reconnaissance d'émotion faciale est de mettre en oeuvre des modèles plus génériques. En effet, une même expression peut varier considérablement d'un humain à un autre, et assez sensiblement pour une même personne mise dans des contextes différents. Dans le but de contrer ces difficultés qui compliquent sévèrement la tâche de reconnaissance, nous avons développé une méthode utilisant des techniques d'apprentissage statistique et des descripteurs d'apparence qu'on compile à travers une grille dense à taille adaptative positionnée sur la partie visage. On se sert des histogrammes de gradients orientés (HoG pour Histograms of oriented gradients) pour extraire les caractéristiques d'apparence à partir des traits de visage. Les histogrammes globaux sont obtenus en compilant les amplitudes de gradients pour un ensemble d'orientations dans des histogrammes 1-D plus locaux définis sur la grille. Comme apprenants de base d'émotions, nous utilisons des machines à vecteurs de support (SVM pour Support Vector Machines).

12h10 Vidéo monitoring automatique des facultés d'une personne par l'analyse du comportement oculaire

François Meunier, UQTR, Claude Desjardins, UQTR

Résumé: Afin d'évaluer les facultés d'une personne, il existe divers tests symptomatiques comme les tests du nystagmus, de convergence oculaire et de réaction à la lumière des pupilles. Ces tests sont effectués manuellement et sont essentiellement appliqués aux yeux. L'objectif de cette recherche est d'améliorer un prototype automatique déjà développé pour automatiser ces tests symptomatiques. Ce nouveau prototype plus compact est aussi constitué d'un système d'analyse d'images vidéo mais avec des fonctionnalités optimisées. Ce dernier prototype comme les premiers, effectue aussi la capture d'images vidéo du mouvement et des réactions de l'œil soumis à des stimuli visuels. Le traitement de ces séquences d'images permet de faire le suivi du déplacement des pupilles ainsi que leurs réactions aux changements d'illumination ambiante et d'ainsi reconstituer les signaux correspondant au déplacement des yeux et des dimensions des pupilles. L'analyse de ces signaux permet d'identifier, entre autres, les saccades correspondant au nystagmus, le degré de convergence des yeux ainsi que les dimensions des pupilles sous divers modes d'illumination qui sont autant de symptômes permettant d'évaluer les facultés d'une personne. Ce prototype est testé sur des volontaires ayant des facultés affaiblies par la consommation de diverses substances ou par la fatigue extrême. L'efficacité du nouvel appareil et des tests automatisés pour déterminer l'état des facultés d'un individu, est ensuite examinée.

» 12h30-13h20: Dîner

 Lunch (offert gracieusement aux auteurs participants)

» 13h20-14h00: Session 2 - Suivi

Président: Pierre-Marc Jodoin, Université de Sherbrooke

13h20 Intégration d'informations spatiales floues dans un filtre particulière pour le suivi mono- et multi-objets dans des séquences d'images 2D

Nicolas Widynski, Université de Montréal

Résumé: Dans cet exposé, nous nous intéressons au suivi d'objets par filtrage particulière dans des séquences d'images 2D. En effet, les méthodes séquentielles de Monte-Carlo, ou filtres particuliers, ont connu un succès grandissant depuis la fin des années 1990 dans la communauté du suivi d'objets, de par leur faculté

d'adaptation à ce type de problématique, leur efficacité en termes de performances ainsi que leur mise en œuvre souvent directe. Dans le cadre d'un thèse réalisée à Télécom ParisTech, nous avons proposé une méthode originale d'intégration d'informations floues dans le filtre particulaire, afin de guider le suivi d'objets en ajoutant des informations spatiales, qui peuvent être de type structurel ou de forme. Cela permet de mieux guider les particules pendant la phase prédictive, et grâce au modèle hybride défini, d'intégrer au filtre particulaire une information synthétisée des k dernières estimations ainsi que la dernière observation. Nous avons également proposé d'intégrer des contraintes spatiales floues dans le filtre particulaire dans le cas multi-objets. Ce type d'information permet de représenter des contraintes complexes entre plusieurs objets. A cette contribution, nous avons ajouté celle d'un nouveau modèle de filtre particulaire pour le suivi multi-objets, qui est une amélioration de l'échantillonnage partitionné.

13h40 Suivi 3D monoculaire appliqué à la détection de chutes

Caroline Rougier, Université de Montréal, Jean Meunier, Université de Montréal, Alain St-Arnaud, CSSS Lucille-Teasdale, Jacqueline Rousseau, Centre de recherche IUGM

Résumé: Suivre la tête d'un individu peut être très intéressant dans les systèmes de vidéosurveillance. De nombreux travaux ont été faits dans ce domaine que ce soit pour le suivi 2D de la tête dans une séquence d'images ou pour le suivi 3D de la tête en utilisant plusieurs caméras. Dans la même optique, nous proposons une méthode originale pour extraire la trajectoire 3D de la tête d'une personne dans une pièce, mais cette fois en se limitant à une seule caméra calibrée. Concrètement, la tête est modélisée par une ellipsoïde 3D et est suivie à l'aide d'un filtre à particules hiérarchique. Les informations utilisées pour le suivi sont basées sur l'histogramme de couleur de la tête et sur la silhouette de la personne. Notre méthode fonctionne quasi en temps réel et les erreurs 3D des trajectoires obtenues sont raisonnables pour un système monoculaire. Nous montrerons un exemple d'application dans le cas de la détection de chutes où une chute est détectée grâce à la vitesse verticale de la tête ou grâce à sa hauteur par rapport au sol, ces informations étant calculées à partir de la trajectoire 3D de la tête.

» 14h00-15h00: Session 3 - Vidéosurveillance

Président: Pierre-Marc Jodoin, Université de Sherbrooke

14h00 Détection automatique de personnes courant dans des images de vidéosurveillance

Marc Lalonde, Centre de recherche informatique de Montréal (CRIM), Samuel Foucher, CRIM, Langis Gagnon, CRIM

Résumé: Notre équipe développe depuis plusieurs années des technologies de vision par ordinateur à des fins de surveillance. Typiquement, des caméras sont reliées à un ordinateur qui analyse les images reçues afin de détecter la présence d'objets d'intérêt (piétons, véhicules). Or avec l'accumulation de données liées à des mouvements d'objets dans un endroit particulier, la machine pourrait-elle apprendre des comportements moyens et ainsi détecter les comportements 'déviant's? Nous avons développé un algorithme qui regroupe spatialement les trajectoires d'objets en déplacement en formant des enveloppes appelées routes et composées de nœuds discrets avec des attributs comme le type d'objet, sa taille et sa vitesse moyenne, etc. On obtient alors un modèle de la scène surveillée qui renseigne sur le comportement usuel des objets à observer. Pour les scènes avec déplacements d'objets moins structurés, les trajectoires accumulées peuvent fournir beaucoup d'informations même sans l'étape d'agrégation spatiale, permettant par exemple le calcul d'un champ de vitesse moyenne. Cette dernière technique a été utilisée pour le concours TRECVID-SED, dont un des objectifs était de concevoir un logiciel capable de détecter, à partir des images de vidéosurveillance, des personnes en train de courir dans un aéroport, donc avec une vitesse de déplacement plus élevée que la moyenne. Nous discuterons des résultats de ces deux approches dans le cadre de TRECVID ainsi que leurs limitations.

14h20 Dénombrement de foule lors d'événements au Quartier des spectacles de Montréal

Samuel Nault, École de technologie supérieure (ÉTS), Thierry Blaise Tasse Fotso, ÉTS, Richard Lepage, ÉTS, Éric Granger, ÉTS

Résumé: Le Quartier des spectacles désire dénombrer le nombre de personnes assistant à ses événements. Il dispose à cette fin de caméras installés sur le toit de 4 édifices en périphérie. Le dénombrement manuel d'une foule étant très long et peu précis, le dénombrement automatique de foule s'impose. Cette étude explore deux méthodes de résolution distinctes, une approche globale et une approche locale. La méthode globale est basée

sur l'analyse des textures des images de foule acquises par caméra. Les images couleurs sont d'abord converties en tons de gris et divisées en plusieurs petites imagerie. Quatre matrices de cooccurrences sont ensuite calculées pour chaque imagerie. Des mesures de contraste, d'énergie, d'homogénéité et d'entropie calculées sur chacune des matrices permettent d'identifier différentes textures. Un classificateur permettra ensuite de déterminer le nombre de personnes se trouvant dans une image selon les textures présentes. L'approche locale est basée sur l'identification des contours de tête de personnes qui sont présentes sur l'image prise par la caméra. Une égalisation d'histogramme permet de minimiser l'influence provoquée par l'illumination; ensuite une transformée de HAAR est appliquée afin d'extraire des coefficients qui seront analysés pour déterminer des vecteurs caractéristiques selon un modèle de foule. Ces vecteurs sont par la suite classés selon la technique SVM pour permettre le comptage du nombre total de foule.

14h40 Recherche de contenu appliquée à la surveillance vidéo

André Caron, Université de Sherbrooke, Pierre-Marc Jodoin, Université de Sherbrooke

Résumé: Cet article porte sur une méthode de recherche par contenu adaptée au domaine de la surveillance vidéo. Notre méthode s'appuie sur une représentation hiérarchique spatio-temporelle de la forme, de la taille et de la direction des objets en mouvement. À l'aide de cette représentation, une table de hachage de type locality-sensitive hashing (LSH) est construite. Légère en mémoire, cette table non seulement résume le contenu utile de la vidéo, mais elle facilite également la navigation à travers de longues séquences (des heures voire des jours de vidéo). La navigation se fait via des requêtes entrées par un utilisateur. Ces requêtes portent sur la forme, la direction et la position des événements recherchés. Une fonction de hachage transforme par la suite la requête en une clé permettant de retrouver en temps réel les sections de la vidéo dont le contenu s'apparente aux critères recherchés. De par la nature d'une table de hachage, la complexité de la recherche est $O(1)$. Notons que la table est trois ordres de magnitude plus légère sur disque que la vidéo d'origine. Cela correspond à environ 55Mo pour une vidéo 320x240 d'une durée de 12 heures dont le contenu brut correspond à plus de 20Go.

» 15h00-15h30: Pause café



Café & assortiment de pâtisseries

» 15h30-16h10: Tutoriel - la Kinect

Président: Jean Meunier, Université de Montréal

15h30 Tout ce que vous avez toujours voulu savoir sur la Kinect

Edouard Auvinet - Université de Montréal, Université de Rennes II, M2S

Résumé: La vive compétition entre les constructeurs de consoles pour jeux vidéo les a poussés à se surpasser. Au niveau de l'interface utilisateur, Microsoft a créé la surprise en proposant la Kinect, une manette humaine où le corps de l'utilisateur est lui-même l'outil de contrôle. Durant la présentation, une description du fonctionnement interne de cette caméra active sera faite. Les domaines d'application à des fins scientifiques seront aussi discutés. Enfin une démo montrant un exemple d'utilisation de la Kinect sera présentée..

» 16h10-16h50: Session 4 - Applications vidéo

Président: Jean Meunier, Université de Montréal

16h10 Rectification métrique de la démarche d'une personne à partir d'une seule caméra calibrée

Frédéric JEAN, Université Laval, Robert BERGEVIN, Université Laval, Alexandra BRANZAN, University of Victoria

Résumé: La démarche humaine contient une foule d'informations sur un individu, autant au niveau de son identité qu'au niveau de sa condition physique. En effet, chaque individu a sa propre façon de marcher et il est donc possible de l'identifier uniquement en observant sa démarche à l'aide de systèmes de vision par ordinateur. Ces systèmes non intrusifs trouvent également des applications dans le domaine médical puisqu'ils simplifient le diagnostic de pathologies reliées à la locomotion humaine. Ces systèmes se limitent toutefois aux vues de profil des marcheurs. Dans le cas de points de vue différents, les mesures effectuées dans les images provenant

de la caméra sont faussées par les effets de projection de perspective. Une façon de pallier à cette limitation consiste à décomposer la marche en ses demi-cycles et à déterminer à chaque demi-cycle l'image du plan dans lequel est effectué le mouvement de chaque partie du corps (tête, pieds). Connaissant les paramètres intrinsèques ainsi que la ligne d'horizon de la surface de marche, il est possible de connaître l'orientation de ces plans par rapport à la caméra, et donc de rectifier les mouvements effectués dans une vue de profil métrique. Les résultats obtenus avec cette méthode sont prometteurs puisqu'il est possible de recouvrer des mesures telles que la longueur des foulées et la grandeur des individus observés. Les travaux futurs porteront sur la modélisation et la comparaison de démarches à partir des mouvements rectifiés.

16h30 Analyse des textures en temps réel exploitant une architecture parallèle multi-cœurs et GPU

Moulay AKHLOUFI, Université Laval, Gilles CHAMPAGNE, Centre de Robotique et de Vision Industrielles, Mario LAFRAMBOISE, Centre de Robotique et de Vision Industrielles

Résumé: Ces dernières années on assiste à un intérêt croissant pour l'analyse des textures dans des applications industrielles. Dans ce type d'application, le traitement en temps réel est primordial, surtout avec la croissance de la résolution des caméras et des images traitées. Plusieurs solutions pour optimiser le temps de traitement sont disponibles : DSP, FPGA, cartes dédiées, systèmes parallèles, clusters, etc. La majorité de ces solutions sont très coûteuses et peu évolutives. Plus récemment, la communauté scientifique s'est intéressée à l'utilisation des unités de traitement graphique disponibles commercialement pour accélérer les traitements des données. Ceci a donné naissance au domaine du GPGPU (General-Purpose computation on Graphics Processing Units) qui vise à utiliser la puissance des processeurs graphiques pour accélérer divers types de traitements de données : mathématique, visualisation 3D, traitement des images, etc. Dans ce travail, nous présentons l'utilisation du GPGPU dans l'optimisation et le traitement en temps réel de certains descripteurs utilisés pour l'analyse des textures. Les techniques suivantes ont été optimisées : LBP (Local Binary Pattern), LTP (Local Ternary Pattern), LATP (Local Adaptive Ternary Pattern), Noyaux de Laws, filtres de Gabor et GLCM (Gray Level Co-Occurrence Matrix). L'architecture choisie pour ce travail repose sur la technologie CUDA (Compute Unified Device Architecture) proposée par NVIDIA. Cette architecture permet d'augmenter de manière significative la performance des traitements en exploitant la puissance des unités de traitement graphique (GPU) et une architecture de programmation parallèle. Les optimisations GPU sont comparées aux optimisations CPU utilisant les technologies MMX-SSE et la programmation parallèle sur des processeurs multi-cœurs. De plus, l'architecture CUDA de nouvelle génération « Fermi » sera utilisée dans nos tests. Les résultats expérimentaux montrent des gains importants lorsque l'architecture GPU est utilisée, plus particulièrement pour des images de grande taille. Les figures incluses dans le fichier ci-joint, montrent quelques exemples de gains importants obtenus pour des descripteurs de textures avec l'optimisation sur des architectures parallèles CPU multi-cœurs et GPU.

» 16h50-17h00: Mot de clôture

16h50 Jean Meunier, Université de Montréal, responsable du colloque