ETS-RT - 2011-002

MÉTHODES D'ÉVALUATION DE L'ACHALANDAGE DU QUARTIER DES SPECTACLES
Étape 1 : Revue de documentation

RICHARD LEPAGE
ÉRIC GRANGER

ETS-RT – 2011-002

MÉTHODES D'ÉVALUATION DE L'ACHALANDAGE DU QUARTIER DES SPECTACLES
Étape 1 : Revue de documentation

RAPPORT TECHNIQUE DE L'ÉTS

RICHARD LEPAGE
ÉRIC GRANGER

Génie de la production automatisée

ÉCOLE DE TECHNOLOGIE SUPÉRIEURE
UNIVERSITÉ DU QUÉBEC

MONTRÉAL, LE 17 JANVIER 2011

ETS-RT – 2011-002

MÉTHODES D'ÉVALUATION DE L'ACHALANDAGE DU QUARTIER DES SPECTACLES

Étape 1 : Revue de documentation

RICHARD LEPAGE
ÉRIC GRANGER
Génie de la production automatisée
ÉCOLE DE TECHNOLOGIE SUPÉRIEURE

La version électronique de ce rapport technique est disponible sur le site Web de l'École de technologie supérieure (http://www.etsmtl.ca).

Pour se procurer une copie papier, s'adresser à :

Service de la bibliothèque
École de technologie supérieure
1100, rue Notre-Dame Ouest
Montréal (Québec)
H3C 1K3

Téléphone : (514) 396-8946
Télécopieur : (514) 396-8633
Courriel : biblio@etsmtl.ca

© École de technologie supérieure 2011

La citation d'extraits ou la reproduction de courtes sections est permise à la condition de mentionner le nom de l'auteur et la source. Toute reproduction de parties quantitativement ou qualitativement importantes requiert l'autorisation du titulaire des droits d'auteur.

ISBN 978-2-921145-77-0

Dépôt légal : Bibliothèque nationale du Québec, 2011
Dépôt légal : Bibliothèque nationale du Canada, 2011

MÉTHODES D'ÉVALUATION DE L'ACHALANDAGE DU QUARTIER DES SPECTACLES
Étape 1 : Revue de documentation

RICHARD LEPAGE
ÉRIC GRANGER

RÉSUMÉ

Le Partenariat du Quartier de spectacles souhaite disposer d'outils d'évaluation de l'expérience du visiteur au moyen de sondages qualitatifs et d'outils d'évaluation de l'achalandage. Ces derniers outils serviront à mesurer l'évolution de l'achalandage aux différents pôles du quartier et à permettre ainsi d'utiliser cet achalandage par exemple à des fins d'affichage publicitaire.

Le Partenariat a confié à l'ÉTS un mandat pour, dans un premier temps, procéder à une revue complète de documentation pour identifier les pistes de solution suivies jusqu'à présent par les chercheurs et éventuellement identifier un produit commercial déjà existant qui nécessiterait peu ou aucune modification pour être adaptée aux conditions d'opération du Partenariat.

La revue de documentation a porté sur l'étude de la littérature scientifique de même que sur l'étude des brevets et logiciels construits pour dénombrer des foules. Notre étude montre qu'il n'y a actuellement aucun logiciel ou brevet qui réponde aux spécifications et contraintes posées par le Partenariat. Nous nous sommes dès lors concentrés sur l'étude de la littérature scientifique pour bien dégager les diverses approches développés par des scientifiques et identifier les approches les plus prometteuses pour solutionner notre problématique.

Notre étude a permis d'identifier 2 approches qui nous semblent particulièrement intéressantes. Une première approche, plus globale, utilise la notion de texture afin d'évaluer la densité d'une foule. La texture permet de quantifier la rugosité de surface d'un motif sur l'image et la mesure de divers paramètres de texture permet d'associer

une densité de foule d'une zone dans l'image aux divers paramètres mesurés de texture de cette même zone. La seconde approche, plus locale, consiste à identifier le plus finement possible les têtes qui apparaissent dans la zone étudiée pour ensuite estimer le nombre de personnes présentes dans la zone. Ces deux scénarios seront étudiés et expérimentés durant le trimestre d'hiver 2011 par deux étudiants en fin de B.Ing. qui sont inscrits en PFE (projet de fin d'études).

TABLE DES MATIÈRES

Résumé ..... **iv**

Table des matières ..... **vi**

Liste des figures ..... **vii**

Liste des tableaux ..... **viii**

1 Introduction **1**
- **1.1 Résumé du mandat** **1**
- **1.2 Contexte de l'étude** **2**
  - 1.2.1 Acquisition des images de foule 2
  - 1.2.2 Conditions d'opération 3
  - 1.2.3 Images de test 3

2 Techniques pour le comptage de Foule **6**
- **2.1 Estimation manuelle** **6**
- **2.2 Analyse automatique** **9**
  - 2.2.1 Prétraitement 10
  - 2.2.2 Extraction de descripteurs 16
  - 2.2.3 Classification et comptage 23
- **2.3 Discussion** **25**

3 Approches commerciales **27**
- **3.1 Logiciels** **27**
  - 3.1.1 *ExactCrowd* de Lynce 27
  - 3.1.2 *Automated Footfall Monitoring* de Springboard 29
- **3.2 Brevets** **30**

4 Approches proposées **33**
- **4.1 Méthodes prometteuses** **33**
- **4.2 Évaluation expérimentale** **36**

5 Conclusion **37**

Bibliographie **38**

LISTE DES FIGURES

Figure 1 : Photos prises du toit de l'UQAM, 2 de jour et 1 en soirée. 4

Figure 2 : Photos prises de jour, du toit de la Maison des Festivals. 4

Figure 3 : Images instantanées prises à partir de séquences vidéo haute définition. 5

Figure 4 : Tourniquets de comptage. 7

Figure 5 : Estimation manuelle de la densité de foule. 8

Figure 6 : Système général pour le comptage de foule. 10

Figure 7 : Soustraction d'une image de référence. 11

Figure 8 : Redressement d'une image. 12

Figure 9 : Pavage de l'image avec des grilles de dimension variable. 13

Figure 10 : Points singuliers détectés dans deux images contenant de la foule. 15

Figure 11 : Équipement utilisé par Lynce pour le comptage de foule : (1) caméra vidéo haute définition; (2) ballon aérien; (3) appareil photo numérique haute définition. 29

Figure 12 : Structure du logiciel de comptage de foule. 33

LISTE DES TABLEAUX

Tableau 1 : Sélection des algorithmes pour chaque étape du système de comptage de foule.....34

1 INTRODUCTION

Ce premier rapport technique comprend une synthèse des techniques pertinentes qui sont proposées dans la littérature scientifique, les brevets et les technologies commerciales pour le comptage de foules dans les espaces publiques. Il contient aussi des recommandations et une analyse des différentes pistes prometteuses pour cette application. Nous abordons dans cette section les principaux termes de notre mandat, la portée de notre étude de même que le contexte de fonctionnement de notre étude.

1.1 Résumé du mandat

Le Partenariat du Quartier de spectacles (PQS) a notamment pour mandat de gérer la programmation et l'animation culturelles sur différents espaces publics du centre-ville de Montréal. Le Partenariat souhaite disposer d'outils quantitatifs pour l'évaluation de l'achalandage. Les données quantitatives serviront à mesurer l'évolution de l'achalandage aux différents pôles du quartier, et à « vendre » cet achalandage, par exemple à des fins d'affichage publicitaire.

Le PQS n'a pas identifié une technologie commerciale répondant à ses besoins qui permettrait, à partir des images numériques à haute résolution, de dénombrer le plus précisément possible (avec une marge d'erreur constante) les foules présentes. Dans un premier temps, le PQS a confié à l'ÉTS un mandat d'expertise abordant la première étape de ce projet, soit une revue de la documentation (littérature scientifique, brevets, produits commerciaux, etc.) afin d'évaluer :

- s'il existe un produit commercial répondant aux besoins de PQS;
- s'il existe un produit susceptible de répondre aux besoins de PQS moyennant une adaptation simple;
- l'inventaire des pistes suivies jusqu'à présent par les chercheurs.

Au début du mandant, nous avons analysé des images statiques captées par Mattera avec une caméra numérique Nikon D700 de 12,1 Mpixel et les trames d'une vidéo HD captée sur le toit de l'UQAM. Nous jugeons qu'il est possible à court terme d'utiliser les plateformes de développement à notre disposition (Aphelion, OpenCV, MATLAB, etc.) afin de valider la fonctionnalité et les performances d'approches prometteuses. Cette vérification ne produirait pas un logiciel performant et robuste pour le comptage de foules, ni un logiciel pour commander les caméras. Nous allons entamer cette vérification sous peu avec la réalisation d'algorithmes de détection de foule car elle constitue la partie délicate du projet, qui nous permettra de confirmer la viabilité du projet de comptage de foule.

1.2 Contexte de l'étude

Notre mandat consiste à évaluer diverses méthodes de comptage de foule dans le contexte prévu d'opération du Quartier des spectacles. Les contraintes spécifiques au Quartier des spectacles sont détaillées ci-dessous.

1.2.1 Acquisition des images de foule

Lorsque le Quartier des spectacles sera pleinement opérationnel, 4 caméras vidéo à haute définition seront disposées en périphérie du Quartier des spectacles afin de couvrir au mieux l'ensemble des diverses scènes extérieures d'un évènement comme le Festival international de Jazz de Montréal ou les Francofolies. Ces caméras seront disposées sur le toit de 4 immeubles et chaque caméra pourra être commandée pour ajuster le zoom et l'angle d'incidence par rapport à la verticale. Le comptage de foule pour chaque caméra sera réalisé en différé par l'analyse informatique d'un instantané provenant de la séquence vidéo d'une des caméras. S'il s'avère que la résolution de l'image instantanée n'est pas suffisante pour l'analyse automatique, il faudra alors envisager l'achat et l'installation couplée à chaque caméra vidéo de 4 appareils photos numériques à haute résolution.

1.2.2 Conditions d'opération

L'analyse automatique des images de foule tentera d'estimer le nombre de piétons dans diverses zones d'une image à haute résolution. Le logiciel de comptage de foule devra être en mesure d'opérer pour les diverses conditions suivantes :

1. La densité de la foule est très variable, pouvant passer d'espaces presque vides à des zones fortement peuplées lors de spectacles extérieurs.
2. Les images pourront être acquises de jour comme de soir, dans des conditions météorologiques et d'éclairage artificiel variables.
3. La précision des images devra être suffisante afin que la zone couverte par chaque caméra soit suffisamment étendue pour que les 4 caméras couvrent l'ensemble des scènes extérieures de l'évènement.
4. La technique pour le comptage de foule devra composer avec les contraintes d'un taux d'erreur maximum fixe. Lors de son analyse, cette technique n'exploitera pas simultanément les caméras multiples pour analyser une même scène.
5. La foule devant un spectacle étant relativement statique, le mouvement des individus dans une foule, pouvant être capté sur de trames vidéo successives, ne sera pas exploité.

1.2.3 Images de test

Afin d'étudier la faisabilité d'un logiciel robuste de comptage de foule, nous avons obtenu des images numériques du studio Mattera et une séquence vidéo filmée par le réalisateur Charles-Olivier Bleau.

Les images numériques fournies par le Studio Mattera ont été acquises le 5 juillet 2010, de jour et de soir à partir du toit de l'UQAM et de jour à partir du toit de la Maison des Festivals. Cette date correspond à l'avant-dernière journée du Festival International de

Jazz de Montréal. L'appareil photo utilisé est un appareil professionnel à visée reflex de marque Nikon D700 avec une résolution de 12,1 Mpixel. Les figures ci-dessous illustrent des photos typiques captées par le photographe.

Figure 1 : Photos prises du toit de l'UQAM, 2 de jour et 1 en soirée.

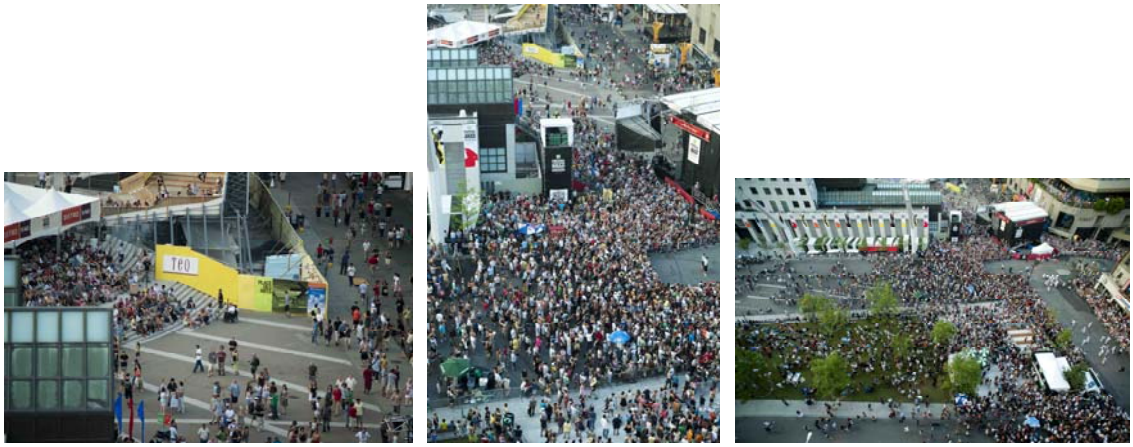

Figure 2 : Photos prises de jour, du toit de la Maison des Festivals.

Quelques séquences vidéo haute définition ont été enregistrées à partir du toit de l'UQAM le dimanche 18 juillet 2010 de 20h30 à 21h20, lors du festival Juste pour rire. La vidéo a été captée par le réalisateur Charles-Olivier Bleau à l'aide d'une caméra

professionnelle louée pour l'occasion. La figure ci-dessous illustre 3 instantanés capturés à différents instants de la vidéo.

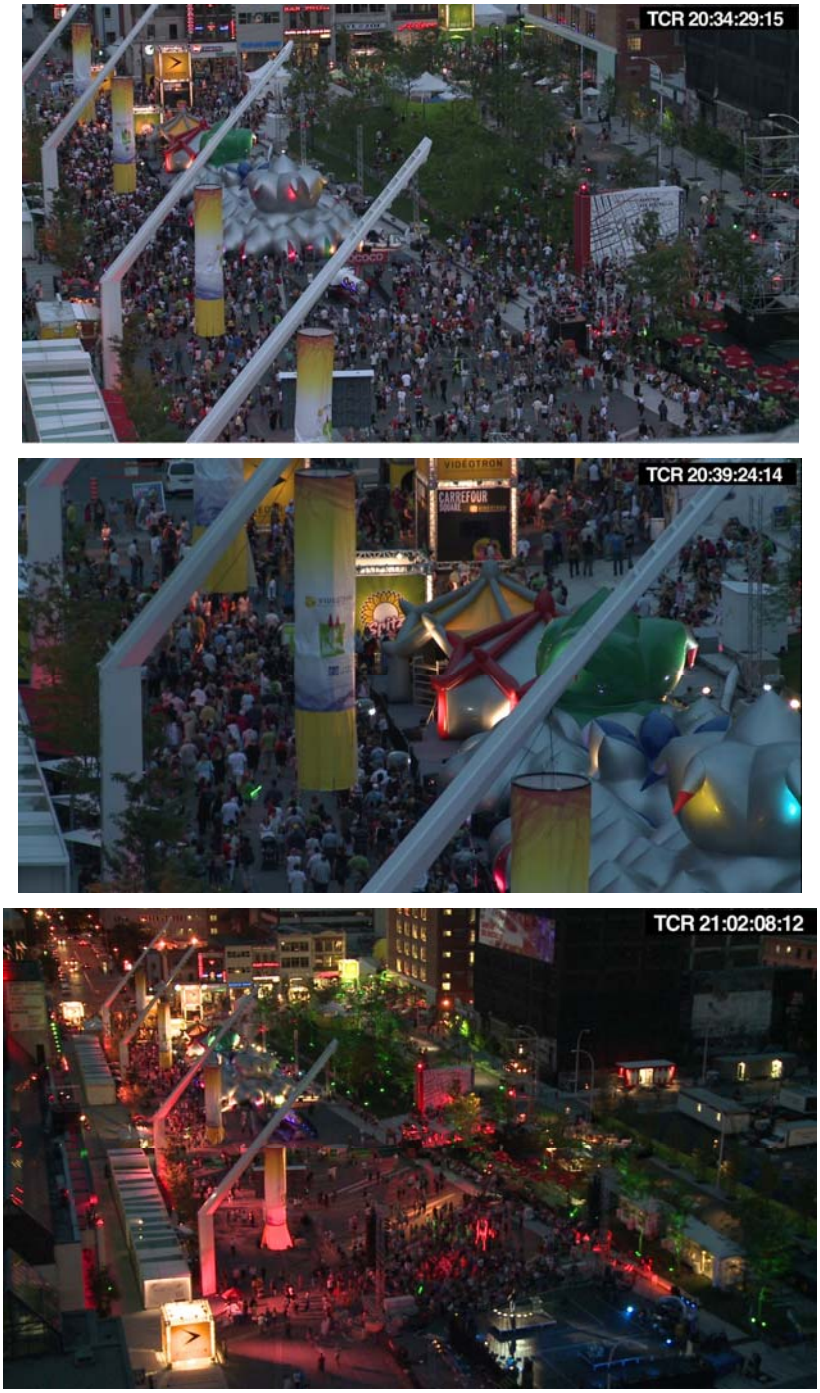

Figure 3 : Images instantanées prises à partir de séquences vidéo haute définition.

2 TECHNIQUES POUR LE COMPTAGE DE FOULE

L'estimation du nombre de personnes assistant à un spectacle en plein air, à une manifestation ou à une inauguration a, depuis toujours, intéressé les organisateurs, les journalistes et/ou les autorités policières. Cette estimation pourra être utilisée par les organisateurs pour promouvoir un évènement, par les journalistes afin de rapporter l'évènement le plus fidèlement possible et par les autorités policières pour organiser les mesures de sécurité appropriées. Jusqu'à présent, la grande majorité des estimés de densité de foule est calculée manuellement, offrant de très grandes différences dépendant qui réalise l'estimation. Par exemple, lors d'une manifestation contre la réforme de la retraite à Lyon le 13 octobre 2010, la Police estimait la foule à quelque 19 000, alors que les Syndicats organisateurs de l'évènement l'estimaient à 45 000! Ces différences significatives peuvent être amoindries grâce à des méthodes, toujours manuelles, mais plus rigoureuses et qui nécessitent des moyens et de l'équipement plus sophistiqués et plus onéreux. Une autre piste de solution consiste à automatiser l'opération de comptage de la foule grâce aux percées récentes de la vision par ordinateur (Zhan et al., 2008).

2.1 Estimation manuelle

Lorsque leur installation est possible, les tourniquets d'accès à un site demeurent les outils de comptage les plus simples et les plus précis pour déterminer le nombre de personnes présentes sur un site (Figure 4). Il n'est malheureusement pas possible d'utiliser cette technique pour les sites du Quartier des spectacles parce qu'il existe de multiples points d'entrée et que la présence de tourniquets ralentirait considérablement le flux des spectateurs.

Figure 4 : Tourniquets de comptage

Des méthodes manuelles d'estimation de foule plus rigoureuses reposent sur les travaux du professeur de journalisme Herbert Jacob, développés à la fin des années 60 et encore largement utilisées de nos jours. Jacob, à la fin des années 60, à l'occasion d'une manifestation monstre sur le campus de l'Université de Californie à Berkeley contre la guerre au Vietnam, observait la manifestation à partir de son bureau situé dans une tour. Son bureau surplombait une plazza marquée de lignes formant des grilles régulières bien identifiées. Il eut alors l'idée d'estimer le nombre de personnes en évaluant approximativement la densité de la foule à l'intérieur d'un rectangle de superficie connue (les grilles de la plazza) (Doig, 2009, 15 january). Jacob utilisait trois niveaux de densité :

- Densité faible : 1 personne / m^2. Les participants sont à un bras tendu de distance les uns des autres.

- Densité moyenne : 2,5 personnes / m^2. Chaque personne occupe 0,4 m^2.

- Densité forte : 4,5 personnes / m^2. Chaque personne n'occupe plus que 0,2 m^2.

Le nombre de personnes dans une grille est calculé par le produit de la densité estimée par la surface de la grille. Le nombre total de participants est obtenu en additionnant les réponses obtenues pour chaque grille. (Melina, 2010, 3 September) propose une extension de la méthode de Jacob en utilisant des photos aériennes (Figure 5).

Figure 5 : Estimation manuelle de la densité de foule.

La photographie est d'abord quadrillée par des carrés ou des rectangles de dimensions connues. La densité de chaque grille est évaluée puis le nombre approximatif de participants dans chaque grille est calculé.

2.2 Analyse automatique

La Figure 6 présente le diagramme bloc du système générique qui vise à une estimation automatique du nombre d'individus dans une foule. Ce diagramme bloc s'apparente à un système de reconnaissance de formes classique (Duda, Hart et Stork, 2001). Étant donnée une image numérique **I** qui est captée par une caméra vidéo à haute résolution, ce système parcourt quatre étapes principales de traitement afin d'estimer le nombre d'individus qui occupent l'image.

Dans un premier temps, le système applique un *prétraitement* à l'image **I** afin d'isoler une ou plusieurs régions d'intérêt et/ou de conditionner l'image, pour le rendre e.g. invariante aux variations d'illumination et de perspective. Ensuite, lors de l'*extraction des descripteurs*, des caractéristiques pertinentes et discriminantes sont extraites des régions d'intérêt ou de l'image complète. Les caractéristiques sont souvent regroupées en un vecteur $\mathbf{v}_i$ qui est de nature locale ou globale.

La troisième étape consiste à *classifier* le vecteur de caractéristiques pertinentes selon une des classes de reconnaissance. La classification doit généralement faire appel à un modèle de foule qui correspond aux vecteurs caractéristiques $\mathbf{v}_i$. Les classes peuvent correspondre à différents niveaux de densité de la foule dans une des régions d'intérêt de **I**, ou bien à l'occurrence d'un mouvement ou d'un contour de tête. Finalement, la quatrième étape vise à estimer le nombre total d'individus dans l'image, en tenant compte de différentes informations contextuelles.

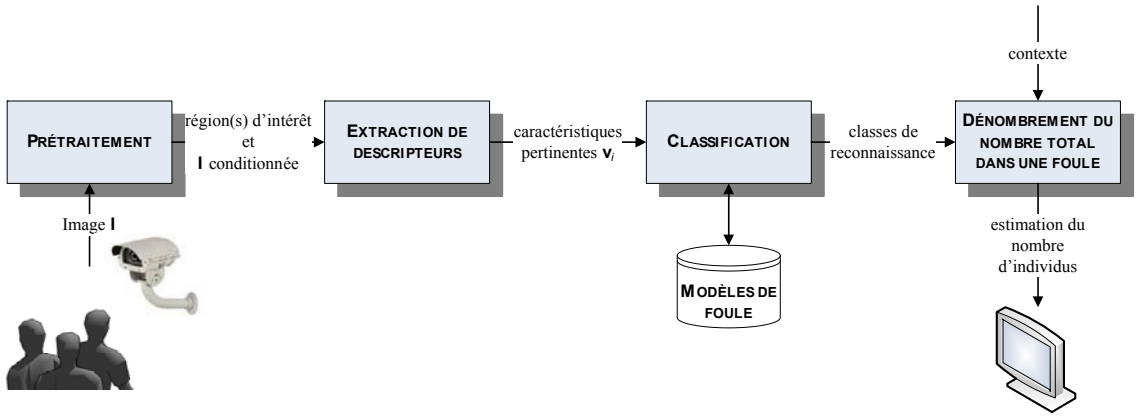

Figure 6 : Système général pour le comptage de foule.

Dans la littérature scientifique, un certain nombre d'auteurs ont proposé des systèmes automatiques pour estimer le comptage de foules. La suite de cette section présente une synthèse de ces techniques en fonction des différentes étapes du traitement.

2.2.1 Prétraitement

L'étape du prétraitement comprend habituellement un traitement de l'image **I** pour faire ressortir les régions ou surfaces d'intérêt qui sont occupées par des individus. Le prétraitement permet également de corriger certains défauts de la capture d'image comme la distorsion de perspective et l'illumination non-uniforme de la scène.

Soustraction de l'arrière plan sans foule (Davies, Yin et Velastin, 1995) proposent de capter une image de la scène sans la présence de la foule. Cette image devient l'image de référence **R**. L'image de référence **R** est par la suite soustraite de l'image avec foule **I**. L'image obtenue ne contient plus que les individus qui composent la foule à évaluer (Figure 7). Pour qu'une telle méthode fonctionne bien, il faut disposer de période dans la journée où il y a très peu ou pas du tout de participants et que les paramètres de la prise d'images (élévation et orientation de la caméra, zoom, zone imagée) demeurent les mêmes.

(Denman, Chandran et Sridharan, 2007) proposent l'analyse du mouvement de la foule pour identifier les zones peuplées. Cette technique ne peut être utilisée dans notre cas puisque les densités de foule estimées sont principalement statiques.

Figure 7 : Soustraction d'une image de référence.

Une image de référence **R** est d'abord obtenue sans la présence de la foule à évaluer. Cette image **R** est ensuite soustraite de l'image de foule à évaluer. Les pixels de la zone d'observation appartiennent tous à des individus.

Effet de perspective L'effet de perspective se manifeste parce que la prise d'images n'est pas à la verticale. La caméra est située à une hauteur h et est orientée à un angle θ par rapport à l'horizontale. Cette disposition fait en sorte que les personnes plus près de la caméra occuperont plus d'espace sur l'image que les personnes plus éloignées. (Ryan et al., 2009) corrigent cet effet de perspective en calculant un facteur de correction basé sur l'espace occupé par une personne à l'avant-plan et l'espace occupé par une personne à l'arrière-plan. (Davies, Yin et Velastin, 1995) proposent de redresser l'image afin d'obtenir l'image résultante comme si la caméra avait été placée à la verticale de la scène observée. Le résultat de la correction géométrique appliquée pour redresser

l'image apparaît ci-dessous. Les auteurs ne précisent pas l'algorithme utilisé, mentionnant que de tels algorithmes se retrouvent dans le monde de la publicité ou de la télévision où de tels redressements sont réalisés de façon routinière pour obtenir de bons effets vidéos. Nous croyons pouvoir adapter des algorithmes de redressement largement utilisés en télédétection afin de pouvoir recaler les images provenant d'un satellite d'observation de la Terre et les cartes géographiques géoréférencés.

Figure 8 : Redressement d'une image.

L'image de gauche est redressée sur l'image de droite. La caméra semble maintenant située à la verticale.

(Kong, Gray et Tao, 2005) tiennent compte de l'orientation de la caméra (ils supposent que la caméra est suffisamment élevée) pour calculer l'homographie entre le plan image et la scène 3D correspondante afin de rendre le système de comptage de foule invariant à l'effet de perspective.

(Wu et al., 2009) superposent sur l'image un ensemble de grilles d'évaluation, chaque grille étant formé d'un rectangle dont les dimensions diminuent au fur et à mesure que

l'on s'éloigne de la caméra pour tenir compte de la projection de perspective. Ainsi, pour une densité fixe de personnes, chaque cellule contiendra approximativement le même nombre de personnes, les petites grilles situées loin de la caméra entourant des personnes qui apparaissent elles aussi plus petites à cause de l'effet de perspective. La Figure 9 ci-dessous illustre le pavage de l'image par un grillage dont les dimensions des fenêtres diminuent en s'éloignant de la caméra.

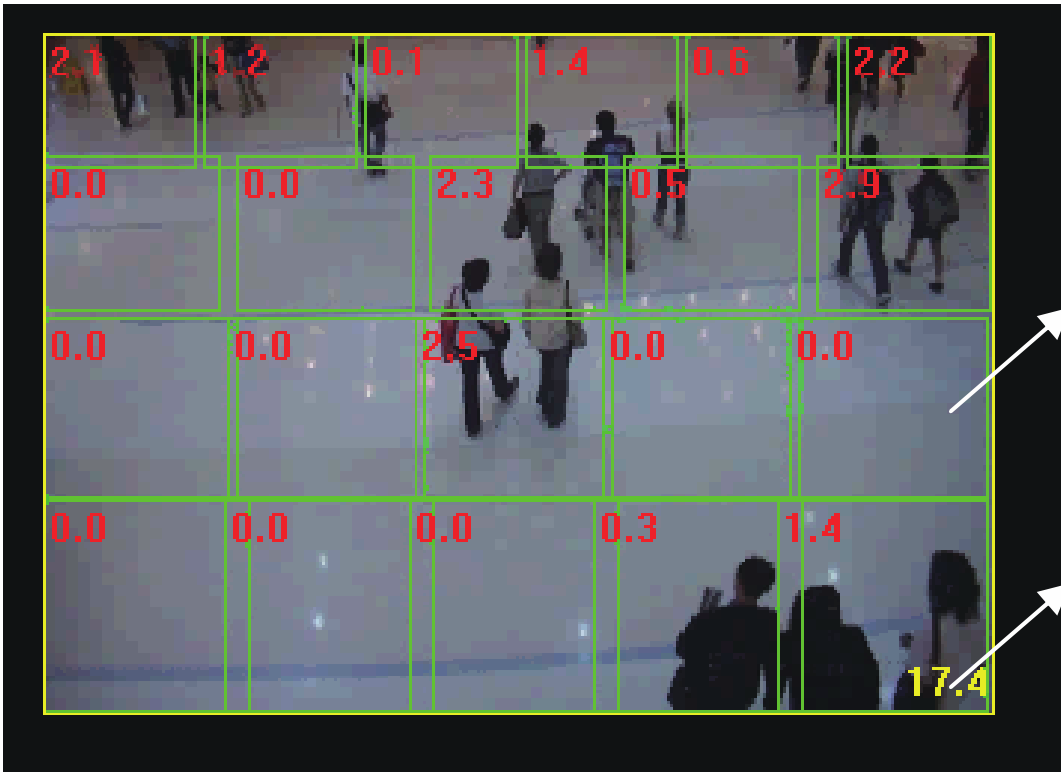

Figure 9 : Pavage de l'image avec des grilles de dimension variable.

Les dimensions des grilles diminuent en s'éloignant de la caméra pour compenser les effets de la projection de perspective (d'après (Wu et al., 2009)). Le chiffre en jaune indique le nombre total de personnes sur l'image alors que les chiffres en rouge indiquent le nombre de personnes dans chaque grille.

Éclairage La scène à l'étude offre une apparence différente selon qu'on est en début de journée, sur l'heure du midi ou en début de soirée. (Lin, Chen et Chao, 2001) proposent d'égaliser l'histogramme des niveaux de gris de l'image pour minimiser l'influence de l'éclairage. L'histogramme d'une image est un graphique représentant la distribution des valeurs de niveau de gris de tous les pixels de l'image. L'égalisation de l'histogramme permet une distribution uniforme des niveaux de gris, permettant d'atténuer l'éclairement des zones sur-éclairées de l'image. L'égalisation d'histogramme permet une certaine invariance à l'illumination, ou éclairage, de la scène.

Invariance à l'échelle (Arandjelovic, 2008) propose un système de segmentation d'une image en zones de foule et en zones où la foule est absente. Sur une image, la foule se présente sous 2 différentes formes dépendamment de l'échelle de la photo. À une petite échelle (résolution élevée), les pixels de l'image indiquant la présence de foule sont composés d'éléments périodiques de base, chaque élément répétitif ressemblant faiblement à des personnes (faiblement parce que les vêtements peuvent être de couleur différente, l'attitude varie, présence d'occlusion, etc.). À une plus grande échelle, la foule sur l'image apparait comme un ensemble d'éléments répétitifs à peu près uniformes en apparence. Arandjelovic propose de détecter les points singuliers invariants sur l'image, et ce à plusieurs échelles au moyen d'une fenêtre d'analyse glissante constituée d'une pyramide multirésolution. Les points singuliers correspondent à certains points caractéristiques d'objets sur l'image qui seront détectés quelque soit l'échelle, l'orientation ou l'illumination. Pour un objet en particulier, il n'y aura que quelques points singuliers qui répondront aux critères. Par exemple, pour un cube, il n'y aura que les coins du cube qui seront détectés comme points singuliers. La détection des points singuliers invariants est basée sur l'algorithme SIFT (*Scale Invariant Feature Transform*) (Lowe, 2004). Chaque point singulier détecté est représenté par un vecteur de grande dimension (160 éléments) qui permet de décrire les caractéristiques du point singulier et son échelle caractéristique. La Figure 10 ci-dessous illustre l'extraction des

points singuliers d'une image de foule. Chaque point singulier est représenté par un cercle dont le diamètre est lié à l'échelle caractéristique du point.

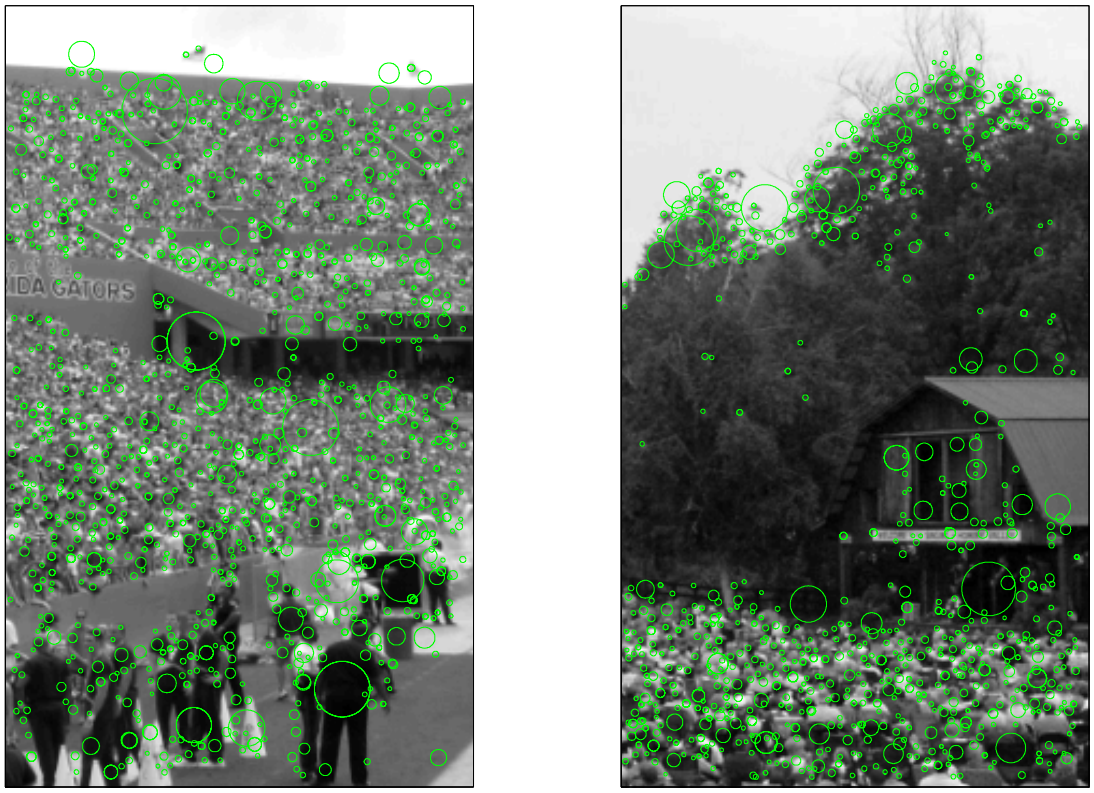

Figure 10 : Points singuliers détectés dans deux images contenant de la foule.
Les points singuliers sont représentés par des cercles. Le diamètre d'un cercle représente l'échelle caractéristique du point singulier. (d'après (Arandjelovic, 2008))

Certains points singuliers détectés sur les personnes composant la foule possèdent des vecteurs descriptifs qui sont voisins dans l'espace vectoriel de 160 dimensions et qui sont discriminants par rapport aux autres points singuliers. Cette proximité des vecteurs descriptifs permet d'identifier les zones dans l'image qui contiennent des portions de la foule, et ce quelque soit le zoom ou la distance de la caméra qui capte la scène.

2.2.2 Extraction de descripteurs

Il existe dans la littérature plusieurs techniques permettant d'extraire des caractéristiques pertinentes et discriminantes à partir d'une zone de l'image **I**. Parmi les descripteurs d'une foule, on retrouve la densité des pixels, la texture, la forme (apparence de la tête) et le mouvement. Ces descripteurs peuvent être extraits selon différentes techniques proposées en traitement d'image et en reconnaissance de formes. Par contre, des approches spécialisées sont requises pour l'extraction de caractéristiques dans un contexte d'analyse des scènes denses.

La majorité des techniques décrites dans la littérature propose d'extraire des caractéristiques de nature holistique ou globale (basées sur l'ensemble de la foule) afin d'estimer la densité d'un foule en termes du nombre de personnes par unité de surface (Chan, Liang et Vasconcelos, 2008; Davies, Yin et Velastin, 1995). Les techniques basées sur la densité des pixels, la texture, la morphologie et les points invariants exploitent des caractéristiques dites globales. Par contre, certaines autres approches proposent d'extraire des caractéristiques locales, basées sur une seule personne ou sur de petits groupes d'individus, pour ensuite compter plus directement le nombre d'individus. Les techniques basées sur l'analyse de mouvement et l'apparence (i.e. le contour d'une tète) exploitent des caractéristiques dites locales (Ryan et al., 2010).

2.2.2.1 Densité des pixels classés « foule »

Cette catégorie d'algorithmes repose sur une estimation directe de la densité d'individus dans les régions d'intérêt à partir du nombre de pixels classés « foule ». Cette densité est ensuite reliée à un nombre équivalent de personnes par une fonction linéaire (Davies, Yin et Velastin, 1995).

La segmentation de l'image I pour identifier les pixels étiquetés « foule » s'obtient en lui soustrayant l'image de référence R captée en l'absence de foule. La densité de foule est

ensuite calculée soit en comptant le nombre de pixels classés « foule » par rapport au nombre total de pixels de la zone d'intérêt, soit en extrayant le contour aminci des individus et en comptant le nombre de pixels appartenant aux contours d'individus par rapport au nombre total de pixels de la zone d'intérêt.

L'hypothèse fondamentale est que le nombre de pixels étiquetés « foule » ou « contour d'individu » liés aux régions d'intérêt est linéairement proportionnel au nombre de personnes dans la foule. L'hypothèse n'est pas vérifiée pour des foules très denses, où on retrouve beaucoup d'occlusion et de chevauchement entre les individus (Zhan et al., 2008). D'après (Marana et al., 1997), la relation devient non linéaire lorsque le nombre de personnes dans la zone d'intérêt dépasse environ 30 personnes, ce qui, pour nous, est une condition habituelle lors des spectacles extérieurs. De plus, cette approche est sensible aux variations d'illumination et aux nouvelles formes dans la scène qui ne sont pas des personnes. L'hypothèse est également mise à mal lorsque la région d'intérêt est soumise à des effets de perspective non-négligeables : les personnes près de la caméra occupent un espace plus grand sur l'image que les personnes en arrière-plan.

(Hou et Pang, 2008) situent la caméra loin de la zone étudiée, ce qui élimine en bonne partie les effets de perspective. Ils introduisent la notion intéressante d'image de référence adaptative, ce qui permettrait de modifier l'image de référence sans foule au fur et à mesure de la progression du spectacle. Les auteurs se servent également du mouvement dans la foule pour parfaire l'identification des pixels étiquetés « foule ».

2.2.2.2 Texture

L'identification des différents éléments composant une image numérique n'est souvent pas aisée, en particulier lorsque ces éléments ne sont pas des compositions géométriques simples comme par exemple des maisons, des blocs, des fenêtres, etc. Dans ces cas, l'extraction des contours des objets, qui apparaissent comme des éléments géométriques simples comme des lignes ou des courbes, et leur organisation permettent de les

identifier. Une image de foule n'apparait pas comme composé d'objets géométriques simples, mais plutôt comme présentant une surface globale complexe qui peut être mieux qualifiée par la notion de texture de la surface. Une texture fine décrit une foule dense alors qu'une texture grossière décrit plutôt une foule moins dense, plus éparpillée. Nous cherchons donc des mesures sur l'image qui nous permettent de caractériser la texture. Il devient dès lors possible de dériver une relation entre le niveau de texturisation et le nombre de personnes dans différentes régions d'une image.

Les descripteurs de texture associés aux images de foule se regroupent en 4 grandes catégories :

1. Matrices de cooccurrence
2. Spectre de fréquences
3. Dimension fractale de Minkowski
4. Moments invariants orthonormaux de Chebyshev

Matrices de cooccurrence Une matrice de cooccurrence $H(i, j \mid d, \theta)$ indique la distribution des couples de pixels de niveau de gris i et j séparés d'une distance de d pixels et formant un angle de θ entre les deux pixels. La matrice $H(i, j \mid d, \theta)$ est donc un tableau dont l'axe vertical représente le niveau de gris i du premier pixel du couple et dont l'axe horizontal représente le niveau de gris j du second pixel, et ce, pour une valeur d et une valeur θ données. En général, une distance d unitaire est utilisée et 4 orientations sont étudiées, $\theta = 0^o$, 45°, 90° et 135°. On calcule donc 4 matrices de cooccurrence. Prenons un exemple pour illustrer la construction de la matrice pour d=1 et $\theta = 0^o$: La case de la matrice de coordonnées $i = 10$ et $j = 15$ indique le nombre de fois que l'on retrouve 2 pixels voisins sur une même ligne avec le pixel à gauche affichant une valeur de niveau de gris de 10 et celui à droite, une valeur de 15.

Les matrices obtenues sont difficiles à interpréter. (Haralick, 1979) propose différents paramètres calculés sur les matrices obtenues afin de caractériser différents types de

texture. Quatre de ces paramètres figurent parmi les plus utilisés pour caractériser la texture : Le contraste, l'homogénéité, l'énergie et l'entropie. Ces 4 mesures permettent de quantifier et caractériser les principales textures.

(Marana et al., 1998; Wu et al., 2009) utilisent ces 4 paramètres pour caractériser les textures de foule en 5 groupes de densité :

- TF densité très faible 0-15 personnes
- F densité faible 16-30 personnes
- M densité moyenne 31-45 personnes
- H haute densité 46-60 personnes
- TH très haute densité > 60 personnes

Quatre matrices de cooccurrence (d = 1, θ = 0°, 45°, 90° et 135°) sont formées pour chaque fenêtre de 20x20 sur l'image, avec un déplacement de 10 pixels entre 2 fenêtres consécutives dans (Marana et al., 1998). Les résultats obtenus sont précis à 76,5% pour la haute et très haute densité, ce qui constitue une amélioration par rapport aux résultats obtenus par (Davies, Yin et Velastin, 1995) (section 2.2.2.1) avec la méthode de densité de pixels classés « foule ». Pour leur part, (Wu et al., 2009) utilisent des fenêtres d'analyse de dimension variable pour tenir compte de l'effet de perspective et l'échelle caractéristique de chaque fenêtre est précisément calculée en se basant sur l'algorithme SIFT discuté brièvement dans la section précédente sur l'invariance à l'échelle. Les auteurs en arrivent ainsi à une erreur d'estimation de foule dans chaque cellule inférieure à 5% et à une erreur inférieure à 12% pour toute l'image de foule.

Spectre de fréquence Une image peut être représentée par son contenu fréquentiel. La transformée de Fourier discrète appliquée sur l'image de foule produit une image $S(r, \theta)$ qui contient les fréquences spatiales de l'image de foule exprimées dans un système de

coordonnées polaires. Des mesures calculées sur le spectre permettent de différencier les différentes textures présentes. Les textures ainsi identifiées selon le contenu spectral correspondent à différentes densités de foule. Les foules denses correspondent à des zones de fréquences élevées tandis que les foules peu denses ont un contenu spectral de plus basses fréquences. (Marana et al., 1997) ont comparé la méthode des matrices de cooccurrence avec celle du spectre de fréquence pour conclure à des résultats comparables pour les deux méthodes, avec un léger avantage pour la méthode spectrale pour des foules de densité faible et moyenne.

Dimension fractale de Minkowski La dimension fractale est une mesure qui permet de caractériser des objets de forme irrégulière et permet de quantifier leur rugosité, et donc leur texture. La dimension fractale pour une image est un chiffre compris entre 1 et 2 qui permet de décrire une image irrégulière. En pratique, la dimension fractale est évaluée en extrayant tout d'abord toutes les arêtes de l'image. Les arêtes sont dilatées par un élément structurant circulaire pour former des arêtes dilatées nommées *saucisses de Minkowski*. Un graphique logarithmique du nombre de pixels activés en fonction du rayon de l'élément structurant est ensuite tracé. La dimension fractale est obtenue comme étant 2 fois la pente de la droite obtenue.

(Marana et al., 1999) ont comparé 3 méthodes d'estimation de densité de foule : les matrices de cooccurrence, le spectre de fréquence et la dimension fractale de Minkowski. Les meilleurs résultats ont été obtenus avec les matrices de cooccurrence, suivi de la dimension fractale de Minkowski.

Moments invariants orthonormaux de Chebyshev Les moments sont de puissants outils statistiques reconnus comme des descripteurs globaux. Cette formulation particulière des moments permet le calcul pratique de moments d'ordre relativement élevé sur des images discrètes. L'analyse des différents moments calculés permet de classer les zones de l'image en coefficient de densité de foule.

(Rahmalan, Nixon et Carter, 2006) comparent leur nouvel algorithme basé sur les Moments invariants orthonormaux de Chebyshev aux méthodes de matrice de cooccurrence et de dimension fractale discutés précédemment et introduits par Marana et son équipe. La méthode des moments s'avère la meilleure méthode parmi les trois testées, mais les foules évaluées ne doivent pas être trop denses, ce qui pourrait occasionner des problèmes si on choisissait cette méthode pour le Quartier des spectacles. Les matrices de cooccurrence suivent de très près du point de vue performance, mais leur temps de calcul est d'un ordre de grandeur plus élevé que pour la méthode des moments. De son côté, l'algorithme basé sur la dimension fractale s'est retrouvé déclassé dans ces essais.

2.2.2.3 Mouvement

Les approches locales qui exploitent le mouvement sont très répandues et sont souvent appliquées dans le domaine de la surveillance vidéo. Elles sont basées sur la détection des changements et l'analyse des trajectoires reliées aux nuées dans une scène. Les techniques reposent sur l'analyse de séquences chronologiques d'images ou de trames captées par une caméra vidéo (Kong, Gray et Tao, 2005).

Dans un premier temps, on applique typiquement une segmentation afin de soustraire l'arrière-plan de l'image **I** et repérer les formes pertinentes dans l'avant-plan. Ensuite, on détecte les formes correspondant à une seule personne ou à un petit groupe de personnes. La majorité des auteurs dans la littérature appliquent alors une poursuite des caractéristiques détectées. C'est-à-dire que les caractéristiques similaires mesurées dans une séquence de trames sont catégorisées selon la trajectoire d'une forme dans la scène (Antonini et Thiran, 2006).

Plusieurs techniques sont décrites dans la littérature pour représenter les formes par des caractéristiques simples, telles que la surface de la forme (somme des pixels) ou son périmètre (somme des pixels du contour) (Haritaoglu, Harwood et Davis, 2000). Afin

d'améliorer la qualité des diverses catégorisations, (Kong, Gray et Tao, 2005) proposent des caractéristiques plus robustes basées sur des histogrammes de dimension et d'orientation de contours. Les auteurs proposent aussi une normalisation des caractéristiques pour compenser les variations associées à l'orientation et la perspective de projection d'une caméra.

(Rittscher, Tu et Krahnstoever, 2005) extraient des points d'ancrage selon les propriétés locales du contour. Un algorithme spécialisé de type E-M est utilisé pour regrouper ces caractéristiques en rectangle ayant la dimension d'une tête. (Hou et Pang, 2011) veulent estimer le nombre de personnes et localiser chaque personne dans la scène. Ils appliquent un seuillage aux pixels de l'avant-plan afin de les binariser, ainsi que différentes techniques pour extraire les relations entre pixels et ressortir leur profil de mouvement.

(Rabaud et Belongie, 2006) proposent de partitionner des foules denses en formes semi-rigides par le biais de la catégorisation et de la poursuite de points caractéristiques à bas niveau. L'algorithme de Kande-Lucas-Tomasi est utilisé pour déduire les trajectoires d'une séquence vidéo et ensuite pour effectuer une analyse de mouvement dans le domaine spatio-temporel. Les auteurs proposent de catégoriser les trajectoires avec des contraintes de rigidité locales.

Dans (Reisman et al., 2004), les auteurs proposent une système pour détecter le mouvement de piétons d'une foule en temps réel. Une analyse spatio-temporelle de séquences vidéo permet de déduire le patron de mouvement correspondant aux différentes formes dans la scène. Un algorithme pour estimer les flux des champs de vitesse est proposé dans (Cong et al., 2009) afin de compter le nombre de pixels et de contours (i.e., piétons) traversant une zone de la scène.

Finalement, certains auteurs comme (Viola, Jones et Snow, 2005) et (Ryan et al., 2009; Ryan et al., 2010) exploitent des caractéristiques d'apparence et de mouvement pour

détecter des formes. Dans (Hou et Pang, 2011), on détecte les têtes avec un modèle 3D ellipsoïde pour vérifier les pics locaux sur le contour de la forme en avant-plan. Ainsi, on modélise conjointement le processus de détection et de poursuite comme un problème Maximum A Posteriori (MAP).

2.2.2.4 Apparence - contour de têtes

Un autre exemple de descripteurs locaux est présenté dans l'article de (Lin, Chen et Chao, 2001). Les auteurs ont proposé une approche pour le comptage de foules qui est basée sur la détection de contours de tête dans une image **I**. Les formes dans **I** ressemblant à un contour de tête sont extraites via un transformée en ondelette de type Haar. Lors d'une phase préliminaire, un modèle de l'apparence d'un contour de la tête est obtenu à partir d'une base d'environ 1000 images réelles en tons de gris (16x16 pixels). On extrait les coefficients en exploitant différentes représentations en ondelettes de Haar (fines et grossières), suivi d'une analyse statistique afin d'isoler les coefficients les plus importants. Cette technique tire avantage des informations sur l'apparence d'une tête pour extraire des caractéristiques locales.

Une technique semblable qui exploite aussi l'apparence d'une tête selon une transformée de Haar (en plus d'informations sur le mouvement, réf. Section 2.2.2.4) pour la détection de personnes en mouvement est présentée dans (Viola, Jones et Snow, 2005). Il s'agit d'une extension de l'algorithme Viola-Jones, initialement proposé pour la détection de visages. La technique nécessite par contre le passage d'une fenêtre de détection avec représentation en ondelettes de Haar sur 2 trames consécutives dans une séquence vidéo.

2.2.3 Classification et comptage

Suite à l'extraction des descripteurs, les caractéristiques pertinentes sont classifiées par le système pour le comptage de foule. Dans une phase préliminaire, un modèle de la foule est dérivé, permettant au classificateur de comparer les vecteurs caractéristiques $\mathbf{v}_i$

extraits de l'image aux classes de reconnaissance. L'étape de classification est souvent fortement couplée avec la dernière étape du traitement qui consiste à estimer le nombre total d'individus dans l'image en tenant compte de différentes informations contextuelles, telles que les variations dans les dimensions, densité et distributions de la foule, afin d'éliminer les ambigüités.

Pour les *approches globales* (densité de pixels, texture, etc.), on applique généralement un algorithme de classification multi-classe ou un modèle de régression, et les sorties (classes de reconnaissance ou variables de régression, respectivement) correspondent à différents niveaux de densité de la foule dans une des régions d'intérêt de **I.** Dans ce cas, un expert humain doit à priori estimer le nombre de personnes par unité de surface selon les densités perçues. Pour les *approches locales* (mouvement et apparence), on applique généralement un classificateur à 1- ou 2-classes, et les classes de reconnaissance correspondent à l'occurrence (ou non) d'un mouvement humain ou d'un contour de tête. Un expert humain doit à priori caractériser le mouvement ou l'apparence qui seront détectées, souvent par le biais d'une base de données réelle.

Le classificateur interagit avec son modèle de la foule afin de réaliser les comparaisons entre vecteurs caractéristiques $\mathbf{v}_i$ d'images et classes de reconnaissance. Cette modélisation peut être de nature *explicite*, et un expert humain doit alors caractériser la loi de classification avec un système d'équations. On retrouve typiquement cette modélisation explicite dans certaines techniques qui adoptent l'approche globale, pour décrire la relation entre le nombre de pixels classés « foule » (par unité de surface) et le nombre de personnes. Il est possible de dériver la relation entre le nombre pondéré de pixels et de contours de régions et le nombre de personnes grâce à une analyse par régression ou une technique de correction géométrique (Davies, Yin et Velastin, 1995).

Une approche plus commune dans la littérature pour établir une correspondance entre les caractéristiques et le nombre de personnes est la modélisation *implicite*. Dans ce cas, les paramètres d'un algorithme de classification ou de régression (statistique ou neuronique)

sont estimés ou appris à partir d'une base de données réelle. Plusieurs auteurs ont appliqués des algorithmes d'apprentissage de type supervisé : Système à vaste marge (SVM) (Arandjelovic, 2008; Lin, Chen et Chao, 2001; Wu et al., 2009), Perceptron multicouche (Kong, Gray et Tao, 2005), AdaBoost (Viola, Jones et Snow, 2005), analyse par régression (Chan, Liang et Vasconcelos, 2008; Kong, Gray et Tao, 2005; Lu et al., 2010; Ryan et al., 2009; Ryan et al., 2010). Certains auteurs ont aussi utilisé des algorithmes d'apprentissage de type non-supervisé comme la carte auto-organisatrice de Kohonen (*SOM – Self-Organizing Map*) (Marana et al., 1998) (Rahmalan, Nixon et Carter, 2006) afin de réduire la dimensionnalité de l'espace des caractéristiques, et pour classifier les régions d'intérêt selon des catégories de densité de foule. On peut ensuite compter le nombre total de personnes car ces catégories correspondent à des nombres de personnes par unité de surface.

2.3 Discussion

Il n'existe présentement aucune technique robuste dans la littérature pour l'estimation du nombre d'individus qui occupent une scène dense à partir d'une seule image statique.

Les approches basées sur des descripteurs globaux sont plus communes et permettent un dénombrement plus rapide d'une foule. Cependant, dans des scènes très denses, il est difficile de segmenter la foule de l'arrière plan, et la présence d'occlusion et de chevauchement entre les individus complique l'isolement des individus pour le décompte. Le défi avec ce genre d'approche est le manque d'informations spatiales dans le modèle de foule. Les personnes plus loin de la caméra contribuent moins à l'estimation de la foule. Ces approches demandent aussi des stratégies complexes pour le comptage final dû aux différents comportements, distributions, densités et dimensions de la foule.

La performance des approches locales basées sur l'analyse du mouvement (détection de changements et poursuite de nuées) dégénère rapidement avec la densité des foules et

l'occlusion car l'extraction précise d'objets de l'arrière-plan devient alors très difficile. Les individus doivent typiquement être en mouvement. De plus, il est très difficile de suivre la trajectoire de plusieurs nuées simultanément dans une scène complexe. La complexité algorithmique de ces approches devient problématique pour les systèmes en temps-réel. Dans les approches locales basées sur l'apparence (détection de contour de tête), le détecteur peut se dégrader avec des changements d'illumination et de perspective, et peuvent aussi avoir un cout computationnel élevé. Dans (Lin, Chen et Chao, 2001), on propose d'appliquer une transformée sur l'image afin de stabiliser les effets de la perspective. Par contre, une fois que les têtes ont été détectées, soit par le mouvement ou l'apparence, le comptage de foules s'obtient plus directement qu'avec les approches globales.

Un défi pour l'application pratique de toutes les techniques de comptage de foules en littérature est le besoin de collecter et d'analyser des données à partir d'images représentatives. Ces données servent pour la soustraction d'arrière-plans sans foule et pour la conception des classificateurs. La performance des classificateurs de pointe est très dépendante de la quantité et de la qualité des données utilisées pour l'estimation des paramètres du classificateur.

3 APPROCHES COMMERCIALES

Notre mandat comprend le recensement de produits commerciaux qui pourraient être utilisés directement ou adaptés afin d'estimer le nombre de personnes présentes sur les différents espaces publics du Quartier des spectacles. Des caméras haute définition disposées sur le toit d'édifices situés dans le périmètre du quartier des spectacles fourniront des images à haute définition de la foule présente. L'estimé du nombre de personnes sera produit à partir de l'analyse informatique de ces images.

Nous recensons dans cette section les logiciels commerciaux qui proclament pouvoir estimer le nombre de personnes dans une foule, de même que les brevets délivrés en Occident qui abordent la problématique du comptage de foule.

3.1 Logiciels

Il n'existe que fort peu de logiciels destinés au dénombrement de foule. Une raison possible demeure qu'il n'existe qu'un petit marché pour un tel produit. Les groupes intéressés par de telles données se contentent souvent de données très approximatives, avec un biais qui dépend de qui demande les statistiques de foule : l'organisme qui organise une manifestation, le corps policier qui veille à la sécurité ou la presse qui rapporte l'évènement.

Nous n'avons retrouvé que 2 logiciels disponibles pour évaluer des foules : Le logiciel *ExactCrowd* de l'Espagnole Lynce et le système *Automated footfall monitoring* de l'Anglaise Springboard.

3.1.1 *ExactCrowd* de Lynce

Un débat intense sévit depuis quelques mois en France sur le comptage des manifestants lors de grandes manifestations. Des écarts considérables sont constatés régulièrement entre les chiffres fournis par les organisateurs de la manifestation, ceux des corps

policiers et les chiffres utilisés par la Presse, qui ne sait plus trop qui croire. Le journal France-Soir (Gonzalez, 2010, 14 octobe) a voulu en avoir le cœur net et a fait appel à la société espagnole Lynce, spécialisée dans le comptage de manifestants, pour mesurer le rassemblement parisien du mardi 12 octobre 2010 à l'occasion de l'une des manifestations sur l'âge de la retraite. Pour cette manifestation en particulier, Lynce-ExactCrowd a délivré les chiffres mesurés à France-Soir : « *Nous avons mesuré un total de 73.027 personnes, avec une correction éventuelle à la hausse de 10 %, soit un total de 80.330 personnes.* » Les organisateurs de la manifestation ont compté 330 000 personnes et la police a estimé la foule à 89 000 personnes !

La méthode de comptage offerte par la société Lynce se base sur un logiciel, *ExactCrowd*, développé en 2008 par une équipe d'ingénieurs espagnols et qui comptabilise les manifestants un à un à partir de photographies prises en hauteur (Lynce, 2009). Pour remplir son mandat, Lynce a déployé toute une équipe pour décompter la foule présente à cette manifestation parisienne du 12 octobre. L'instrumentation déployée comprend en général des ballons aériens munis de caméras et plusieurs photographes disposés en hauteur sur des immeubles le long de la manifestation (Figure 11). Pour cette manifestation en particulier, les ballons ne pouvaient être utilisés à cause de rafales de vent. Les photographes s'installent à leur poste d'observation environ 2 heures avant le début du défilé. Les photographies prises doivent être d'une très grande netteté car elles devront être zoomées afin de différencier chaque manifestant. Près de 1 000 photos à très haute résolution seront captées durant toute la manifestation. Cinq heures d'analyse par une seconde équipe composée de 3 ingénieurs experts et du président de Lynce seront nécessaires pour produire les résultats finaux. Le logiciel *ExactCrowd* analyse chacune des photos, numérote chaque manifestant afin d'éviter les dédoublements d'une photo à la suivante. Un algorithme d'analyse du mouvement est également utilisé afin de suivre à la trace chaque individu sur une séquence d'images captées le long du parcours de la manifestation. L'équipe d'ingénieurs tentent de réduire les ambigüités entre les séquences d'images afin qu'une même personne ne soit pas

comptée plus d'une fois. Ceci devient difficile et couteux pour des foules denses avec la présence d'occlusions.

Figure 11 : Équipement utilisé par Lynce pour le comptage de foule : (1) caméra vidéo haute définition; (2) ballon aérien; (3) appareil photo numérique haute définition.

La méthode développée par Lynce s'avère fort utile pour dénombrer avec précision de grandes foules mais s'avère trop onéreuse, trop précise et beaucoup trop longue pour les besoins du PQS.

3.1.2 *Automated Footfall Monitoring* de Springboard

La compagnie Springboard (Springboard high street, 2002) offre sur internet un service de comptage d'une zone définie par le demandeur. La technologie offerte par Springboard permet de surveiller de façon continue les véhicules ou les piétons qui traversent la zone désignée. Une caméra est installée en hauteur (sur le mur d'un édifice par exemple) pour enregistrer les piétons qui franchissent la zone. Le nombre de piétons est enregistré grâce à un logiciel propriétaire basé sur la poursuite de cible spécifique.

Les données sont vérifiées quotidiennement par Springboard avant d'être transférées et stockées sur un serveur internet. Le demandeur accède aux rapports de surveillance directement sur le site internet.

Le logiciel de surveillance est bâti essentiellement sur la poursuite de piétons en mouvement. De plus, la zone couverte est beaucoup plus restreinte que les zones que le PQS aimerait décompter. Le logiciel proposé par Springboard est utilisé pour la surveillance complète de petites zones. Le logiciel a été développé à l'origine pour les milieux de la défense et de la sécurité pour obtenir des résultats précis sur de petits groupes de piétons en mouvement. Il ne nous apparait pas adapté à notre projet de comptage de foule dans de grands espaces publics.

3.2 Brevets

Nos recherches parmi les brevets délivrés au Canada, aux États-Unis et en Europe ont permis de trouver trois brevets qui abordent la problématique du comptage de foule. Comme avec les systèmes commerciaux, deux de ces brevets exploitent une approche locale basée sur le mouvement et sont ciblés pour des applications en surveillance vidéo. Les techniques reposent sur la détection des changements et l'analyse des trajectoires reliées aux formes dans des séquences vidéo.

Dans (Kilambi, Masoud et Papanikolopoulos, May 22, 2008), on propose une technique pour détecter et suivre la trajectoire associée avec des personnes dans une séquence vidéo. Cette technique est ciblée pour un contrôle plus sécuritaire du trafic piétonnier en temps-réel. Suite à une soustraction de l'arrière-plan, on détermine les zones de l'image avec chevauchement. Les formes dans une des images sont identifiées et suivies avec un filtre de Kalman. Un classificateur est entrainé avec des données réelles afin de corréler les zones de chevauchement avec le nombre de personnes. Enfin, le nombre de personnes est estimé à partir de la surface occupée par l'ensemble des formes.

Dans (Rittscher et al., September 29, 2009), on propose un système pour détecter et compter le nombre d'individus dans une foule non-stationnaire, à partie d'une seule caméra. Un modèle géométrique statique est conçu à partir de données (images de référence) afin de représenter les caractéristiques d'image correspondant à une personne dans la foule. En opération, la trajectoire de chaque personne est déterminée afin de réduire les occlusions entre personnes. Enfin, une porte virtuelle est définie pour compter le nombre de personnes qui se déplacent à travers une zone spécifique de l'image.

Dans (Gonzalez-Banos, Yang et Guibas, January 17, 2006), on propose un système pour déterminer le nombre de personnes dans une foule à l'aide d'informations sur la convexité visuelle. Dans un premier temps, on extrait les images silhouettes à partir des images captées avec un réseau de caméras. Le plan de projection est déterminé selon la convexité visuelle de la foule – on projette sur une plan l'intersection entre le cône de chaque image silhouette et le volume analysé. Les projections planaires des images sont ensuite agrégées selon chaque intersection afin de produire un patron de subdivisions. Finalement, ces patrons servent de support pour construire une structure en arbre de la scène. Cette structure évolue dans le temps, en fonction de nouvelles images. Le nombre de personnes dans une foule est estimé grâce aux nœuds de cette structure.

Tel que mentionné au chapitre 2, les approches locales basées sur l'analyse du mouvement ne sont pas performantes pour des scènes denses et statiques. Les deux premiers brevets délivrés (Kilambi, Masoud et Papanikolopoulos, May 22, 2008; Rittscher et al., September 29, 2009) ont été proposés pour des scénarios en vidéosurveillance, pour la détection et la poursuite d'un nombre restreint de personnes en mouvement. Ils ne sont pas pertinents pour le comptage de foules lorsque les scènes sont denses et les individus dans la foule ne sont pas en mouvement. Le troisième brevet (Gonzalez-Banos, Yang et Guibas, January 17, 2006) doit exploiter simultanément des

caméras multiples pour analyser une même scène, et ne respecte donc pas les contraintes de l'application.

4 APPROCHES PROPOSÉES

Devant l'absence de solutions commerciales au comptage de foule généralement dense, nous proposons le développement et le test sur des images réelles de foule au QS des algorithmes les plus prometteurs identifiés lors de la revue de documentation détaillée au chapitre 2.

4.1 Méthodes prometteuses

Nous reprenons le schéma de la Figure 6 : Système général pour le comptage de foule. Le schéma illustre les différentes étapes du logiciel de comptage de foule.

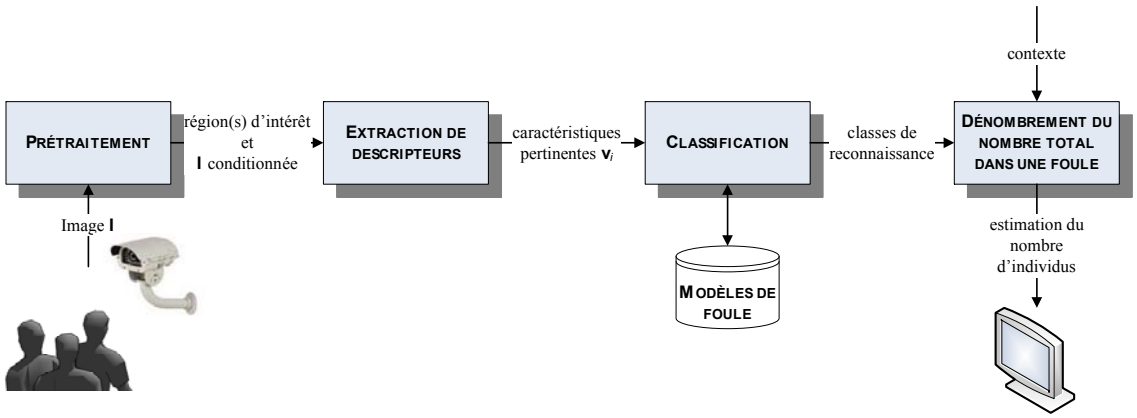

Figure 12 : Structure du logiciel de comptage de foule.

Tableau 1 : Sélection des algorithmes pour chaque étape du système de comptage de foule.

Étape	Algorithmes	Références	Note
Prétraitement	Calibration et grillage de la zone étudiée	(Wu et al., 2009)	1
	Soustraction de l'arrière-plan sans foule ou segmentation invariante à l'échelle en zones « foule » et « non-foule »	(Davies, Yin et Velastin, 1995; Arandjelovic, 2008)	2
	Corrections apportés à l'image : 1. Égalisation d'histogramme (effets de l'éclairage) 2. Redressement de l'image (caméra en hauteur et de biais) 3. Effet de perspective	1. (Lin, Chen et Chao, 2001) 2. (Davies, Yin et Velastin, 1995) Algorithmes de redressement issus de la télédétection 3. (Kong, Gray et Tao, 2005)	
Extraction de descripteurs	Texture : Matrices de cooccurrence	(Marana et al., 1998)	
	Texture : Moments invariants orthonormaux de Chebyshev	(Rahmalan, Nixon et Carter, 2006)	
	Apparence : Contour de têtes	(Lin, Chen et Chao, 2001)	
Classification	Système à vaste marge (SVM)	(Arandjelovic, 2008; Lin, Chen et Chao, 2001; Wu et al., 2009)	3
Dénombrement	Densité d'occupation de la foule × surface de la zone		4

Notes :

1. Nous devrons être en mesure de calculer l'équivalence entre les distances mesurées sur l'image et les distances physiques sur le Quartier des Spectacles. Cette relation sera linéaire si les effets de perspective ne sont pas importants,

mais dans le cas probable contraire, il faudra calibrer les images à l'aide de mesures réalisées sur le terrain, possiblement avec du matériel d'arpentage.

2. La méthode de soustraction de l'arrière-plan sans foule est la plus simple à réaliser en principe. Dans notre cas, cette mesure pourrait s'avérer plus difficile à réaliser que prévu. Les caméras seront munies de zoom ajustable, et donc les photographies d'arrière-plan sans foule qui doivent couvrir les mêmes superficies que les images de foule seront vraisemblablement difficiles à obtenir. De plus, les décors et accessoires disposés sur les scènes extérieures changent fréquemment et peuvent être déplacés à l'intérieur d'un même festival. Il n'est pas évident non plus d'obtenir des images sans foule. De plus, idéalement, nous devrions pouvoir disposer de telles images à différentes périodes de la journée. Pour toutes ces raisons, nous devons envisager de coder l'algorithme de segmentation invariante à l'échelle proposé par (Arandjelovic, 2008). L'auteur fournit peu de détails pour l'implémentation pratique de l'algorithme, mais nous devrons probablement pousser dans cette direction.

3. Peu importe l'approche utilisée pour le dénombrement de foules, il sera nécessaire de paramétrer un classificateur lors d'une phase préliminaire d'apprentissage. Cette phase fait appel à une base représentative de données et produit un modèle de foule.

4. Une fois la densité de foule estimée pour des parcelles bien identifiées de la scène, le nombre de personnes présentes sur la parcelle s'obtient directement par la connaissance de la surface de la parcelle. D'où la nécessité de pouvoir mesurer sur l'image des périmètres de parcelle que l'on puisse faire correspondre précisément aux mesures réelles et physiques sur le terrain.

4.2 Évaluation expérimentale

Nous profiterons du trimestre en cours (Hiver 2011) pour évaluer 2 des 3 méthodes proposées d'extraction de descripteurs. Deux étudiants en fin de B.Ing. qui sont inscrits en PFE (projet de fin d'étude) vont s'attaquer à cette problématique. Les résultats de leurs travaux seront exposés sous forme d'affiche à la fin du trimestre et chacun produira un rapport décrivant sa démarche et les résultats obtenus.

Les deux méthodes évaluées sont :

1. Caractérisation de la texture représentant la foule sur l'image au moyen de matrices de cooccurrence.

2. Dénombrement de la foule par la détection des contours de tête.

5 CONCLUSION

Dans cette première étape du projet, nous avons réalisé une revue complète de la littérature scientifique, ainsi que des brevets et des produits commerciaux. Il n'existe aucune solution commerciale répondant aux besoins du PQS.

La revue exhaustive de la littérature laisse présager quelques pistes intéressantes de solution. Nous avons identifié 2 méthodes qui peuvent être rapidement programmées pour traiter les images réelles fournies par le PQS. Une première méthode associe chaque texture mesurée sur une portion de l'image à une mesure de densité de foule, et la somme de toutes les densités partielles de l'image fournira le nombre recherché de personnes présentes. La seconde méthode modélise les têtes des participants au moyen d'une transformée appliquée sur l'image, ce qui devrait permettre de compter un nombre de personnes suffisamment précis pour les besoins de l'application. Ces deux méthodes font présentement l'objet du projet de fin d'études de 2 étudiants finissants en génie de la production automatisée. Les résultats seront présentés au colloque « Vidéo numérique : Traitement, analyse et reconnaissance » qui se tiendra vendredi le 13 mai 2011 dans le cadre du 79ème Congrès de l'Acfas à Sherbrooke du 9 au 13 mai 2011.

BIBLIOGRAPHIE

Antonini, G., et J. P. Thiran. 2006. « Counting Pedestrians in Video Sequences Using Trajectory Clustering ». IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, n° 8, p. 1008-1020.

Arandjelovic, Ognjen. 2008. « Crowd Detection from Still Images ». In *19th British Machine Vision Conference - BMVC* (Leeds, UK, 1-4 september). p. 10 p.

Chan, A. B., Z. S. J. Liang et N. Vasconcelos. 2008. « Privacy preserving crowd monitoring: Counting people without people models or tracking ». In *IEEE Conference on Computer Vision and Pattern Recognition - CVPR* (23-28 June 2008). p. 1-7.

Cong, Yang, Haifeng Gong, Song-Chun Zhu et Yandong Tang. 2009. « Flow Mosaicking: Real-time Pedestrian Counting without Scene-specific Learning ». In *Computer Vision and Pattern Recognition - CVPR* (Miami (Florida), 22-24 June). p. 1093-1100.

Davies, Anthony C., Jia Hong Yin et Sergio A. Velastin. 1995. « Crowd Monitoring Using Image Processing ». IEE Electronic and Communications Engineering Journal, vol. 7, n° 1 (February), p. 37-58.

Denman, Simon, Vinod Chandran et Sridha Sridharan. 2007. « An adaptive optical flow technique for person tracking systems ». Pattern Recognition Letters, vol. 28, n° 10, p. 1232-1239.

Doig, Steve. 2009, 15 january. *How big will inaugural crowd be? Do the math - When people gather in vast numbers, 'official' estimates often run wild*. Journal en ligne msnbc.com. <http://www.msnbc.msn.com/id/28662672/>. Consulté le 10 décembre 2010.

Duda, Richard O., Peter E. Hart et David G. Stork. 2001. *Pattern classification*. New York ; Toronto: New York ; Toronto : John Wiley & Sons.

Gonzalez, Alexandra. 2010, 14 octobe. *Comment nous avons compté les manifestants du mardi 12 octobre | France Soir.pdf*. Journal France Soir en ligne.

Gonzalez-Banos, Hector H., Danny B. Yang et Leonidas J. Guibas. January 17, 2006. *Systems and methods for using visual hulls to determine the number of people in a crowd*. Patent, United States (inv.). 24 p. US 6,987,885 B2.

Haralick, R.M. 1979. « Statistical and structural approaches to texture ». Proceedings of the IEEE*,* vol. 67, n° 5, p. 786-804.

Haritaoglu, I., D. Harwood et L. S. Davis. 2000. « W^4: real-time surveillance of people and their activities ». IEEE Transactions on Pattern Analysis and Machine Intelligence*,* vol. 22, n° 8, p. 809-830.

Hou, Ya-Li, et G. K. H. Pang. 2011. « People Counting and Human Detection in a Challenging Situation ». IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*,* vol. 41, n° 1, p. 24-33.

Hou, Ya-li, et Grantham K. H. Pang. 2008. « Automated people counting at a mass site ». In *IEEE International Conference on Automation and Logistics - ICAL* (1-3 Sept. 2008). p. 464-469.

Kilambi, Prahlad, Osama T. Masoud et Nikolaos Papanikolopoulos. May 22, 2008. *Crowd counting and monitoring*. Publication, United States Patent Application (inv.). 20 p. US 2008/0118106 A1.

Kong, Dan, Doug Gray et Hai Tao. 2005. « Counting Pedestrians in Crowds Using Viewpoint Invariant Training ». In *British Machine Vision Conference*.

Lin, Sheng-Fuu, Jaw-Yeh Chen et Hung-Xin Chao. 2001. « Estimation of number of people in crowded scenes using perspective transformation ». IEEE Transactions

on Systems, Man and Cybernetics, Part A: Systems and Humans, vol. 31, n° 6, p. 645-654.

Lowe, David G. 2004. « Distinctive Image Features from Scale-Invariant Keypoints ». International Journal of Computer Vision, vol. 60, n° 2, p. 91-110.

Lu, Zhenyu, Jungho Im, Lindi Quackenbush et Kerry Halligan. 2010. « Population estimation based on multi-sensor data fusion ». International Journal of Remote Sensing, vol. 31, n° 21, p. 5587-5604.

Lynce. 2009. *Exactcrowd*. Logiciel de dénombrement de foule. Espagne: Lynce.

Marana, A. N., L. Da Fontoura Costa, R. A. Lotufo et S. A. Velastin. 1999. « Estimating crowd density with Minkowski fractal dimension ». In *IEEE International Conference on Acoustics, Speech, and Signal Processing. ICASSP '99. Proceedings.* (15-19 Mar 1999). Vol. 6, p. 3521-3524.

Marana, A. N., S. A. Velastin, L. F. Costa et R. A. Lotufo. 1998. « Automatic estimation of crowd density using texture ». Safety Science, vol. 28, n° 3, p. 165-175.

Marana, A.N., S.A. Velastin, L.F. Costa et R.A.Lotufo. 1997. « Automatic Estimation of Crowd Density Using Texture ». In *International Workshop on Systems and Image Processing (IWSIP'97)* (Polland, May 28-30).

Melina, Remy. 2010, 3 September. *How is Crowd Size Estimated?* Journal en ligne : Life's Little Mysteries. <http://www.lifeslittlemysteries.com/how-is-crowd-size-estimated--1074/>.

Rabaud, V., et S. Belongie. 2006. « Counting Crowded Moving Objects ». In *IEEE Computer Society Conference onComputer Vision and Pattern Recognition* (17-22 June 2006). Vol. 1, p. 705-711.

Rahmalan, H., M. S. Nixon et J. N. Carter. 2006. « On Crowd Density Estimation for Surveillance ». In *The Institution of Engineering and Technology Conference on Crime and Security* (13-14 June 2006). p. 540-545.

Reisman, P., O. Mano, S. Avidan et A. Shashua. 2004. « Crowd detection in video sequences ». In *IEEE Intelligent Vehicles Symposium* (14-17 June 2004). p. 66-71.

Rittscher, Jens, Peter Henry Tu, Nils Oliver Krahnstoever, Amitha Perera, Thomas Baby Sebastian et Siaoming Liu. September 29, 2009. *System and method for automatic person counting and detection of specific events*. Publication, United States Patent Application (inv.). 15 p. US 7,596,241 B2.

Ryan, D., S. Denman, C. Fookes et S. Sridharan. 2009. « Crowd Counting Using Multiple Local Features ». In *Digital Image Computing: Techniques and Applications. DICTA '09.* (1-3 Dec. 2009). p. 81-88.

Ryan, David, Simon Denman, Clinton Fookes et Sridha Sridharan. 2010. « Crowd Counting Using Group Tracking and Local Features ». In *Seventh IEEE International Conference on Advanced Video and Signal Based Surveillance*. p. 218-224.

Springboard high street. 2002. *Automated footfall monitoring in towns and cities*. Logiciel d'estimation de mouvement de foule. United Kingdom: Springboard.

Viola, Paul, Michael J. Jones et Daniel Snow. 2005. « Detecting Pedestrians Using Patterns of Motion and Appearance ». International Journal of Computer Vision, vol. 63, n° 2, p. 153-161.

Wu, Xinyu, Guoyuan Liang, Ka Keung Lee et Yangsheng Xu. 2009. « Crowd Density Estimation Using Texture Analysis and Learning ». Bulletin of Advanced Technology Research, vol. 3, n° 5 (May), p. 35-41.

Zhan, Beibei, Dorothy N. Monekosso, Paolo Remagnino, Sergio A. Velastin et Li-Qun Xu. 2008. « Crowd analysis: a survey ». Machine Vision and Applications, vol. 19, n° 5-6, p. 345-357.