

**Titre:** Recalage automatique de séquences vidéo infrarouge et visible  
Title: basé sur les trajectoires des objets en mouvement

**Auteur:** François Morin  
Author:

**Date:** 2009

**Type:** Mémoire ou thèse / Dissertation or Thesis

**Référence:** Morin, F. (2009). Recalage automatique de séquences vidéo infrarouge et visible basé sur les trajectoires des objets en mouvement [Mémoire de maîtrise, École Polytechnique de Montréal]. PolyPublie. <https://publications.polymtl.ca/143/>

## Document en libre accès dans PolyPublie

Open Access document in PolyPublie

**URL de PolyPublie:** <https://publications.polymtl.ca/143/>  
PolyPublie URL:

**Directeurs de recherche:** Guillaume-Alexandre Bilodeau  
Advisors:

**Programme:** Génie informatique  
Program:

UNIVERSITÉ DE MONTRÉAL

RECALAGE AUTOMATIQUE DE SÉQUENCES VIDÉO INFRAROUGE ET  
VISIBLE BASÉ SUR LES TRAJECTOIRES DES OBJETS EN MOUVEMENT

FRANÇOIS MORIN

DÉPARTEMENT DE GÉNIE INFORMATIQUE ET GÉNIE LOGICIEL  
ÉCOLE POLYTECHNIQUE DE MONTRÉAL

MÉMOIRE PRÉSENTÉ EN VUE DE L'OBTENTION  
DU DIPLÔME DE MAÎTRISE ÈS SCIENCES APPLIQUÉES  
(GÉNIE INFORMATIQUE)  
AOÛT 2009

UNIVERSITÉ DE MONTRÉAL

ÉCOLE POLYTECHNIQUE DE MONTRÉAL

CE MÉMOIRE INTITULÉ :

RECALAGE AUTOMATIQUE DE SÉQUENCES VIDÉO INFRAROUGE ET  
VISIBLE BASÉ SUR LES TRAJECTOIRES DES OBJETS EN MOUVEMENT

présenté par : MORIN François

en vue de l'obtention du diplôme de : Maîtrise ès sciences appliquées

a été dûment accepté par le jury d'examen constitué de :

M. GUIBAULT François, Ph.D., président

M. BILODEAU Guillaume-Alexandre, Ph.D., membre et directeur de recherche

M. BOYER François-Raymond, Ph.D., membre

*À la mémoire d'une Hélène passée.*

*Dédié à une Hélène en devenir.*

## **REMERCIEMENTS**

J'aimerais d'abord remercier les membres du jury pour avoir bien voulu participer à l'évaluation de ce travail. J'aimerais plus spécifiquement remercier mon directeur de recherche, M. Guillaume-Alexandre Bilodeau, pour les nombreuses heures d'aide qu'il m'a accordée ainsi que la patience dont il a toujours fait preuve. Non seulement la réalisation de ce mémoire aurait été impossible sans ses conseils et son enseignement, mais il m'aurait été aussi très difficile de décrocher mon emploi.

Un merci tout spécial aussi à mes collègues ayant aussi œuvré au LITIV soit Pier-Luc St-Onge, Soufiane Ammour, Guillaume Desjardins, Atousa Torabi et Parisa Darvish Zadeh Varcheie qui m'ont aidé à réaliser mes séquences vidéo, mais aussi qui m'ont promulgués de précieux conseils.

Je remercie aussi mes amis et ma famille pour leur appui indéfectible et leurs encouragements qui m'ont permis de terminer ce que j'avais entrepris. Finalement, je remercie infiniment ma copine Amanda Fahey qui m'a appuyé et encouragé dans les moments les plus difficiles en plus de me faire cadeau de ma plus grande source d'inspiration, ma nouvelle fille Hélène. Hélène, tu es trop jeune pour t'en rendre compte maintenant, mais tu es vraiment la principale raison qui m'a poussé à terminer ce mémoire.

## RÉSUMÉ

Le projet présenté dans ce document porte sur une nouvelle méthode pour effectuer le recalage entre deux séquences vidéo de spectres différents soit l'une du spectre visible et l'autre du spectre infrarouge. Un recalage consiste à retrouver la matrice de transformation qui permet de passer d'un référentiel à un autre. Dans ce projet en particulier, il s'agit de trouver la matrice de transformation affine qui permet de passer du référentiel de la caméra infrarouge à celui de la caméra couleur. Ce recalage est effectué pour chaque trame. Pour y arriver, la méthode proposée se base sur les trajectoires des objets en mouvement ainsi que sur des combinaisons d'images d'avant-plan en recouvrement.

Plusieurs méthodes de recalage multispectral ont été proposées dans la littérature. Cependant, il n'est pas toujours possible d'appliquer ces techniques dans le cas spécifique d'un recalage entre des images couleur et infrarouge. Il existe cependant une méthode qui semble avoir un bon potentiel. Il s'agit de la méthode de Caspi, Simakov, & Irani (2006). Cette méthode utilise uniquement les trajectoires des objets en mouvement pour retrouver la matrice de transformation entre les deux caméras. Cet article fut la principale source d'inspiration de ce mémoire.

Pour pouvoir utiliser les trajectoires des objets contenus dans la scène, la méthode proposée dans les prochains chapitres effectue une soustraction d'arrière-plan par moyenne temporelle (Shoushtarian & Bez, 2005). La méthode extrait ensuite les « blobs » (groupes de pixels de l'avant-plan reliés entre eux) et trouve le sommet de chacun. Pour construire les trajectoires, elle utilise la méthode du recouvrement des rectangles englobants de « blobs », afin de relier entre eux les différents blobS. Pour le calcul de la matrice de transformation, un algorithme de type RANSAC est utilisé. Comme la matrice de transformation utilisée est de type affine, 3 points sont nécessaires pour en trouver une première estimation. L'algorithme utilisé choisit la meilleure matrice possible à travers tous les points d'une paire de trajectoires donnée en évaluant le pointage de recouvrement des images de pointages. Ces images de pointages sont

formées à partir d'une combinaison d'images d'avant-plan sauvegardée tout le long de la séquence. On ne conserve qu'un maximum de 5 images d'avant-plan dans ces images de pointages. Les trajectoires sont aussi appariées en utilisant un algorithme RANSAC. L'algorithme d'appariement des trajectoires utilise lui aussi le recouvrement des images de pointages pour vérifier si les trajectoires sont bien appariées. L'algorithme maximise le nombre de trajectoires bien appariées et minimise, en cas d'égalité, le pourcentage d'erreur de recouvrement. Un autre choix est fait après chaque trame afin de comparer la nouvelle matrice de transformation avec la précédente. Après avoir calculé le nouveau pointage de l'ancienne matrice avec la nouvelle image de pointage, les pointages des deux matrices sont comparés et la matrice avec le plus petit pointage est sélectionnée.

Cette nouvelle méthode a été comparée à une méthode dite de base qui suit l'idée de Caspi et al. (2006) ainsi qu'à un recalage effectué manuellement. Les résultats montrent que la nouvelle méthode surpassé la méthode de base sur 7 des 9 séquences vidéo testées. Les situations qui semblent privilégier la méthode de base sont les situations où le nombre de trajectoires est très grand et lorsque les erreurs de soustraction d'arrière-plan sont très fréquentes. Par contre, lorsque la soustraction d'arrière-plan et la génération de trajectoires sont bonnes, la nouvelle méthode surpassé toujours la méthode de base, et ce, même si la surface de recouvrement est réduite. De plus, les meilleurs résultats donnent des valeurs qui sont très près de celle du recalage manuel.

Finalement, la nouvelle méthode montre un grand potentiel. L'utilisation d'une meilleure méthode de génération de trajectoires ainsi qu'une soustraction d'arrière-plan plus efficace permettrait sûrement d'améliorer encore plus les résultats obtenus.

## ABSTRACT

This project is about a new registration method between a color video and an infrared video. Registration is to find the transformation matrix from which you can pass from one set of coordinates to another, for example, between two cameras. In this project, this transformation matrix is an affine transformation matrix. This registration is done frame by frame. It uses two kinds of information: the trajectories of moving objects and an overlapped picture made from combinations of foreground pictures.

Multiple multispectral registration methods were found in the literature but not all of these methods can be used for the specific case of an infrared-color registration. Even so, a specific method (Caspi et al., 2006) seems promising to achieve this kind of registration. This method uses only trajectories of moving objects to find the transformation matrix. This article is the principal inspiration of this work.

To use the trajectories in the scene, our new method begins by doing a temporal mean background subtraction. The next step is to extract the « blobs » (groups of interconnected pixels) from the foreground and find the summit of each one. To build the trajectories, our method matches « blobs » across frames with an overlapped bounding box technique. To compute the transformation matrix, a RANSAC algorithm is used. Because we use an affine transformation matrix, 3 points are needed to compute a first estimation. This algorithm chooses the best matrix between the multiple pairs of points forming the trajectory by comparing overlapping score pictures. These score pictures are made from a combination of foreground pictures chosen from several moments in the video. These score pictures are made from up to five of these foreground pictures. The trajectories are also matched using a RANSAC algorithm. This algorithm also uses the same idea of overlapping two score pictures to find the trajectories that are correctly matched. The algorithm maximises the number of well matched trajectories and minimises, in case of a tie, the overlapping error percentage. A last choice is made after each frame between the new transformation matrix found and the one from the previous frame. After updating the score of the old matrix with the new score pictures,

the two scores are compared and the matrix with the lowest score is saved for the future frames.

This new method is compared with the base method that was implemented using Caspi et al. (2006). It is also compared to a manual registration. The results show that the new method is better than the base method in 7 out of 9 videos tested. The base method does better than the new method only when the number of trajectories is really high and when there are too many errors in the background subtraction. However, when the background subtraction and the trajectories are good, the new method is always better than the base one even if the overlapped zone is reduced. More interestingly, the best cases give results really close to the ones from manual registration.

Finally, the new registration method seems promising. Using a better trajectory generating algorithm and a better background subtraction method should give even better results.

## TABLE DES MATIÈRES

DÉDICACE .....	III
REMERCIEMENTS .....	IV
RÉSUMÉ .....	V
ABSTRACT .....	VII
TABLE DES MATIÈRES .....	IX
LISTE DES TABLEAUX.....	XIV
LISTE DES FIGURES.....	XV
LISTE DES SIGLES ET ABRÉVIATIONS .....	XIX
INTRODUCTION .....	1
CHAPITRE 1 REVUE DE LA LITTÉRATURE.....	9
1.1 Matrices de transformation .....	9
1.1.1 Géométrie épipolaire .....	10
1.1.2 Les matrices de transformation homographique isométrique .....	12
1.1.3 Les matrices de transformation homographique de similarité .....	12
1.1.4 Les matrices de transformation homographique affine.....	13
1.1.5 Les matrices de transformation homographique de projection planaire ...	14
1.2 Méthodes classiques de recalage monospectral .....	14
1.2.1 Les méthodes basées sur les pixels .....	15
1.2.2 Les méthodes basées sur les caractéristiques .....	17
1.3 Méthodes de recalage infrarouge-visible basées sur les pixels.....	18
1.4 Méthodes de recalage infrarouge-visible basées sur les éléments caractéristiques d'une image .....	24

1.4.1	Méthodes basées sur le contour.....	24
1.4.2	Méthodes basées sur des points caractéristiques.....	26
CHAPITRE 2	MÉTHODOLOGIE.....	31
2.1	Aperçu de la méthode.....	31
2.2	Prétraitements.....	34
2.2.1	Images binaires d'avant-plan : soustraction d'arrière-plan.....	35
2.2.2	Images binaires d'avant-plan : extraction des grands objets binaires (« blobs ») .....	37
2.2.3	Calcul des trajectoires : algorithmes de suivi.....	38
2.3	Calcul de la matrice de transformation à partir des points de trajectoires .....	40
2.3.1	Justification du choix de la simplification .....	41
2.3.2	Calcul de la matrice de transformation affine .....	43
2.3.2.1	Première étape : identifier les points d'intérêts.....	45
2.3.2.2	Deuxième étape : déduire des correspondances.....	45
2.3.2.3	Troisième étape : estimation de type RANSAC .....	47
2.3.2.4	Quatrième étape : raffinement de la matrice de transformation.....	48
2.3.2.5	Calcul du nombre d'itérations dans un algorithme RANSAC .....	50
2.4	Appariement des trajectoires.....	52
2.4.1	Choix d'une paire de trajectoires .....	53
2.4.2	Calcul du pointage d'une paire de trajectoires .....	54
2.4.2.1	Construction de l'image de pointage .....	57
2.4.3	Recherche des paires de trajectoires participantes .....	62
2.4.4	Évaluation du nombre de bons appariements trouvé .....	62

2.5	Choix à chaque trame de la matrice de transformation.....	64
CHAPITRE 3 RÉSULTATS ET DISCUSSION.....		66
3.1	Méthodologie d'expérimentation .....	66
3.1.1	Montage utilisé.....	67
3.1.2	Création des éléments témoins (ground truth) .....	68
3.1.3	Explication des scénarios .....	72
3.1.4	Explication des graphiques obtenus .....	73
3.1.4.1	Graphiques de performance .....	73
3.1.4.2	Graphiques des gains de la nouvelle méthode .....	75
3.1.4.3	Graphiques des nombres de trajectoires.....	77
3.1.4.4	Paramètres utilisés pour les différents jeux de tests.....	77
3.1.5	Explication de la méthode de recalage de base.....	81
3.2	Variables testées et résultats.....	82
3.2.1	Séquence vidéo numéro 1 .....	83
3.2.1.1	Description du scénario.....	83
3.2.1.2	Présentation des résultats de la séquence .....	83
3.2.1.3	Analyse des résultats .....	85
3.2.2	Séquence vidéo numéro 2 .....	87
3.2.2.1	Description du scénario.....	88
3.2.2.2	Présentation des résultats de la séquence .....	88
3.2.2.3	Analyse des résultats de la séquence.....	90
3.2.3	Séquence vidéo numéro 3 .....	93
3.2.3.1	Description du scénario.....	93

3.2.3.2	Présentation des résultats de la séquence .....	93
3.2.3.3	Analyse des résultats de la séquence.....	95
3.2.4	Séquence vidéo numéro 4 .....	97
3.2.4.1	Description du scénario.....	98
3.2.4.2	Présentation des résultats de la séquence .....	98
3.2.4.3	Analyse des résultats de la séquence.....	100
3.2.5	Séquence vidéo numéro 5 .....	102
3.2.5.1	Description du scénario.....	102
3.2.5.2	Présentation des résultats de la séquence .....	102
3.2.5.3	Analyse des résultats de la séquence.....	104
3.2.6	Séquence vidéo numéro 6 .....	105
3.2.6.1	Description du scénario.....	106
3.2.6.2	Présentation des résultats de la séquence .....	106
3.2.6.3	Analyse des résultats de la séquence.....	108
3.2.7	Séquence vidéo numéro 7 .....	110
3.2.7.1	Description du scénario.....	110
3.2.7.2	Présentation des résultats de la séquence .....	110
3.2.7.3	Analyse des résultats de la séquence.....	112
3.2.8	Séquence vidéo numéro 8 .....	113
3.2.8.1	Description du scénario.....	113
3.2.8.2	Présentation des résultats de la séquence .....	114
3.2.8.3	Analyse des résultats de la séquence.....	116
3.2.9	Séquence vidéo numéro 9 .....	118

3.2.9.1	Description du scénario.....	118
3.2.9.2	Présentation des résultats de la séquence .....	118
3.2.9.3	Analyse des résultats de la séquence.....	120
3.3	Analyse des résultats globaux .....	123
	CONCLUSION .....	129
	RÉFÉRENCES.....	133

**LISTE DES TABLEAUX**

Tableau 3.1 Paramètres généraux utilisés pour les différents tests .....	78
Tableau 3.2 Paramètres de l'algorithme de soustraction d'arrière-plan utilisés pour les différents tests avec la caméra infrarouge .....	80
Tableau 3.3 Paramètres de l'algorithme de soustraction d'arrière-plan utilisés pour les différents tests avec la caméra couleur.....	81
Tableau 3.4 Tableau des données statistiques des performances des 2 méthodes sur la durée totale de chaque séquence vidéo .....	124
Tableau 3.5 Tableau des gains sur le pourcentage d'erreur de la nouvelle méthode sur la méthode de base .....	127

## LISTE DES FIGURES

Figure 1.1 Géométrie épipolaire de base d'un point dans l'espace (Laganière, 2009) ...	10
Figure 1.2 Relations coplanaires des épipoles, des projections et des points caractéristiques (Laganière, 2009) .....	11
Figure 2.1 Aperçu global de l'algorithme de recalage automatique .....	32
Figure 2.2 Schéma de transformation des données par les prétraitements .....	35
Figure 2.3 Images infrarouge (A) et couleur (B) avant la soustraction d'arrière-plan ...	36
Figure 2.4 Images infrarouge (A) et couleur (B) d'avant-plan de la méthode de la moyenne temporelle .....	37
Figure 2.5 Images infrarouge (A) et couleur (B) d'avant-plan après le filtre de taille de « blob ».....	38
Figure 2.6 Images infrarouge (A) et couleur (B) binaire d'avant-plan .....	38
Figure 2.7 Vue schématique en plongée du montage des caméras .....	43
Figure 2.8 Illustrations des définitions des termes paire de trajectoires et paire de points .....	46
Figure 2.9 Images infrarouge (A) et couleur (B) ainsi que leurs trajectoires .....	53
Figure 2.10 Images de pointage infrarouge (A) et couleur (B) ainsi que l'image des différences de recouvrement (C) .....	55
Figure 2.11 Images de pointage infrarouge (A) et couleur (B) ainsi que l'image des différences de recouvrement (C) pour une situation où le grossissement de l'image infrarouge pose problème.....	56
Figure 2.12 Schéma de l'algorithme d'acceptation de modification des images de pointage .....	61
Figure 3.1 Images en recouvrement à la trame 63 (A) et 88 (B) .....	69

Figure 3.2 Images des erreurs des éléments témoins en recouvrement lorsqu'ils sont bien (A) et mal répartis (B) .....	69
Figure 3.3 Images couleur (A) et infrarouge (B) des éléments témoins utilisés dans la majorité des scénarios .....	71
Figure 3.4 Images couleur (A) et infrarouge (B) des éléments témoins utilisés dans le scénario n'utilisant pas des cartons .....	71
Figure 3.5 Exemple de graphique de performance de la nouvelle méthode de recalage	74
Figure 3.6 Exemple de graphique de gains de la nouvelle méthode de recalage.....	76
Figure 3.7 Exemple de graphique de nombres de trajectoires .....	77
Figure 3.8 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 1 .....	84
Figure 3.9 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 1 .....	84
Figure 3.10 Graphique des gains de la nouvelle méthode pour la séquence numéro 1 ..	87
Figure 3.11 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 2.....	89
Figure 3.12 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 2 .....	89
Figure 3.13 Graphique des gains de la nouvelle méthode pour la séquence numéro 2 ..	91
Figure 3.14 Graphique du nombre de trajectoires par trame pour les spectres infrarouge et couleur sur la séquence vidéo numéro 2 .....	92
Figure 3.15 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 3.....	94
Figure 3.16 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 3 .....	94

Figure 3.17 Graphique des gains de la nouvelle méthode pour la séquence numéro 3 ..	96
Figure 3.18 Images fusionnées des caméras couleur et infrarouge montrant une inclusion des champs de vision de 23 % (A) et de 100 % (B).....	98
Figure 3.19 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 4.....	99
Figure 3.20 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 4 .....	99
Figure 3.21 Graphique des gains de la nouvelle méthode pour la séquence numéro 4	101
Figure 3.22 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 5.....	103
Figure 3.23 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 5 .....	103
Figure 3.24 Graphique des gains de la nouvelle méthode pour la séquence numéro 5	105
Figure 3.25 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 6.....	107
Figure 3.26 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 6 .....	107
Figure 3.27 Graphique des gains de la nouvelle méthode pour la séquence numéro 6	108
Figure 3.28 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 7.....	111
Figure 3.29 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 7 .....	111
Figure 3.30 Graphique des gains de la nouvelle méthode pour la séquence numéro 7	112
Figure 3.31 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 8.....	115

Figure 3.32 Graphique des performances en pourcentage d’erreur de la méthode de base sur la séquence vidéo numéro 8 .....	115
Figure 3.33 Graphique des gains de la nouvelle méthode pour la séquence numéro 8	116
Figure 3.34 Graphique du nombre de trajectoires par trame pour les spectres infrarouge et couleur sur la séquence vidéo numéro 8 .....	117
Figure 3.35 Graphique des performances en pourcentage d’erreur de la nouvelle méthode sur la séquence vidéo numéro 9.....	119
Figure 3.36 Graphique des performances en pourcentage d’erreur de la méthode de base sur la séquence vidéo numéro 9 .....	119
Figure 3.37 Graphique des gains de la nouvelle méthode pour la séquence numéro 9	121
Figure 3.38 Graphique du nombre de trajectoires par trame pour les spectres infrarouge et couleur sur la séquence vidéo numéro 9 .....	122

### **LISTE DES SIGLES ET ABRÉVIATIONS**

<b>blob</b>	Terme anglophone qui désigne groupe de pixels interconnectés
<b>RANSAC</b>	RANdom SAMple Consensus

## INTRODUCTION

Depuis son invention au début du siècle dernier, la caméra et les séquences vidéo qu'elle produit n'ont cessé de fasciner les gens de tous âges. Aujourd'hui, on retrouve des séquences vidéo partout. La télévision, le cinéma et récemment des sites de partages de séquences vidéo comme YouTube ne sont que quelques exemples d'applications utilisées par énormément de gens sur la planète tous les jours. Cette fascination pour la vidéo est compréhensible puisqu'elle repose sur notre sens le plus utilisé, la vue. Notre premier réflexe pour comprendre et analyser une situation est de l'observer. Il n'est donc pas surprenant de constater que la première chose qui nous vienne à l'esprit quand on parle de système de sécurité soit l'installation de caméras. Cette utilisation de caméras dans le domaine de la sécurité porte un nom, la télésurveillance.

La télésurveillance n'est cependant pas seulement utilisée que dans des situations où l'on tente d'empêcher un crime ou une intrusion dans un domicile. Elle peut aussi être utilisée avec des personnes vulnérables comme des personnes âgées. En effet, il est de plus en plus courant que des personnes âgées fassent affaire avec des compagnies de sécurité afin qu'elles installent des systèmes de télésurveillance pour les observer et réagir promptement si quelque chose devait leur arriver. Il est cependant très fastidieux pour des employés d'observer sans relâche des écrans qui diffusent des séquences vidéo. C'est pour cette raison que de plus en plus de recherches sont effectuées pour analyser automatiquement les séquences vidéo.

L'analyse automatique de séquences vidéo est cependant beaucoup plus facile à dire qu'à faire. En effet, malgré les milliers d'articles publiés et les centaines de logiciels développés, aucune technique ni aucun logiciel ne peuvent seulement approcher les performances d'analyse de l'être humain. Même des techniques de bases comme la soustraction d'arrière-plan automatique pour laquelle des centaines d'articles ont été publiés n'atteignent pas les niveaux de performance des techniques de soustraction d'arrière-plan manuelles. Par contre, même l'être humain a ses limites. En effet, nous ne voyons qu'un spectre de fréquences très limité. Le meilleur exemple de limitation de

notre vision est bien sûr notre incapacité à voir dans le noir. Plus généralement, l'utilisation d'images ou de séquences vidéo issues d'un seul spectre comporte toujours des limitations insurmontables amenées par la nature même du capteur utilisé. Dans notre exemple précédent de soustraction d'arrière-plan, des problèmes causés par l'ombrage sont très souvent rencontrés avec les caméras de spectre visible. Par contre, la même technique de soustraction d'arrière-plan utilisée avec des caméras infrarouge (qui capte le rayonnement thermique des objets) n'aurait pas ces problèmes d'ombrage puisque cette dernière n'existe tout simplement pas. Par contre, l'utilisation de caméras infrarouges a aussi ses problèmes qui lui sont propres. Citons ici l'exemple d'une personne en mouvement qui passe devant un écran d'ordinateur. Il est surprenant de constater que dans ces situations, la méthode de soustraction d'arrière-plan aura de la difficulté à différencier la personne de l'écran puisque les 2 sont approximativement à la même température. La solution semble donc résider dans la combinaison de plusieurs spectres différents puisque leurs forces et faiblesses sont souvent complémentaires. C'est en exploitant cette idée que de nombreux chercheurs tentent maintenant de rejoindre les performances de l'être humain.

Aujourd'hui, il existe de nombreuses sortes de caméras qui utilisent d'autres spectres différents que celui du visible pour produire des images. Dans ce genre de caméras, c'est sans doute la caméra infrarouge qui est la plus connue. Une caméra infrarouge produit des images à partir de la chaleur qui est émise par les corps qui se trouvent dans son champ de vision. Ce qui rend ce genre de caméra si intéressant est qu'il fonctionne aussi bien dans le noir qu'à la lumière puisque la chaleur n'est pas dépendante de la luminosité (sauf bien sûr dans le cas où le soleil plomberait intensément dans la pièce). Un autre facteur qui incite de nombreux chercheurs à se tourner vers les caméras infrarouges est leur prix qui est en chute libre depuis quelques années. Ces prix font en sorte que les solutions basées sur ce type de caméra sont aujourd'hui abordables. L'idée qui a vite germé dans la tête des chercheurs est simple; pourquoi ne pas combiner les avantages des caméras infrarouges avec ceux des caméras classiques couleur? Une solution possible pour y arriver est d'utiliser ce qu'on appelle le recalage multispectral.

Faire un recalage signifie trouver la matrice de transformation qui permet de passer d'un référentiel à un autre. Dans le cas d'un recalage multispectral dans le domaine de la vision par ordinateur, on fait plus particulièrement référence à trouver la matrice de transformation qui permet de transformer une image d'un spectre donné à l'image correspondante dans un autre spectre.

### **Problématique et travaux antérieurs**

La plus grande difficulté de ce recalage multispectral est de trouver des éléments pouvant être utilisés pour établir des correspondances entre des séquences vidéo tournées avec des caméras de spectres différents, plus précisément dans un cas où ces spectres sont l'infrarouge et la lumière visible. Dans la littérature, il existe principalement 2 types d'informations qui semblent pouvoir être utilisés dans ce cas de recalage précis : les « blobs » et les trajectoires des objets en mouvement. Dans un premier temps, Bhanu & Ju (2007) exposent une méthode qui effectue la soustraction d'arrière-plan afin d'obtenir des « blobs » et comparer le recouvrement de ces « blobs » pour évaluer la qualité du recalage. Le désavantage de cette méthode est qu'elle est très liée à la méthode de soustraction d'arrière-plan utilisée. Même si la soustraction d'arrière-plan est une technique très fréquemment utilisée en traitement vidéo, aucune des différentes méthodes qui permettent de l'effectuer n'est parfaite, car ces méthodes génèrent souvent beaucoup d'erreurs. C'est pour cette raison que Caspi et al. (2006) proposent d'utiliser les trajectoires des objets en mouvement pour effectuer le recalage. Le problème c'est que pour former ces trajectoires, il faut souvent aussi effectuer la soustraction d'arrière-plan. Cependant, comme cette méthode tient compte de toutes les images dans la séquence, elle est moins sensible aux erreurs qui surviennent souvent de façon ponctuelle. Par contre, cette méthode mesure la qualité d'un recalage en observant les distances des points qui composent la trajectoire ce qui est souvent moins précis que l'utilisation des images en recouvrement. L'essentiel de notre projet est de combiner ces deux approches pour tirer profit des avantages de chacune et pallier à leurs défauts.

### **Objectifs**

Le premier objectif qui doit être atteint est de développer une méthode de recalage pour des images visibles et infrarouge pour laquelle la matrice de transformation s'approche de la matrice réelle. Cette matrice réelle est définie ici comme celle obtenue lors d'un recalage manuel. Un deuxième grand objectif est d'évaluer les performances de la nouvelle méthode développée dans différentes situations. Dans ce cadre, voici les objectifs spécifiques :

- Développer une méthode de comparaison de qualité de recalage basée sur les images d'arrière-plan;
- Étant donné que ces images d'arrière-plan proviennent de deux spectres différents, il faut aussi que cette méthode tienne compte des erreurs dues aux différences entre les deux spectres. Il faut donc faire un choix judicieux des éléments d'arrière-plan qui seront utilisés pour que ceux-ci soient visibles dans les deux spectres;
- Abaisser le pourcentage d'erreur de recouvrement de la méthode basée uniquement sur les trajectoires dans les situations où le nombre de trajectoires est relativement petit.

## **Contributions**

On peut distinguer deux grandes contributions à ce travail. Dans un premier temps, la combinaison de l'utilisation des trajectoires et des images en recouvrement dans la même méthode de recalage couleur-infrarouge est, autant que nous sachions, une nouveauté. De plus, il sera démontré que cette combinaison de deux méthodes permet de compenser les défauts de la méthode basée uniquement sur les trajectoires et ainsi d'améliorer les performances.

Dans un deuxième temps, la technique utilisée pour évaluer la qualité du recalage semble aussi novatrice. La technique utilisée repose sur un pourcentage d'erreurs de recouvrement obtenu à partir d'un recalage effectué manuellement. Cette mesure donne une meilleure idée des performances de la méthode automatique par rapport à celui

effectué par une personne. Aucun des articles rencontrés ne donne une mesure chiffrée et objective du recalage. Généralement, cette qualité de recalage est évaluée subjectivement par un expert qui classe les différents recalages obtenus comme réussis ou non. On rencontre aussi parfois des articles dans lesquels la qualité est évaluée en donnant le pourcentage d'erreur, mais en comparant celui-ci avec l'image originale et non pas avec un recalage manuel.

### Hypothèses de départ

Le recalage multispectral étant une tâche très complexe, il est très ardu de concevoir une méthode qui fonctionne parfaitement dans toutes les situations possibles. C'est pourquoi certaines simplifications et hypothèses de départ ont été faites. Elles sont décrites dans cette section.

Tout d'abord, il faut être conscient qu'il est plus coûteux d'utiliser plusieurs caméras plutôt qu'une seule. C'est pourquoi nous nous sommes limités à utiliser deux caméras seulement. Cette limitation à deux caméras représente la première condition à respecter pour utiliser la méthode de recalage automatique présentée dans ce travail.

Une autre hypothèse de départ importante est que le système n'a pas été calibré. Il s'agit d'une hypothèse très importante puisque si le système était calibré, le recalage serait presque trivial. Même s'il pourrait être souligné qu'un système calibré pourrait mieux convenir à une application en vidéo surveillance, l'hypothèse de non calibration demeure très utile dans plusieurs situations rencontrées dans les systèmes utilisés dans le domaine de la sécurité. S'il est vrai de dire que la calibration est bonne lorsque les caméras ne bougent pas, les choses se compliquent dans des environnements moins contrôlés. On peut penser aux vibrations causées par des véhicules sur une caméra qui filme une autoroute ou au vent qui déplacerait la même caméra comme source possible d'instabilité de la caméra. Dans ces situations, la calibration ne fonctionnerait qu'à court terme. Les systèmes non calibrés sont donc plus intéressants dans des situations où le système doit fonctionner longtemps ou dans des conditions moins idéales. De plus, il peut être très fastidieux de calibrer chaque caméra une à une dans un système en

comportant des centaines, voir des milliers dans certains gratte-ciels. De plus, le recalage basé sur un système non calibré est souvent utilisé comme étape préliminaire ou complémentaire à d'autres techniques plus avancées comme la reconstruction trois dimensions d'une scène où le suivi d'objet.

Le premier problème qui sera rencontré lorsque l'on tente de combiner deux séquences vidéo est d'abord la synchronisation de ces deux séquences. Pour nos expérimentations, nous avons utilisé un petit logiciel qui s'occupait de synchroniser les deux caméras lors du tournage des séquences tests. Malgré tout, la synchronisation en question n'étant pas parfait, car logicielle, il est possible que certaines erreurs de recalage soient expliqués en partie par ces défauts de synchronisation. Néanmoins, la synchronisation des deux séquences vidéo représente une hypothèse de base nécessaire au fonctionnement de la méthode de recalage présenté ici. Cette hypothèse est raisonnable car il existe des modules matériels pour synchroniser les caméras. Ce type de module est d'ailleurs utilisé dans le domaine de la vidéo surveillance. Par contre, comme nous ne disposons pas d'un tel équipement, nous avons été forcés de nous rabattre sur une synchronisation logicielle.

Une autre hypothèse de départ est reliée aux choix du type de matrice de transformation utilisé. En effet, il sera question un peu plus loin des différents types de matrice de transformation ainsi que des difficultés à surmonter pour trouver les paramètres de chacun. Pour simplifier quelque peu la tâche, nous nous sommes inspirés de (Caspi et al., 2006) et avons posé comme hypothèse que tous les objets d'avant-plan filmés par les caméras doivent être situés à peu près tous dans le même plan. Il s'agit d'ailleurs d'une hypothèse souvent utilisée en vidéo surveillance puisque les caméras sont souvent en plongée. Il sera question des impacts de cette hypothèse sur le choix de la matrice de transformation un peu plus loin. Pour respecter cette restriction, les séquences vidéo utilisées ont été tournées en plongée.

Le type de lentille utilisé peut aussi amener son lot de problème. Par contre, ces problèmes sont généralement bien connus et les solutions abondent dans la littérature.

C'est pour cette raison que nous avons décidé de faire l'hypothèse que les images obtenues ne sont pas victimes de distorsions géométrique causées par les lentilles. Ainsi donc, seul des lentilles ayant très peu de distorsion ont été utilisés lors du tournage des séquences vidéo testées.

Les positions relatives des caméras représentent aussi un obstacle majeur aux méthodes de recalage automatique. Par conséquent, l'hypothèse de l'utilisation de la matrice homographique pour décrire la transformation entre les deux caméras doit être respectée pour que la méthode présentée dans ce travail soit utilisable. Les impacts sur le montage qui doit être utilisé pour respecter cette hypothèse seront présentés ultérieurement. Il est à noter que malgré ces restrictions, les positions relatives des caméras utilisées pour tester la méthode demeurent valides dans le contexte de la vidéo surveillance puisque ces contraintes sont, somme toute, assez faibles.

Enfin, il faut qu'une certaine portion de la scène soit visible par les deux caméras puisque, dans le cas contraire, il n'y aurait pas d'information commune. L'impact de la portion en inclusion d'une des images dans l'autre sera évalué dans les tests de performances de la méthode.

### **Plan du mémoire**

Ce document est composé de trois chapitres, d'une introduction et d'une conclusion. Le premier fait une tournée des différentes techniques de recalage multispectral qui sont offertes dans la littérature scientifique. Comme ces techniques sont nombreuses, il sera surtout question des catégories dans lesquelles ces techniques peuvent être regroupées. Il sera aussi question des différentes matrices de transformation qui peuvent être utilisées pour effectuer ce genre de recalage.

Dans le deuxième chapitre, c'est la méthode de recalage infrarouge-visible qui a été développée dans ce travail qui sera exposée. Tous les détails des prétraitements qui doivent être appliqués à l'image seront d'abord présentés. Ces prétraitements tournent tous autour du concept d'extraction des trajectoires et des images d'avant-plan qui sont à la base de la nouvelle technique de recalage. Ensuite, l'algorithme de recherche des

appariements des trajectoires sera présenté. Il sera aussi question des raisons qui ont motivé le choix de la matrice de transformation affine qui est utilisée ainsi que de l'algorithme utilisé pour calculer cette matrice à partir des paires de trajectoires trouvées.

Pour le troisième chapitre, il traite de tous les tests qui ont été effectués afin de mesurer les performances du nouvel algorithme développé. La méthodologie suivie pour effectuer ces tests est aussi exposée dans ce chapitre. Ces tests consistent à exécuter deux méthodes de recalage sur 9 séquences vidéo tournées avec des caméras infrarouge et visible. La première méthode exécutée est notre propre implantation de la méthode exposée dans Caspi et al. (2006). La deuxième méthode est la méthode développée dans le cadre de ce projet de recherche. Dans ce même chapitre, une discussion sur les causes et les indices qui tendent à montrer une supériorité de la nouvelle méthode dans ces séquences vidéo sera aussi présentée. Pour finir, une conclusion résumant le travail effectué et les résultats importants obtenus sera incluse à la fin de ce document. Il y sera aussi question de l'atteinte des objectifs fixés précédemment ainsi que de travaux futurs pouvant être menés à partir de la nouvelle méthode de recalage multispectral développée. Mais dans un premier temps, dressons un portrait des différentes techniques de recalage multispectral dans la littérature.

## CHAPITRE 1 REVUE DE LA LITTÉRATURE

Dans ce chapitre, il sera question des différentes méthodes développées dans la littérature pour effectuer le recalage entre des images infrarouge et visible. Mais avant de parler de recalage, il sera question des différentes matrices de transformation à deux dimensions qui sont à la base du recalage. Ensuite, il sera question des différentes méthodes de recalage en commençant par les différentes techniques normalement utilisées pour effectuer un recalage entre deux images de même spectre. Deux grands types de méthodes de recalage seront exposés soit le recalage basé sur les pixels et le recalage basé sur les caractéristiques. Des méthodes de recalage pouvant être classées dans ces types seront décrites dans les sections 1.2.1 et 1.2.2. Pour sa part, la section 1.3 traitera de différentes méthodes existantes permettant d'effectuer un recalage multispectral infrarouge-visible en utilisant seulement l'information des pixels eux-mêmes. La section 1.4 portera elle aussi sur le recalage infrarouge-visible mais en exposant les méthodes qui utilisent plutôt les éléments caractéristiques pouvant être extraits du contenu des images. Cette section a été découpée en deux sous-sections qui illustrent les deux principales caractéristiques utilisées par ces méthodes : les contours (section 1.4.1) et les points caractéristiques (section 1.4.2).

### 1.1 Matrices de transformation

Effectuer un recalage revient à trouver la matrice de transformation homographiques entre deux images. Mais avant de s'attarder aux types de matrices de transformation, il semble important de faire un rappel sur la géométrie épipolaire qui permet de mieux comprendre la provenance des matrices de transformation homographiques et les conditions qui doivent être respectés pour leur utilisation. Ensuite, les matrices de transformation à 2 dimensions dites homographiques seront exposées. Il existe 4 types de ce genre de matrices (Hartley & Zisserman, 2002): les matrices de transformation isométrique, les matrices de transformation de similarité, les matrices de transformation affine et les matrices de transformation de projection. Chacun de ces types de transformation est expliqué dans les sections qui suivent.

### 1.1.1 Géométrie épipolaire

Dans le domaine de la vision par ordinateur, on modélise souvent les caméras par le modèle du sténopé (aussi appelé tête d'épingle). De cette modélisation de caméra s'est développée la géométrie épipolaire. La géométrie épipolaire désigne les théories mathématiques mettant en relations la position d'un point dans un espace en trois dimensions et les coordonnées en deux dimensions de ce même point dans plusieurs plans différents. Ces plans représentent dans les théories de vision par ordinateur les différentes caméras utilisées. Pour mieux comprendre, prenons le cas d'utilisation de seulement deux caméras. La géométrie épipolaire dit que le centre optique de la caméra de droite projetée dans l'image de gauche est l'épipole dans l'image de gauche. De même, le centre optique de la caméra de gauche dans l'image de droite forme l'épipole de l'image de droite. La figure suivante permet de mieux comprendre :

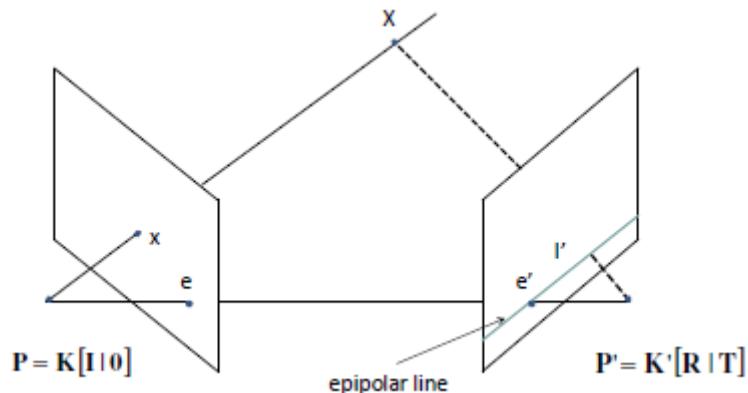


Figure 1.1 Géométrie épipolaire de base d'un point dans l'espace  
(Laganière, 2009)

Dans la figure précédente, on peut voir que le point dans l'espace  $X$  est projeté sur les deux plans en  $x$  pour l'image de gauche et quelque part sur la droite épipolaire nommée  $l'$  dans l'image de droite. On définit une droite épipolaire comme toute droite passant à la fois par l'épipole et par la projection sur le plan considéré d'un point caractéristique. Notons aussi dans la figure précédente les épipoles des deux plans images notées  $e$  et  $e'$ . Sachant ces définitions, il est aisément de constater qu'une relation mathématique existe entre la position d'un point dans l'espace, les projections de ce point sur les plans

images des caméras ainsi qu'avec les épipoles de ces mêmes plans. Tous ces points sont coplanaires. La figure qui suit nous aide à mieux le visualiser :

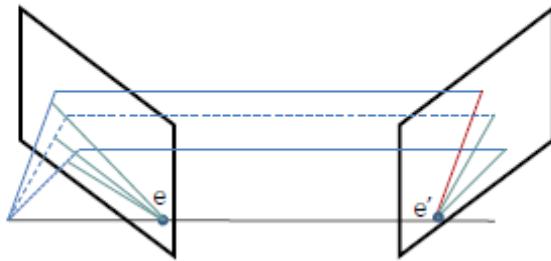


Figure 1.2 Relations coplanaires des épipoles, des projections et des points caractéristiques (Laganière, 2009)

Grâce à ces observations, il est possible de définir l'équation suivante qui met en relation les coordonnées d'un point projeté dans une des images et l'équation de la droite épipolaire correspondante dans l'autre image.

$$l' = Fx \quad (1.1)$$

La matrice notée  $F$  dans cette équation est appelée matrice fondamentale. Les symboles  $l'$  et  $x$  désignent respectivement la droite épipolaire correspondante et les coordonnées du point  $X$  projeté sur un plan image. Il faut cependant noter que l'équation précédente permet de relier un point à une droite et pas un point à un autre point directement. En effet il pourrait y avoir une ambiguïté si la projection de plusieurs points caractéristiques se retrouvent sur la même droite épipolaire mais pas nécessairement à la même profondeur dans l'espace. Pour simplifier le problème, une solution serait d'imposer une restriction comme quoi tous les points de la scène doivent être coplanaires. Dans cette situation, toutes les droites épipolaires deviennent uniques et il devient possible de décrire la transformation des projections de tous les points caractéristiques sur un plan donné en leur correspondant sur l'autre plan par une seule matrice. Cette matrice se nomme la matrice de transformation homographique. Voici l'équation qui décrit le lien entre les points projetés et la matrice homographique.

$$x' = Hx \quad (1.2)$$

Dans cette équation,  $x$  et  $x'$  désignent les deux projections du point  $X$  (en coordonnées homogènes) de la scène sur les deux plans image et  $H$  la matrice homographiques. Il existe plusieurs types de matrices homographiques qui seront présentées dans les prochaines sections.

### 1.1.2 Les matrices de transformation homographique isométrique

Ce type de matrice de transformation est utile dans les cas où les déformations qui sont observées préservent les distances euclidiennes. En d'autres termes, si les déformations observées sont des combinaisons de translations, rotations et réflexions. Voici la matrice qui résume ce type de transformation :

$$H_{\text{isométrique}} = \begin{bmatrix} e \cos \theta & -\sin \theta & t_x \\ e \sin \theta & \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.3)$$

Dans l'équation précédente, les 4 degrés de liberté sont bien visibles. Tout d'abord, la variable  $e$  indique le premier degré de liberté et elle est associée à la réflexion. En effet,  $e$  ne peut prendre que 2 valeurs; 1 et -1 pour indiquer le sens de la réflexion. L'autre variable, qui représente le deuxième degré de liberté est associée à la rotation et elle est notée  $\theta$  dans l'équation précédente. Les deux derniers degrés de liberté sont associés aux variables qui représentent respectivement les translations en X et Y soit  $t_x$  et  $t_y$ . Ce type de transformation est le plus simple et il est hautement restrictif quant à la position de la caméra. C'est pourquoi les prochaines transformations sont présentées pour les cas plus complexes.

### 1.1.3 Les matrices de transformation homographique de similarité

Ce type de transformation est grandement relié à la matrice de transformation isométrique. En fait, il s'agit du même type de transformation auquel on ajoute une homothétie. Voici l'équation qui résume ce type de transformation :

$$H_{\text{similarité}} = \begin{bmatrix} s \cos \theta & -s \sin \theta & t_x \\ s \sin \theta & s \cos \theta & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.4)$$

Dans la précédente équation, on retrouve encore une fois  $\theta$ ,  $t_x$  et  $t_y$  pour la rotation et les translations selon X et Y. À ces 3 transformations, on ajoute une homothétie notée  $s$ . Une homothétie inclut les différents facteurs de grossissement qui peuvent être appliqués. La matrice de transformation de similarité peut être utilisée dans les cas où le recouvrement des champs de vision des caméras est inclus complètement l'un dans l'autre. En effet, une zone réduite d'inclusion des champs de vision fait apparaître des déformations différentes de celles supportées par la matrice de similarité dues aux lentilles. Même si les lentilles utilisées dans nos expérimentations déforment très peu les images (hypothèses de départ dans l'introduction), il y a une augmentation du phénomène sur les bords des images ce qui fait qu'une zone d'inclusion des champs de vision réduite se retrouve presque toujours dans cette zone de déformation accrue. Le résultat est que la matrice de similarité devient pratiquement inutilisable pour effectuer le recalage dans ces cas. Ceci limite le genre de montage que l'on peut utiliser. C'est l'une des raisons qui nous ont fait choisir le prochain type de matrice dans notre projet.

#### 1.1.4 Les matrices de transformation homographique affine

Dans les cas où l'on ne veut pas trop limiter les positions des caméras, il est souvent plus avisé d'utiliser la matrice de transformation affine. Cette dernière est définie par l'équation suivante :

$$H_{\text{affine}} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (1.5)$$

Dans cette équation, on retrouve encore une fois les translations selon les axes X et Y notées par les variables  $t_x$  et  $t_y$ . Cependant, aucune hypothèse n'est faite sur la nature des autres transformations. C'est pourquoi on retrouve les 4 autres degrés de liberté qui sont

simplement ( $a_{11}$ ,  $a_{12}$ ,  $a_{21}$  et  $a_{22}$ ), des nombres à calculer. Il sera abondamment question de cette matrice et de comment l'utiliser dans le prochain chapitre. C'est pourquoi il n'en sera pas question davantage ici.

### 1.1.5 Les matrices de transformation homographique de projection planaire

Il s'agit de la plus générale des transformations à 2 dimensions. Elle peut expliquer n'importe quel type de transformations qui pourraient notamment être dues à certaines lentilles grand-angles qui déforment l'image. Voici l'équation qui définit ce type de matrice de transformation :

$$H_{\text{projection}} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (1.6)$$

Chacune des variables de cette matrice est un degré de liberté. Il s'agit donc d'une matrice à 9 degrés de liberté. Il est très difficile de retrouver cette matrice exacte car on se retrouve souvent dans des situations où plusieurs matrices peuvent expliquer les déformations observées. Il s'agit malgré tout de la matrice qui explique le plus fidèlement un recalage. Néanmoins, sa complexité accrue a fait en sorte qu'elle n'a pas été retenue.

Maintenant que les différents types de matrice de transformation ont été exposés, attardons-nous maintenant aux techniques qui permettent de retrouver ces matrices.

## 1.2 Méthodes classiques de recalage monospectral

Quelques auteurs se sont donnés comme tâche de regrouper l'ensemble des techniques de recalage proposées dans la littérature scientifique. Un des plus vieux et des plus cités est l'article de Brown (1992) qui a servi de base à un article plus récent (Zitova & Flusser, 2003) qui ajoute de nouvelles méthodes. Cet article commence par introduire les grandes étapes qui composent la grande majorité des techniques de recalage :

- La détection des caractéristiques sert à identifier manuellement ou automatiquement des éléments dans les images qui seront utilisés pour effectuer le recalage;
- L'appariement des caractéristiques trouve les correspondances entre les caractéristiques des images en utilisant différentes techniques de mesure de ressemblance;
- L'estimation du modèle de transformation concerne le choix du type et des paramètres de la modélisation de la transformation. Un bon exemple est le choix de la matrice de transformation affine qui sera expliqué dans le chapitre 2 de ce travail;
- Le ré-échantillonnage de l'image et la transformation représente l'étape où l'image est modifiée par la matrice de transformation trouvée. Le ré-échantillonnage fait référence aux possibles méthodes d'interpolation dans les cas de valeurs non entières.

Ensuite, deux grandes familles de méthodes sont identifiées : les méthodes basées sur les pixels et les méthodes basées sur les caractéristiques. L'article enchaîne alors avec différents chapitres basés sur les différentes étapes du recalage. Comme le projet présenté dans ce texte s'attarde essentiellement sur l'appariement des caractéristiques, ce sont ces méthodes exposées dans Zitova & Flusser (2003) qui seront résumées dans cette section.

### **1.2.1 Les méthodes basées sur les pixels**

#### Méthodes de corrélation (Zitova & Flusser, 2003)

Une première famille de méthodes souvent employées appartenant à la catégorie des méthodes basées pixels est la famille des méthodes dites de corrélation. Pour commencer, des fenêtres de recherche sont définies dans les deux images. Ensuite, on déplace ces fenêtres et l'on calcule chaque fois une mesure de similarité entre les deux. Cette mesure de similarité est basée sur une formule de corrélation croisée qui prend en

paramètre l'intensité des pixels situés dans les fenêtres de recherche des deux images. La paire de fenêtres avec la plus grande mesure de corrélation croisée donne la meilleure estimation du recalage. Les coordonnées des fenêtres de recherche sont alors utilisées pour calculer la matrice de transformation entre les deux. Le plus grand désavantage de ces méthodes est la grande quantité de calcul qu'elles nécessitent puisqu'il faut calculer une mesure de corrélation croisée pour chaque possibilité de paires de fenêtres de recherche. La quantité de calcul peut aussi être grandement augmentée si les fenêtres de recherches sont trop petites. Aussi, ces méthodes ne peuvent pas être directement utilisées dans les cas du recalage multispectral puisque les couleurs entre les images ne sont pas les mêmes.

#### Méthodes de Fourier (Zitova & Flusser, 2003)

Si les images sont très bruitées ou que la quantité de calcul est trop grande, les méthodes de Fourier sont souvent utilisées. Rappelons d'abord que la transformation de Fourier permet de passer du domaine temporel au domaine fréquentiel. Il est ainsi plus facile d'appliquer des filtres pour éliminer certaines formes de bruit dans les images. Le recalage s'effectue en tentant d'apparier les différents pics de fréquence entre les images. Les pics de fréquences font référence aux fréquences avec la plus grande amplitude dans les images de Fourier. La disposition spatiale de ces pics de fréquence change avec l'orientation de l'image. Donc théoriquement, la fréquence avec la plus grande amplitude dans l'image infrarouge correspond à la fréquence avec la plus grande amplitude dans l'image couleur. En calculant la distance entre les deux, on peut retrouver la matrice de transformation. Par contre en pratique, l'hypothèse que le plus grand pic de fréquence dans l'image infrarouge correspond au plus grand pic de fréquence en couleur n'est pas toujours respectée ce qui fait que la méthode n'est pas parfaite. Ces méthodes sont aussi souvent nommées corrélation de phases et peuvent être utilisées dans le cas d'un recalage multispectral.

#### Méthodes d'information mutuelle

Cette dernière catégorie de méthodes s'adapte très bien au recalage multispectral. Elle ne sera pas décrite en détail ici, mais il en sera abondamment question à la section 0.

### **1.2.2 Les méthodes basées sur les caractéristiques**

#### Méthodes utilisant les relations spatiales (Zitova & Flusser, 2003)

Cette première catégorie de méthode repose sur un principe très simple : après une bonne transformation, les caractéristiques observées devraient être près de celles calculées. En d'autres mots, est-ce que les caractéristiques de l'image de comparaison se retrouvent au même endroit que dans l'image qui a subi la transformation? Cette catégorie regroupe donc toutes les méthodes qui utilisent des mesures de distances entre des caractéristiques identifiables dans l'image (points, lignes, contours, etc.) pour trouver la matrice de transformation entre les deux images. Elle regroupe aussi les méthodes dites de regroupement (clustering). Les méthodes diffèrent les unes des autres par la mesure de distance utilisée. Ce type de méthode peut très bien être utilisé dans le cas du recalage multispectral.

#### Méthode des descripteurs invariants (Zitova & Flusser, 2003)

Ces méthodes se distinguent des autres par la façon dont elles calculent l'appariement des caractéristiques. Les autres types de méthode essaient souvent d'utiliser des relations spatiales entre les éléments caractéristiques des images pour trouver quelles sont les paires bien appariées. Par exemple, beaucoup de méthodes vont calculer la distance entre 2 points et tenter de la minimiser. Les méthodes de descripteurs invariants tentent plutôt de trouver cet appariement en observant la description des caractéristiques. Par exemple, si les caractéristiques utilisées sont les contours, on pourrait tenter de trouver une paire de contours correspondants en comparant les degrés de courbure de chaque contour. Le degré de courbure serait alors nommé descripteur de la caractéristique qui, dans cet exemple, serait les contours. Idéalement, un descripteur devrait être invariant aux déformations possibles entre les deux images à recaler. Ces descripteurs de caractéristiques doivent respecter des critères comme : l'invariance, l'unicité, la stabilité et l'indépendance mathématique pour être très efficace à apparier les caractéristiques. Il

est très rare que tous ces critères soient respectés. Les méthodes varient donc dans la manière d'évaluer le degré de respect de ces critères. Ce genre de méthode peut être utilisé pour le recalage infrarouge-visible.

#### Méthodes de relaxation (Zitova & Flusser, 2003)

Ce type de méthodes consiste à donner une étiquette à chacune des caractéristiques observées dans les deux images et à calculer les différences dans l'appariement de ces étiquettes. La méthode itère alors tant et aussi longtemps qu'il y a des différences dans l'appariement de ces étiquettes. Prenons par exemple une méthode fictive qui utilise les contours comme éléments caractéristiques dans les images. La première étape d'une méthode de relaxation serait de numérotter chacun de ces contours. Ces numéros sont les étiquettes. Ensuite, on choisit au hasard 2 contours; un dans l'image couleur et l'autre dans l'image infrarouge. On calcule ensuite la matrice de transformation résultante de ce premier appariement. Avec cette matrice, les autres paires de contours sont formées selon une mesure de distance. Par exemple, deux contours pourraient être appariés en choisissant celui le plus près de la position estimée à l'aide de la matrice de transformation. Avec ces nouvelles paires de contours, on tente d'affiner la matrice de transformation en la calculant à partir de toutes ces paires formées. Avec cette nouvelle matrice, on recalcule les paires de contours et l'on observe si ces paires ont changé, soit si elles sont formées des mêmes étiquettes. L'algorithme continue ainsi jusqu'au moment où les paires ne changent plus. Il est possible de trouver certaines méthodes qui s'inspirent de ce concept pour les situations de recalage infrarouge-visible.

### **1.3 Méthodes de recalage infrarouge-visible basées sur les pixels**

En raison de la différence des sources d'informations entre les caméras infrarouges (chaleur émise) et les caméras visibles (lumière réfléchie), les méthodes basées sur les pixels utilisées pour des recalages classiques ne peuvent généralement pas être utilisées. Cependant, les théorèmes statistiques de l'information mutuelle et de l'information mutuelle normalisée ont récemment été utilisés avec un certain succès lors de recalages infrarouge-visible par Chen, Varshney, & Slamani (2003), Kyoung Soo, Jae Hak, &

Jong Beom (2005), Stone & Wolfov (2002) et Krotosky & Trivedi (2007). Intuitivement, la théorie d'informations mutuelles repose sur l'idée d'évaluer l'impact de la connaissance d'une variable aléatoire sur une autre variable aléatoire. En d'autres mots, en assumant que des variables aléatoires (par exemple les valeurs des pixels de deux images) sont dépendantes, alors il devrait exister un moyen statistique de quantifier cette dépendance. Mathématiquement, l'information mutuelle  $I(X;Y)$  est définie par l'équation suivante (Cover & Thomas, 2006) :

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)p(y)} \quad (1.7)$$

Dans cette équation,  $p(x,y)$  représente la probabilité d'intersection des variables aléatoires discrètes  $X$  et  $Y$  et  $p(xy \text{ et } p(y))$ , les fonctions de probabilité de masse de ces variables. Pour mieux comprendre le lien entre cette formule et la définition intuitive de la théorie de l'information mutuelle, l'encyclopédie en ligne Wikipedia (2008) propose un exemple simple. Si  $X$  et  $Y$  sont indépendants, leur probabilité d'intersection  $p(x,y)$  est égale au produit des probabilités de chacune des variables ( $p(x)$  et  $p(y)$ ). Donc, l'information mutuelle donnée par deux variables indépendantes peut être calculée à l'aide de l'équation 1.7 :

$$\text{Si } X \text{ et } Y \text{ sont indépendantes : } p(x,y) = p(x)p(y) \quad (1.8)$$

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x)p(y)}{p(x)p(y)} = \sum_{x,y} p(x,y) \log(1) = 0 \quad (1.9)$$

Cette démonstration est logique puisque si les variables sont indépendantes, aucune des deux ne donne d'information sur l'autre. Traduit en des termes plus proches de notre problématique du recalage d'images, les variables aléatoires  $X$  et  $Y$  seraient les pixels des images couleur et infrarouge. Comme ces deux images (variables aléatoires) décrivent la même scène, elles sont dépendantes. Par conséquent, la transformation qui maximise cette dépendance devrait être celle permettant d'effectuer le recalage. Cependant, l'équation 1.7 n'est pas suffisante pour décrire le comportement des données reliées à

l'information. Pour améliorer le modèle, les théories d'informations mutuelles reposent sur une mesure d'entropie appelée l'entropie marginale de Shannon ( $H$ ) définie par l'équation suivante :

$$H(X) = -\sum_x p(x) \log p(x) \quad (1.10)$$

Ici,  $p(x)$  représente la fonction de masse de la variable aléatoire discrète  $X$ . La même formule pourrait être écrite avec une autre variable aléatoire nommée  $Y$ . Aussi, il est possible de définir l'entropie jointe de deux variables aléatoires discrètes  $(X,Y)$  par l'équation suivante :

$$H(X,Y) = -\sum_{x,y} p(x,y) \log p(x,y) \quad (1.11)$$

Sachant ces informations et quelques propriétés statistiques, il est possible de déduire une nouvelle définition de l'information mutuelle (IM) en fonction de l'entropie. La preuve mathématique est faite au chapitre 2 de Cover & Thomas (2006). L'équation résultante est la suivante :

$$I(X;Y) = H(X) + H(Y) - H(X,Y) \quad (1.12)$$

Dans le cas d'images, pour utiliser ce résultat, les deux travaux utilisent des histogrammes combinés de l'intensité des pixels en ton de gris des deux images. Chaque valeur qui compose un histogramme combiné est obtenue en additionnant le nombre de pixels qui ont une intensité A dans une image et le nombre de pixels qui ont une intensité B dans l'autre image. L'ensemble de ces valeurs additionnées pour chaque valeur possible de A et B nous donne l'histogramme combiné. En normalisant ces histogrammes, il est possible de calculer les différentes fonctions de masse nécessaire au calcul de l'information mutuelle :

$$P(a,b) = \frac{h(a,b)}{\sum_{a,b} h(a,b)} \quad (1.13)$$

$$P(a) = \sum_b P(a,b) \quad (1.14)$$

$$P(b) = \sum_a P(a,b) \quad (1.15)$$

Dans les équations précédentes,  $h(a,b)$  représente l'histogramme combiné des images  $a$  et  $b$ ,  $P(a,b)$  la fonction de masse jointe des images  $a$  et  $b$  et  $P(a)$  ou  $P(b)$  la fonction de masse de l'image  $a$  ou  $b$ . Un histogramme combiné,  $h(a,b)$ , est un histogramme construit à partir de l'information de deux variables. Illustrons par un exemple. Dans notre cas, ces deux variables sont l'intensité des pixels de l'image infrarouge,  $a$ , et l'intensité des pixels de l'image couleur,  $b$ . Ces deux données servent de coordonnées à l'histogramme. Pour connaître la valeur à attribuer à chaque coordonnée, on compte le nombre de pixels ayant l'intensité  $a$  dans l'image infrarouge et on l'additionne au nombre de pixels ayant l'intensité  $b$  dans l'image couleur. À partir de cet histogramme, il est possible de calculer toutes les autres valeurs et ainsi de calculer la valeur de l'information mutuelle qui sera maximisée afin de trouver la meilleure matrice de transformation.

L'information mutuelle a cependant quelques défauts. En effet, elle est dépendante de la taille du recouvrement entre les images. En effet, comme elle n'est composée que d'une addition et d'une soustraction (éq. 1.12), et que les fonctions de masse sont des sommes, si le nombre de pixels augmente, l'information mutuelle sera forcément plus grande. C'est pourquoi Kyoung Soo et al. (2005) propose d'utiliser une mesure relative plutôt qu'une mesure absolue c'est-à-dire une mesure contenant une division. Cette mesure est un dérivé de l'information mutuelle appelé information mutuelle normalisée  $I_N(X;Y)$  et est décrit par l'équation suivante :

$$I_N(X;Y) = \frac{H(X) + H(Y)}{H(X,Y)} \quad (1.16)$$

Dans cette dernière équation, les symboles reliés aux différentes mesures d'entropie ( $H(X)$ ,  $H(Y)$  et  $H(X,Y)$ ) ont la même signification.

Les théories d'informations mutuelles s'intègrent au recalage en étant utilisées comme critère de qualité du recalage. Plus le recalage est bon, plus l'information mutuelle sera grande. Rappelons ici que les images couleur et infrarouge représentent des variables dépendantes et que plus la dépendance est grande, plus l'information mutuelle sera grande. Cependant, une des hypothèses de base de l'information mutuelle est la corrélation des informations partout dans l'image ce qui n'est pas nécessairement le cas dans un recalage infrarouge-visible. Pour éviter de possibles erreurs causées par la violation de cette hypothèse de base, les deux articles présentés ici effectuent leur recalage sur des portions d'images ayant des éléments potentiellement dépendants dans les deux images : des rectangles englobants pour Chen et al. (2003) et une sorte de simplification basée sur les contours pour Kyoung Soo et al. (2005). Malgré tout, un doute subsistait sur la convergence des méthodes basées sur l'information mutuelle dans le cas des recalages visible-infrarouge. C'est pour cette raison que Kern & Pattichis (2007) ont publié un article principalement mathématique qui prouve cette convergence dans cette utilisation précise.

Cependant, un article un peu plus vieux présente une alternative en réussissant à effectuer un recalage basé sur les pixels en utilisant toute l'image et pas seulement des parties. En effet, Stone & Wolfov (2002) propose d'effectuer une opération de prétraitement sur les images pour augmenter la corrélation entre les deux. Il s'agit essentiellement d'un filtre appliqué sur les images qui fait ressortir les arêtes et qui transforme les images en noir et blanc. Ensuite, il propose d'utiliser une méthode basée sur la corrélation. Pour accélérer les calculs, il utilise des équations qui peuvent être résolues avec des transformations de Fourier. Malheureusement, on y mentionne que la qualité du recalage n'est pas nécessairement améliorée quand on la compare avec les méthodes d'informations mutuelles plus classiques et que les images doivent quand même partager une grande ressemblance pour que la méthode fonctionne. Cependant, cette méthode démontre quand même une plus grande rapidité au niveau des calculs.

Une autre approche utilisant l'information mutuelle et démontrant les difficultés d'utiliser celle-ci directement est Krotosky & Trivedi (2007). Dans cet article, il est de

nouveau mention de l'inefficacité des méthodes d'informations mutuelles, lorsqu'utilisées sur des images complètes. C'est pourquoi ils utilisent plutôt les images d'avant-plan obtenues grâce à des techniques de segmentation. On y mentionne aussi que les séquences vidéo rapprochées ne peuvent pas utiliser directement le recalage par une matrice de transformation classique. En effet, des objets se situeront plus près de la caméra que d'autres ce qui complique le calcul. En fait, il faut que la scène soit planaire (que les objets dans l'avant-plan se situent à la même distance de la caméra) pour qu'une seule matrice de transformation puisse expliquer la transformation pour l'ensemble de la scène. Cette observation n'est pas en lien avec la théorie de l'information mutuelle mais est incluse dans cette section puisque cet article est expliqué ici. Pour pallier ce problème, ils calibrent les deux caméras en utilisant la méthode classique de l'échiquier. Ils connaissent ainsi les paramètres intrinsèques et extrinsèques des deux caméras. Le problème de recalage revient donc à compenser la disparité entre les deux caméras dans un montage stéréo. Ils retrouvent cette disparité finale en effectuant un calcul de cette disparité basé sur l'information mutuelle pour chaque objet de l'avant-plan. La disparité finale est donc celle qui revient le plus souvent parmi celles trouvées pour chaque paire d'objets en avant-plan.

La contribution la plus intéressante de cet article est la façon dont il évalue la qualité du recalage retrouvé. En effet, ils utilisent une approche semblable à qui est utilisée dans le projet décrit dans ce mémoire. Ils sélectionnent manuellement les objets en avant-plan et utilisent ces images comme étalon. Ils calculent ensuite manuellement les disparités et effectuent le recalage. En utilisant ces images étalons avec l'algorithme qu'ils ont développé, ils comparent ces résultats avec ceux obtenus lors du recalage manuel. Cependant, ils qualifient un recalage réussi en jugeant de manière subjective un recalage réussi ou non. Cette dernière particularité peut être considérée comme inférieure à celle qui sera proposée dans le chapitre 3 de ce document qui est de quantifier cette différence par un pourcentage.

À la suite de cette section, on s'aperçoit que les méthodes qui se disent basées uniquement sur l'analyse des pixels se contentent très rarement de cette seule

information. En effet, elles doivent souvent extraire des sous-sections de l'image. La sélection de ces sous-sections est souvent basée sur des éléments caractéristiques qui ont été extraits de l'image. Il est donc clair que même pour des méthodes basées sur les pixels, les éléments caractéristiques extraits des images sont souvent essentiels pour effectuer un recalage multispectral. Des méthodes tirant encore plus profit de ce concept d'éléments caractéristiques seront présentées dans la section suivante.

#### **1.4 Méthodes de recalage infrarouge-visible basées sur les éléments caractéristiques d'une image**

Dans cette catégorie de méthodes, on retrouve les méthodes qui ne s'intéressent pas directement aux pixels qui forment l'image (plus précisément à l'intensité de ceux-ci). Il s'agit de la catégorie de méthodes la plus intuitive à utiliser pour effectuer un recalage infrarouge-visible puisque deux vidéos appartenant à ces spectres n'auront pas de couleurs ni de textures semblables. Cette section présentera plusieurs articles regroupés en deux catégories : les méthodes basées sur les contours, les méthodes basées sur des points caractéristiques.

##### **1.4.1 Méthodes basées sur le contour**

Utiliser le contour comme caractéristique pour effectuer un recalage infrarouge-visible est la méthode la plus répandue dans la littérature. Essentiellement, les méthodes proposées reposent toutes sur le même principe : détecter les contours des objets dans les deux images et les utiliser avec des méthodes de recherche de formes pour bien apparier ces contours et ainsi déduire la matrice de transformation. Elles diffèrent essentiellement sur la façon dont ces contours sont obtenus et sur la manière dont ils sont utilisés pour retrouver la matrice de transformation.

Une première façon utilisée dans la littérature pour extraire les contours est la méthode des ondelettes, plus particulièrement les transformations discrètes en ondelette. C'est notamment le cas dans Huang & Chen (2002). Dans cet article, la transformation discrète en ondelettes est utilisée pour extraire les contours dans l'image. Ensuite, pour

apparier les contours de l'image infrarouge avec ceux de l'image couleur, ils ont besoins d'une mesure de similarités invariante à la translation, à la rotation et au grossissement puisqu'ils utilisent une matrice de transformation affine (Hartley & Zisserman, 2002). Dans un premier temps, les moments invariants (Hu, 1962) sont calculés pour chaque contour dans les deux images. Ensuite, on calcule la distance euclidienne entre les moments invariants (il y en a 7) de chacune des paires de contours qui peuvent être formées. On minimise cette distance pour connaître les contours les plus similaires pour en déduire la matrice de transformation. En plus des moments invariants, une mesure complexe basée sur l'orientation des contours est aussi utilisée pour apparié les contours de la première image à la seconde. La paire de contours qui maximise la mesure obtenue en combinant ces deux méthodes de comparaison est finalement utilisée pour connaître la matrice de transformation.

Une autre façon d'utiliser les contours pour effectuer un recalage infrarouge-visible est proposée par Coiras, Santamaria, & Miravet (2000). Eux aussi commencent par extraire les contours des objets dans les deux séquences vidéo. Ils mentionnent avoir utilisé plusieurs méthodes comme les méthodes de Canny (Canny, 1986) et Sobel (Sobel, 1978). Cependant, ils concluent que le choix de la méthode n'a pas une grande influence sur la qualité du recalage. Une fois ce contour obtenu, ils identifient des segments et rassemblent ces segments en triangles. Un triangle est formé de trois segments non colinéaires qui ne se croisent pas en un même point. Ces triangles sont calculés dans les 2 images. Pour chaque paire de triangles pouvant être formée entre les deux images, ils calculent une matrice de transformation affine et calculent un critère de qualité de la transformation trouvée. Ce critère de qualité est basé sur la distance du centre d'un segment ainsi que la différence d'angle avec son correspondant dans l'autre image. Ils additionnent ces distances pour chaque segment et choisissent la transformation avec la plus petite différence. Ils ne quantifient pas les résultats de leur recalage, mais leur méthode semble fonctionner dans les quelques images qu'ils présentent.

Kong et al. (2007) aussi propose une méthode de recalage infrarouge-visible basée sur les contours dans les images. Cet article utilise la méthode de Canny (Canny, 1986) pour

déTECTer les contours dans des images de visages et effectuer le recalage. Comme critère utilisé pour évaluer la qualité de la transformation affine calculée, la distance de Hausdorff est utilisée comme proposée par Huttenlocher, Huttenlocher, Klanderman, & Ruckridge (1993). Les résultats de ce recalage ne sont pas détaillés puisqu'il s'agit d'un article qui s'intéresse plus à la fusion d'images infrarouge-visible et à la reconnaissance de visage. Le recalage est vu comme une étape préalable obligatoire.

Pour terminer sur l'utilisation des contours, Joo, Choi, & Cho (2003) proposent aussi d'effectuer le recalage de cette façon mais introduisent une notion qui sera utilisée dans la méthode proposée dans ce document, les objets en mouvement. En effet, ils n'utilisent que les contours des objets en mouvement pour retrouver une transformation affine. Pour détECTer ces objets en mouvement, ils utilisent une forme de soustraction d'arrière-plan. Ils testent chaque paire possible de contours fermés entre les deux images et retrouvent la meilleure transformation affine possible entre chacune des paires. Pour se faire, ils font varier les différents paramètres de la transformation (translation, rotation et agrandissement) et évaluent la qualité du recalage grâce à une mesure de distance moyenne entre les contours. Plus cette distance est petite, meilleure est la transformation. Les performances ne sont pas quantifiées mais ils concluent que leur méthode fonctionne.

#### **1.4.2 Méthodes basées sur des points caractéristiques**

En théorie, il suffit d'un nombre limité de points ayant une correspondance connue dans l'autre image pour effectuer un recalage. C'est sur cette constatation que reposent les méthodes utilisant les points caractéristiques. La principale difficulté de cette approche est de déterminer la nature de ces points qui peuvent être facilement appariés avec leur correspondant dans des images infrarouges et visibles. Pour cette raison, cette approche est peu utilisée dans les cas de recalage infrarouge-visible utilisant des séquences vidéo générales. En contrepartie, les applications plus spécifiques comme celles en reconnaissance de visage sont plus nombreuses.

Puisque les techniques de reconnaissance de visages utilisent souvent des points caractéristiques pour différencier les visages, elles représentent un bel exemple de recalage infrarouge-visible basé sur les points caractéristiques comme c'est le cas de Wang, Sung, & Venkateswarlu (2004). Dans cet article, on fait mention d'un système calibré qui utilise des points caractéristiques comme les yeux et les coins de la bouche pour retrouver la matrice de transformation affine. Le résultat de ce recalage est ensuite utilisé pour améliorer les résultats de la reconnaissance de visage.

Une autre technique, utilisée pour effectuer un recalage infrarouge-visible, est proposée par St-Onge (2008). Ce travail fait mention de la combinaison de deux méthodes pour générer des points à appariер soit la méthode du squelette et celle du processus DCE. La méthode du squelette consiste d'abord à trouver les axes de symétrie locaux à l'intérieur d'une silhouette humaine obtenue en effectuant une soustraction d'arrière-plan. Les points retenus dans ces squelettes sont ceux aux intersections des arêtes (axes de symétrie locale). Le processus DCE consiste à simplifier le contour d'un « blob » en sélectionner certains points le long de ce contour. Pour apparier tous ces points, l'auteur utilise une combinaison de critères de ressemblance qui tient compte de la distance, de l'orientation et de la nature du point (s'il est du squelette ou du contour). Une paire de points est formée lorsque les deux points ont un pointage de ressemblance très élevé. À partir de ces paires de points bien appariés, l'auteur calcule la matrice fondamentale. Cette matrice ressemble à la matrice homographique, mais elle est plutôt utilisée dans les situations où la profondeur est un facteur ce qui n'est pas le cas dans ce projet.

Un autre article utilise les contours pour obtenir des points caractéristiques. Il s'agit de l'article de Elbakary & Sundareshan (2007). Dans cet article, les contours sont obtenus en utilisant un filtre de Gabor (Forsyth & Ponce, 2003). Ensuite, un certain nombre de points sont sélectionnés de ces contours en utilisant un seuil. Ces points sont ensuite appariés à d'autres en utilisant une méthode de calcul de matrice de transformation robuste face aux données aberrantes.

Un autre article qui effectue un recalage infrarouge-visible en utilisant des points caractéristiques est celui de Hild & Umeda (2005). Dans cet article, le recalage est effectué de deux façons : d'une façon dite locale et d'une façon dite générale. Leur approche locale repose essentiellement sur un montage déterminé des caméras qui permet de déduire la position en trois dimensions de chacun des points de l'image couleur et infrarouge par triangulation. Étant donné que les deux caméras doivent avoir des plans images parallèles, il est possible de calculer la position du point dans un espace en trois dimension grâce à la disparité entre les deux images des caméras. Ainsi, par le même procédé, on peut retrouver où devrait se situer le point correspondant dans l'image infrarouge et calculer la matrice de transformation entre les deux images en se basant sur les différences observées. L'approche générale est pour sa part calculée en utilisant un point de référence répondant à des critères précis. À partir de ce point, une surface est calculée et mise en correspondance dans l'autre image grâce à la position dans l'espace en trois dimensions. On tente ensuite de retrouver la transformation qui permet d'expliquer les différences dans la forme de cette surface. Ces deux approches, locale et générale, sont combinées dans l'espoir de trouver le meilleur recalage possible. Cette méthode est très restrictive et les résultats de la qualité du recalage ne sont pas quantifiés.

Une autre méthode de recalage infrarouge-visible reposant sur des points caractéristiques est celle de Bhanu & Ju (2007). Cette méthode est d'autant plus intéressante par le fait qu'elle propose l'utilisation des mêmes caractéristiques qui sont utilisées par la méthode qui sera présentée dans ce mémoire : les images en avant-plan et des points caractéristiques de ces dernières comme leurs centroïdes. Un centroïde est le point situé au milieu d'un objet. En traitement d'image, on réfère souvent au centroïde comme le point central d'un « blob ». Rappelons que pour trouver la matrice de transformation, il suffit d'obtenir quelques points pour lesquels la correspondance dans l'autre image est connue. Sachant que les objets en mouvement dans une image seront nécessairement aussi en mouvement dans l'autre image, il est possible d'utiliser ces informations comme caractéristique commune. Dans leur article, Bhanu & Ju (2007) se

servent de cette idée et calculent le centroïde d'une personne marchant en ligne droite dans une séquence vidéo. Pour détecter cette personne, ils utilisent une soustraction d'arrière-plan simple gaussienne (McKenna, Jabri, Duric, Rosenfeld, & Wechsler, 2000). Comme cette séquence est composée de plusieurs images, ils obtiennent plusieurs centroïdes et donc plusieurs points pour calculer la matrice de transformation. Comme une seule personne marche dans la séquence, l'appariement des points est très simple. Cependant, le recalage ainsi obtenu est souvent de piètre qualité. Pour améliorer cette qualité, ils utilisent un algorithme génétique qui permet de déplacer le point de correspondance dans une fenêtre. On essaie ainsi de maximiser la ressemblance grâce à une formule dite de produit des similarités. Il est aussi intéressant de constater qu'eux aussi utilisent le recouvrement des deux images pour évaluer la qualité d'un recalage. Cependant, comme ils n'utilisent que les images d'avant-plan, leurs résultats seront liés à la méthode de soustraction d'arrière-plan utilisée.

Une dernière méthode de recalage infrarouge-visible basée sur des caractéristiques est celle de Caspi et al. (2006). Globalement, cet article porte sur le recalage de différentes séquences vidéo qui sont filmées avec des caméras qui peuvent être de spectres identiques ou différents, mais aussi avec des caméras qui sont positionnées dans des angles de vue très différents et sans être nécessairement synchronisées. Cependant, la partie qui s'approche du projet présenté ici est celle sur le recalage infrarouge-visible. À ce sujet, ils proposent d'utiliser les trajectoires des objets en mouvement pour calculer la matrice de transformation entre les deux vidéos. Il est à noter que la matrice retrouvée pour cette application précise est la matrice de transformation générale. Pour calculer ses différents paramètres, ils utilisent les centroïdes ou les sommets des « blobs » d'avant-plan dans l'image. Un sommet de « blob » fait ici référence à un point à deux dimensions dont les coordonnées sont les suivantes :

- La coordonnée en X est la moyenne des coordonnées en X de chacun des pixels qui composent le « blob ».

- La coordonnée en Y est égale à la coordonnée en Y du pixel le plus en haut dans le « blob ».

Cette notion est expliquée plus en détail dans le prochain chapitre puisque ces sommets sont utilisés dans la nouvelle technique de recalage présenté dans ce texte.

L'originalité première de cet article est dans l'appariement de ces points; ils utilisent les trajectoires formées par l'union de ces différents points comme caractéristique d'appariement. En utilisant un algorithme de type RANSAC, ils parcourront chaque paire de trajectoires possible et calculent la somme des erreurs de chaque point formant les trajectoires en comparant les estimations obtenues par la matrice de transformation calculée avec les points observés. Ils choisissent la matrice qui minimise cette mesure d'erreur. Il faut cependant noter que l'approche qu'ils proposent est appliquée une fois la séquence vidéo terminée. Elle n'est donc pas utilisée en temps réel. Les résultats semblent prometteurs mais ne sont pas quantifiés. Malgré tout, c'est en grande partie cette approche qui a inspiré le projet qui sera maintenant exposé en détail.

## CHAPITRE 2 MÉTHODOLOGIE

Dans cette section, il est question de la méthode qui a été développée pour réussir à effectuer automatiquement le recalage entre la séquence vidéo infrarouge et couleur. Dans un premier temps, un résumé de la méthode développée sera proposé à la section 2.1. Ensuite, les différents prétraitements appliqués aux séquences vidéo sont expliqués à la section 2.2. Il est ensuite question, à la section 2.3, des différentes matrices de transformation homographiques existantes et des raisons qui ont motivé le choix de la matrice de transformation homographique dite affine. Pour ce qui est de la section 2.4, elle porte sur l'algorithme développé d'appariement des trajectoires. Pour finir, l'utilité et les différentes étapes nécessaires à l'implantation d'un choix de matrice de transformation à chaque trame sont expliquées plus en détail à la section 2.5.

### 2.1 Aperçu de la méthode

Voici une figure qui résume le fonctionnement du recalage automatique développé.

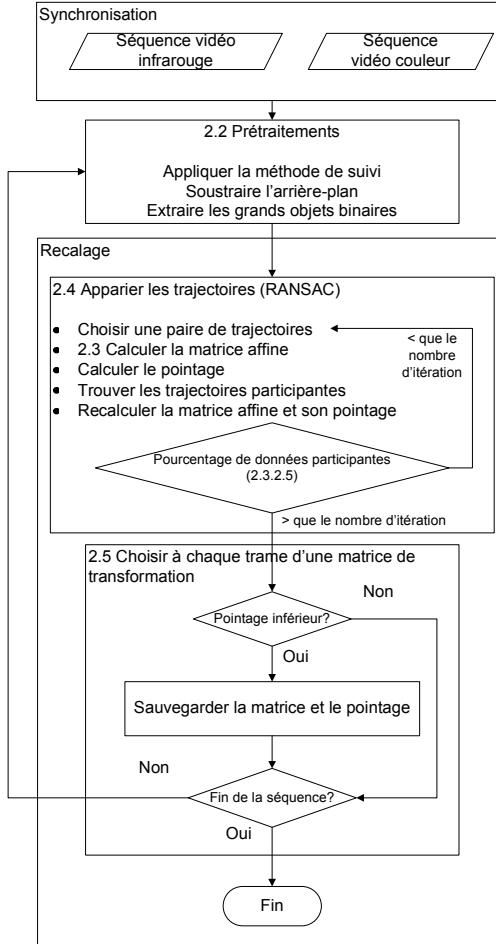


Figure 2.1 Aperçu global de l'algorithme de recalage automatique

Les données utilisées sont des séquences vidéo qui ont été filmées à l'aide de deux caméras opérant dans deux spectres différents : l'infrarouge et la lumière visible. Ces deux séquences vidéo ont aussi été obtenues par une synchronisation logicielle. Cette synchronisation fait partie des hypothèses de départ. Cependant, il sera question, dans le chapitre des résultats, des imperfections de cette synchronisation et de l'impact de ces erreurs sur les performances du recalage. Lors des différents tests, la caméra infrarouge se situait à gauche tandis que la caméra couleur se situait à droite. Il fut supposé aussi que les objets se déplacent à peu près dans le même plan. En d'autres mots, la distance entre les différents objets de la scène et la caméra doit être à peu près identique. Ceci permet de faire le recalage de tous les objets en même temps.

La première opération de prétraitement nécessaire à l'algorithme de recalage est l'application d'une méthode de suivi qui permet d'extraire les trajectoires des objets en mouvement. Ces trajectoires seront utilisées comme éléments de comparaison puisque les méthodes plus classiques peuvent difficilement être utilisées pour comparer deux séquences vidéo de spectres différents. Rappelons ici qu'un recalage consiste essentiellement à mettre en correspondance des points entre deux images pour pouvoir calculer une matrice de transformation qui permet de passer d'un repère image à l'autre. Pour la méthode proposée dans ce texte, ce sont les trajectoires qui sont mises en correspondance et ce sont les points qui composent ces trajectoires qui sont utilisées pour calculer la matrice de transformation. En second lieu, il est nécessaire d'effectuer une soustraction d'arrière-plan pour obtenir l'avant-plan. Ces images d'avant-plan sont utilisées pour calculer le pointage qui donnera une mesure de qualité à une matrice de transformation trouvée. Cette opération est suivie d'une extraction de grands objets binaires (« blob » : terme anglais désignant un groupe de pixels interconnectés) qui sert principalement à filtrer les images d'avant-plan afin d'éliminer les différents défauts qui apparaissent lors de la soustraction d'arrière-plan.

L'appariement des trajectoires est au centre de la méthode de recalage automatique proposée. La méthode proposée est basée sur un algorithme RANSAC (Fischler & Bolles, 1981). Elle pige au hasard une paire de trajectoires et calcule la matrice de transformation affine qui lui est reliée. Elle parcourt ensuite toutes les autres paires de trajectoires pour trouver celles qui améliorent le pointage obtenu lors de la pige initiale. Ces paires sont nommées paires participantes. En ajoutant les points qui composent ces paires, on recalcule une dernière fois la matrice de transformation affine. Ces étapes sont répétées le nombre de fois prescrit par la théorie statistique reliée à l'algorithme RANSAC qui sera expliqué plus loin dans ce chapitre.

Une fois l'étape de l'appariement des trajectoires terminée, la phase dite de choix en temps réel est effectuée. L'expression temps réel est ici utilisée pour résumer qu'une matrice de transformation est calculée à chaque trame. Nous ne nous attarderons pas aux performances de temps de calcul, car le code utilisé pour tester la méthode n'a

nullement fait l'objet d'une phase d'optimisation. De toute façon, le calcul d'une matrice de transformation par trame n'est pas un objectif en soit. En fait, il est très probable que quelques trames seulement soient nécessaires pour obtenir une convergence. Nous continuons simplement de l'appliquer avec l'espoir de voir la précision du recalage augmenter et s'approcher le plus possible du recalage manuel afin de trouver le niveau de performance maximal qui peut être atteint par la méthode.

Durant la phase de choix en temps réel, la qualité du recalage obtenu à un temps donné est comparée avec la qualité de ceux obtenus précédemment. Pour se faire, le pointage est encore une fois utilisé. Après cette étape, il ne reste qu'à vérifier si les séquences vidéo sont terminées. Si elles le sont, le programme se termine, sinon, toutes les opérations précédentes sont reprises à partir des prétraitements pour la prochaine image. Notons aussi que chaque calcul de matrice, effectué à chaque trame, est indépendant des précédents puisque nous ne voulons pas que des erreurs passées viennent fausser les calculs de matrices subséquentes. La seule opération qui tient compte des matrices passées est celle de la phase du choix qui détermine si la nouvelle matrice calculée semble plus précise que la dernière.

## 2.2 Prétraitements

Dans cette section, les différents prétraitements qui doivent être appliqués aux images des séquences vidéo avant qu'elles ne puissent être utilisées par l'algorithme de recalage automatique sont détaillés. Dans un premier temps, la méthode de soustraction d'arrière-plan utilisée et son utilité sont exposées à la section 2.2.1. Ensuite, l'extraction des « blobs » qui est utilisée en combinaison avec la soustraction d'arrière-plan est abordée à la section 2.2.2. Finalement, l'algorithme de suivi utilisé pour trouver les trajectoires des objets en mouvement est expliqué à la section 2.2.3. Par ailleurs, voici un schéma qui résume comment ces prétraitements transforment les séquences vidéo initiales en données utilisables par l'algorithme de recalage automatique.

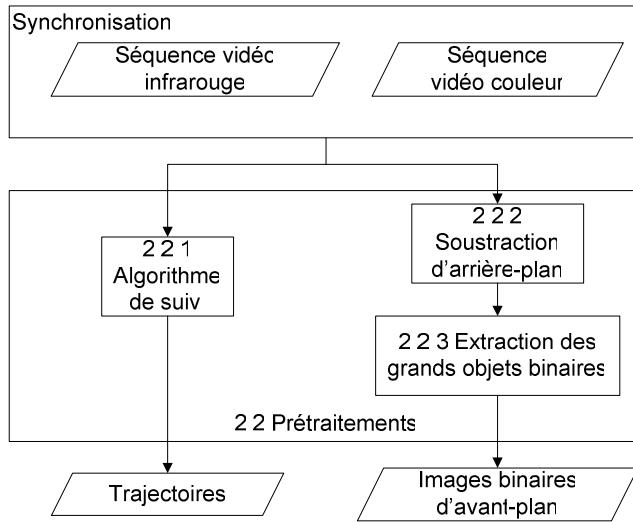


Figure 2.2 Schéma de transformation des données par les prétraitements

### 2.2.1 Images binaires d'avant-plan : soustraction d'arrière-plan

Les images binaires d'avant-plan sont utilisées par le recalage automatique pour évaluer si la matrice de transformation trouvée est de bonne qualité. Les détails du processus qui donne un pointage en fonction de cette qualité seront donnés plus loin. Les images sont transformées en images binaires car la soustraction d'arrière-plan a pour but de classer tous les pixels en 2 groupes : les pixels appartenant à l'avant-plan et les pixels appartenant à l'arrière-plan. Ainsi, les pixels classés comme étant dans l'avant-plan seront blanc et les pixels classés dans l'arrière-plan seront noir. Ceci simplifie beaucoup les opérations subséquentes. Pour obtenir des images d'avant-plan, il faut bien sûr soustraire l'arrière-plan. L'arrière-plan peut être défini comme une image contenant tous les éléments statiques d'une scène. Par opposition, l'avant-plan est composé de tous les objets qui se déplacent dans la scène. Plusieurs méthodes ont déjà été proposées dans la littérature. Elles ont toutes leurs avantages et leurs inconvénients. Il faut donc identifier les éléments requis et les limitations imposées par les séquences filmées ainsi que par l'algorithme de recalage utilisé. Les séquences qui ont été utilisées ont toutes été filmées à l'intérieur. Donc il y a absence de petits objets en mouvement comme des feuilles qui bougent au vent. Bien sûr, cette affirmation n'est vérifiable que s'il y a bel et bien absence de courants d'air qui pourraient déplacer de petits objets dans la pièce filmée ce

qui était le cas dans les séquences utilisées. Aussi, ces séquences sont plutôt courtes. Il est donc nuisible que les objets qui arrêtent de bouger s'intègrent trop rapidement à l'arrière-plan. Cependant, la plus grande limitation vient de l'algorithme de recalage. En effet, ce dernier est relativement gourmand en temps de calcul. Il faut donc que l'algorithme de soustraction d'arrière-plan utilisé soit très rapide et qu'il ne consomme pas beaucoup de mémoire. À la lumière de ces requis et suite à de nombreux essais, l'algorithme de la moyenne temporelle (Shoushtarian & Bez, 2005) s'est révélé un choix acceptable dans plusieurs situations. Cette méthode fait la moyenne de la valeur de chaque pixel lorsqu'il n'y a pas d'objet dans la scène, et l'avant-plan est détecté en comparant la valeur courante d'un pixel avec sa moyenne calculée dans les trames initiales. Si l'écart est grand, cela signifie que le pixel a changé et provient d'un objet en mouvement. Le processus de soustraction d'arrière-plan est illustré par les deux figures suivantes. À la Figure 2.3, deux images à traiter sont montrées suivies par le résultat à la Figure 2.4 des avant-plans après la soustraction d'arrière-plan.

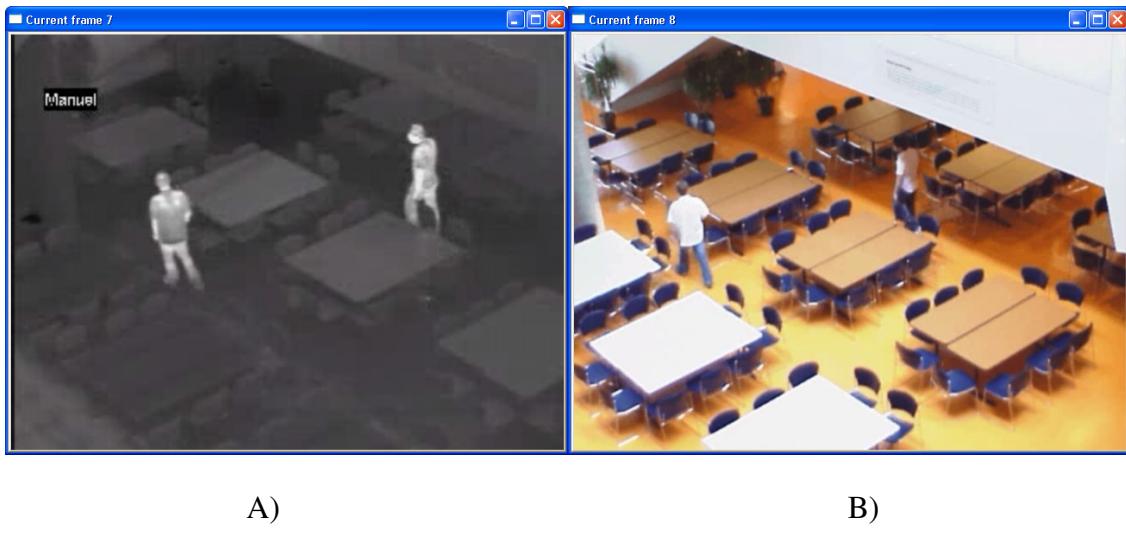


Figure 2.3 Images infrarouge (A) et couleur (B) avant la soustraction d'arrière-plan

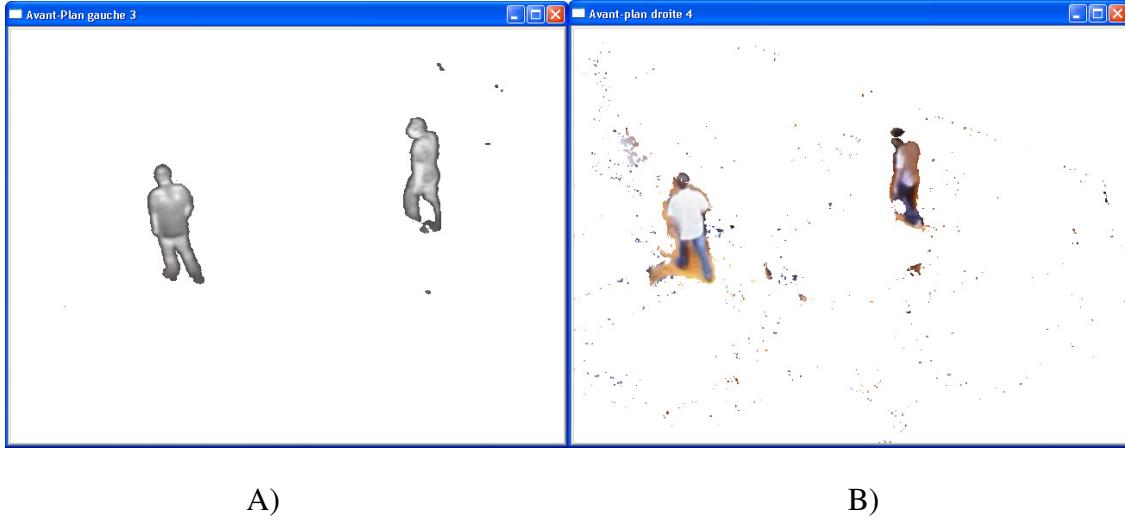
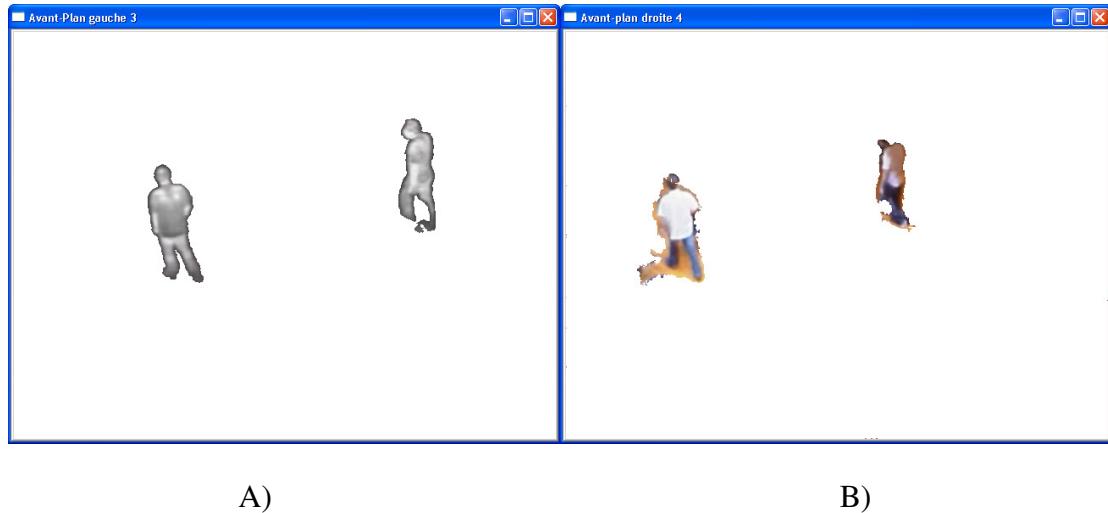


Figure 2.4 Images infrarouge (A) et couleur (B) d'avant-plan de la méthode de la moyenne temporelle

## 2.2.2 Images binaires d'avant-plan : extraction des grands objets binaires (« blobs »)

Une fois la soustraction d’arrière-plan effectuée, des images d’avant-plan sont obtenues en assignant la couleur blanche à tout pixel faisant partie de l’arrière-plan. Il faut ensuite les rendre binaires afin de simplifier le calcul du pointage de recouvrement détaillé à la section 2.4.2. Même s’il avait été plus simple d’appliquer directement un filtre binaire (0 si le pixel est blanc et 1 sinon) sur les images, appliquer d’abord une méthode d’extraction des grands objets binaires rend la procédure plus robuste. En effet, les erreurs dans la soustraction d’arrière-plan sont fréquentes pour quelques pixels isolés (voir Figure 2.4 B). En utilisant un algorithme de connexion de pixels, il est possible d’extraire les « blobs » (groupes de pixels connectés) mais aussi de supprimer les pixels isolés. De plus, un deuxième filtre qui rejette les « blobs » trop petits a aussi été utilisé. De cette façon, des bouts d’ombres non rattachés au « blob » principal ont pu être supprimés. Cette extraction de « blobs » a été réalisée à l’aide de la fonction *cvFindContour()* de la bibliothèque OpenCV (Intel, 2006). Il est possible de voir le résultat après ces traitements à la Figure 2.5. Une fois ces filtres appliqués, il ne reste

plus qu'à rendre binaires les images comme décrit précédemment. Ces images binaires d'avant-plan sont illustrées à la Figure 2.6.



A)

B)

Figure 2.5 Images infrarouge (A) et couleur (B) d'avant-plan après le filtre de taille de « blob »



A)

B)

Figure 2.6 Images infrarouge (A) et couleur (B) binaire d'avant-plan

### 2.2.3 Calcul des trajectoires : algorithmes de suivi

Comme expliqué à la figure précédente, les points des trajectoires sont les données qui sont utilisées lors du calcul de la matrice de transformation affine. Trouver des trajectoires revient à détecter les objets en mouvement et les situer dans l'image. C'est

aussi la définition d'un algorithme de suivi. Mathématiquement, la méthode de recalage automatique développée définit une trajectoire comme une matrice contenant les coordonnées homogènes en deux dimensions des positions des « blobs » dans l'image en fonction du temps. L'équation suivante résume cette définition.

$$Ti_{image} = \begin{bmatrix} X_1 & X_2 & \dots & X_n \\ Y_1 & Y_2 & \dots & Y_n \\ 1 & 1 & \dots & 1 \end{bmatrix} \quad (2.1)$$

Dans l'équation 2.1,  $Ti_{image}$  représente la trajectoire  $i$  pour l'image de gauche ou de droite. Aussi, les coordonnées ( $X_j$  et  $Y_j$ ) du « blob » dans le système de coordonnées de l'image (donc en pixels) sont données. De plus, le nombre  $j$  qui suit la coordonnée représente le numéro de trame pour laquelle ce point de la trajectoire a été observé. De plus, comme ces trajectoires sont la matière première de la méthode de recalage automatique développée, leur qualité a des conséquences sur le recalage. Les problèmes typiques rencontrés dans les méthodes de suivi sont les suivants : problèmes d'occlusion, problèmes de sortie et d'entrée dans la scène et problèmes de fusion/séparation de « blobs ». Il est donc possible de relier la qualité d'une trajectoire à l'influence que ces différents problèmes auront sur elle. En résumé, un algorithme de suivi qui est capable de gérer correctement tous ces problèmes donnera une trajectoire de très grande qualité. À l'inverse, un algorithme ne prenant pas en compte ces problèmes donnera une trajectoire de mauvaise qualité.

Les résultats montrent que la qualité d'une trajectoire affecte les résultats du recalage (voir CHAPITRE 3). Malheureusement, une seule méthode de suivi a été testée par manque de temps. Celle-ci est plutôt rudimentaire. Pour commencer, cette méthode effectue une soustraction d'arrière-plan suivie d'une extraction de « blobs ». Ces deux étapes sont réalisées de la même façon que dans les sections 2.2.1 et 2.2.2. À partir des images binaires obtenues, un point caractéristique pour chaque « blob » est calculé. Ce sont ces points caractéristiques qui forment les trajectoires (éq. 2.1). Dans ce projet, c'est le sommet de la partie du haut des « blobs » qui a été utilisé comme point

caractéristique. Le sommet de la partie du haut est calculé en trouvant la moyenne des coordonnées en X des pixels du « blob » et la coordonnée maximale en Y de ces mêmes pixels. L'équation suivante résume ces calculs.

$$\begin{aligned}
 Sommet &= (S_X, S_Y) \\
 S_X &= \frac{\sum X}{nbPixel} \\
 S_Y &= MAX(Y)
 \end{aligned} \tag{2.2}$$

Il faut appairer ces « blobs » avec ceux de l'image précédente pour construire les trajectoires. Pour se faire, la méthode des rectangles englobants en recouvrement est utilisée (Fuentes & Velastin, 2006). Les trajectoires sont construites selon le recouvrement des « blobs » au temps  $t$  avec ceux au temps  $t-1$ . Si aucun recouvrement n'est trouvé pour un « blob » donné, une nouvelle trajectoire est créée. Si plus d'un « blob » au temps  $t-1$  est en recouvrement avec un « blob » au temps  $t$ , un critère basé sur la surface du « blob » est utilisé. En effet, ce « blob » au temps  $t$  sera associé à la trajectoire qui a une moyenne d'aire de « blob » la composant le plus près de ce « blob » au temps  $t$ . Cet algorithme n'est donc pas robuste aux erreurs classiques de suivi mentionnées précédemment.

À partir de ces trajectoires, il devient possible de calculer la matrice de transformation comme le montre la prochaine section.

### 2.3 Calcul de la matrice de transformation à partir des points de trajectoires

Par définition, effectuer le recalage entre deux images revient à trouver la matrice de transformation qui permet de passer du système de coordonnées d'une image à celui d'une autre. Mathématiquement, ce processus est résumé par l'équation suivante :

$$X_{DROIT} = H \times X_{GAUCHE} \tag{2.3}$$

Le symbole  $X_{DROIT}$  représente les coordonnées de tous les pixels de l'image de droite,  $H$  la matrice de transformation homographique et  $X_{GAUCHE}$  les coordonnées de tous les pixels de l'image de gauche. Déterminer cette matrice peut devenir très complexe si l'on

ne connaît pas la correspondance entre les points des deux images. Et même si cette correspondance est connue, plusieurs matrices de transformation peuvent expliquer mathématiquement cette correspondance. C'est pour éviter ces situations de solutions multiples que quelques simplifications ont été faites comme expliquées dans la section 2.3.1. Aussi, le choix des points utilisés pour calculer la matrice de transformation s'avère primordial. C'est pourquoi plusieurs méthodes de calcul de matrice de transformation sont présentes dans la littérature. Celle qui a été retenue, ainsi que les raisons qui ont motivé ce choix sont expliquées à la section 2.3.2.

### 2.3.1 Justification du choix de la simplification

Comme le problème qui nous intéresse ne porte pas sur la position des objets dans le monde réel, la problématique est simplifiée en passant d'un problème tridimensionnel à un problème bidimensionnel. Étant donné que les coordonnées homogènes d'un espace bidimensionnel nécessitent des vecteurs à trois données, la matrice de transformation générale qui permet de passer d'une image à l'autre est donc une 3 par 3 ayant 9 coefficients à évaluer. Cette matrice est présentée par l'équation 2.4. Elle est représentée par le symbole  $H_{générale}$  et ses coefficients par les symboles  $h_{ij}$  où  $i$  représente le numéro de ligne de la variable dans la matrice et  $j$  son numéro de colonne.

$$H_{générale} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \quad (2.4)$$

Cette matrice est souvent appelée matrice homographique. Les 9 variables qui la composent demandent au moins 9 équations linéaires afin de déterminer de manière unique la valeur de chacune de ces variables. Comme les coordonnées des points dans les images sont en deux dimensions, elles fournissent donc deux équations chacune. Il faudrait donc théoriquement 5 points pour retrouver la matrice homographique générale. Cependant, le choix de ces points est primordial. Par exemple, des points alignés entraîneraient une situation de dégénérescence et ne permettraient pas de retrouver la matrice recherchée. Aussi, divers problèmes reliés à la déformation de l'image causée

par la disposition des caméras complexifient encore plus le calcul de la matrice homographique. Tous ces différents problèmes sont abondamment traités dans la littérature notamment dans Hartley & Zisserman (2002).

Pour simplifier quelque peu le problème, des restrictions ont été imposées comme mentionné précédemment dans la section sur les hypothèses de base de l'introduction. Ainsi, aucune lentille grand-angle n'a été utilisée, car elles déforment les images. Aussi, les deux caméras ont été fixées sur des barres où seule une translation en X est possible. Par contre, une rotation le long de l'axe des Z (voir figure 2.7) est aussi tolérée entre les deux caméras. De plus, les caméras peuvent avoir un facteur de grossissement différent. Aussi, il faut qu'une certaine portion de la scène soit visible par les deux caméras puisque, dans le cas contraire, il n'y aurait pas d'information commune. Cette portion ne semble pas avoir un grand impact sur la qualité du recalage puisque des résultats satisfaisants sont obtenus même avec une zone d'inclusion des champs de vision aussi petite que 23 % (chapitre 3). En plus des restrictions sur les caméras, il faut aussi faire une supposition sur la position relative des objets de l'avant-plan par rapport à la caméra. En effet, ces objets doivent faire partie d'un même plan, soit être à peu près à la même distance de la caméra. Cette restriction permet de recaler tous les objets en même temps (Caspi et al., 2006). Pour respecter cette restriction, les séquences vidéo utilisées ont été tournées en plongée. Voici maintenant une figure qui résume le système d'axe que doivent respecter les caméras :

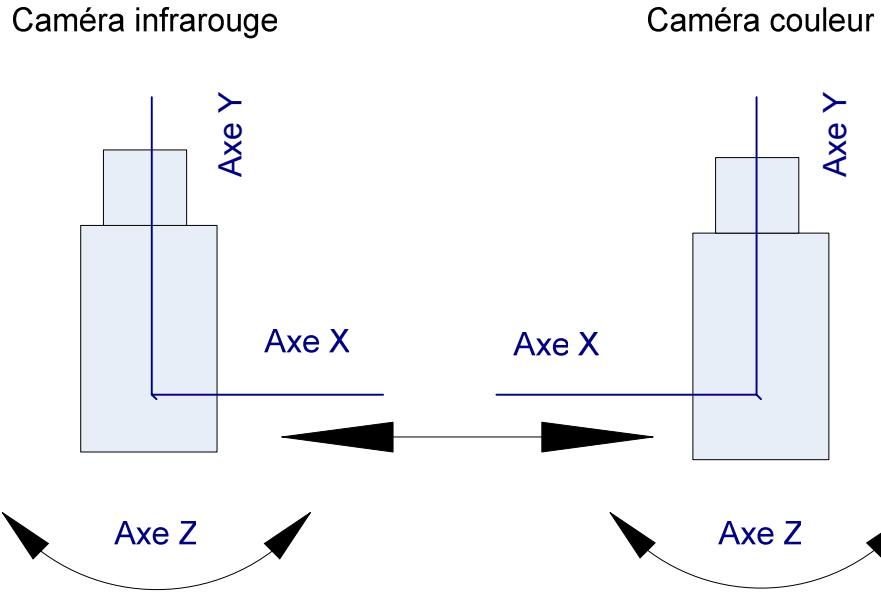


Figure 2.7 Vue schématique en plongée du montage des caméras

Par ces restrictions, ce n'est plus la matrice homographique générale qui est recherchée, mais plutôt une simplification de celle-ci appelée la matrice de transformation affine (Hartley & Zisserman, 2002). Mathématiquement, elle est définie par l'équation 2.5 et ne nécessite que 3 points pour être résolue. En effet, rappelons que chaque point donne 2 équations (points en 2 dimensions) et que la matrice est formée de 6 inconnus.

$$H_{affine} = \begin{bmatrix} a_{11} & a_{12} & t_x \\ a_{21} & a_{22} & t_y \\ 0 & 0 & 1 \end{bmatrix} \quad (2.5)$$

### 2.3.2 Calcul de la matrice de transformation affine

Pour calculer cette matrice de transformation affine, rappelons que les points utilisés sont obtenus d'une paire de trajectoires. Cette paire est composée d'une trajectoire de la séquence vidéo de gauche et d'une trajectoire de celle de droite. Ces trajectoires sont composées d'au moins 3 points chacune pour répondre à la restriction posée par l'utilisation d'une matrice affine. Plusieurs méthodes capables de calculer les matrices de transformation homographiques sont proposées dans la littérature. Un livre particulièrement intéressant à ce sujet est Hartley & Zisserman (2002). On y fait

mention des méthodes les plus utilisées et les situations où elles s'appliquent le mieux. Dans la situation qui nous intéresse, l'architecture en temps réel développée combinée avec l'utilisation de séquences vidéo qui fournissent un grand nombre d'images amène à conclure que le nombre de points qui sera disponible sera beaucoup plus grand que les trois points nécessaires pour déterminer une matrice de transformation affine. Il faut donc privilégier une méthode capable de tirer avantage de tous ces points. De plus, il est très probable que ces points ne soient pas sans erreur. Il faut donc une méthode robuste qui permet de minimiser l'impact de données aberrantes. Il s'agit du principal avantage des algorithmes de type RANSAC (Fischler & Bolles, 1981). C'est pourquoi la méthode proposée dans Hartley & Zisserman (2002) à la section 4.8 a été retenue. Voici, dans un premier temps, les grandes lignes de l'algorithme proposé. Chacune de ces étapes ainsi que leurs applications seront détaillées dans les sections suivantes.

1. Trouver des points d'intérêt;
2. Déduire des correspondances;
3. Estimation de type RANSAC;
  - a. Calculer une estimation de la matrice de transformation;
  - b. Calculer la distance entre l'estimation et l'observation pour chaque correspondance;
  - c. Calculer le nombre de correspondances participantes;
4. Raffinement de la matrice de transformation.
  - a. Nouveau calcul de la matrice de transformation en utilisant toutes les correspondances participantes;
  - b. Identifier les nouvelles correspondances participantes et le nombre total de correspondances participantes;

### **2.3.2.1 Première étape : identifier les points d'intérêts**

L'expression points d'intérêt signifie ici les points qui sont utilisés pour calculer la matrice de transformation affine. Dans le cas présent, deux types de points ont été testés : les coordonnées des centroïdes et les coordonnées du sommet des « blobs » qui forment les trajectoires. Des tests préliminaires ont rapidement montré que les sommets des « blobs » donnent de meilleurs résultats. En effet, les centroïdes étant beaucoup plus sensibles aux changements de forme des « blobs », ils induisent souvent des déviations dans la trajectoires qui ne peuvent être retrouvés dans les deux images. Ceci a comme conséquence de rendre beaucoup plus complexe l'appariement des trajectoires et cet appariement devient aussi beaucoup moins précis, voir carrément erroné. Citons quelques sources courantes qui modifient la forme des « blobs » comme l'ombrage ou les occlusions partielles. Par conséquent, les tests dont il sera question au chapitre 3 utilisent tous les sommets des « blobs ».

### **2.3.2.2 Deuxième étape : déduire des correspondances**

Avant de se lancer plus en détail dans l'algorithme, rappelons la définition de quelques termes à l'aide de la figure suivante.

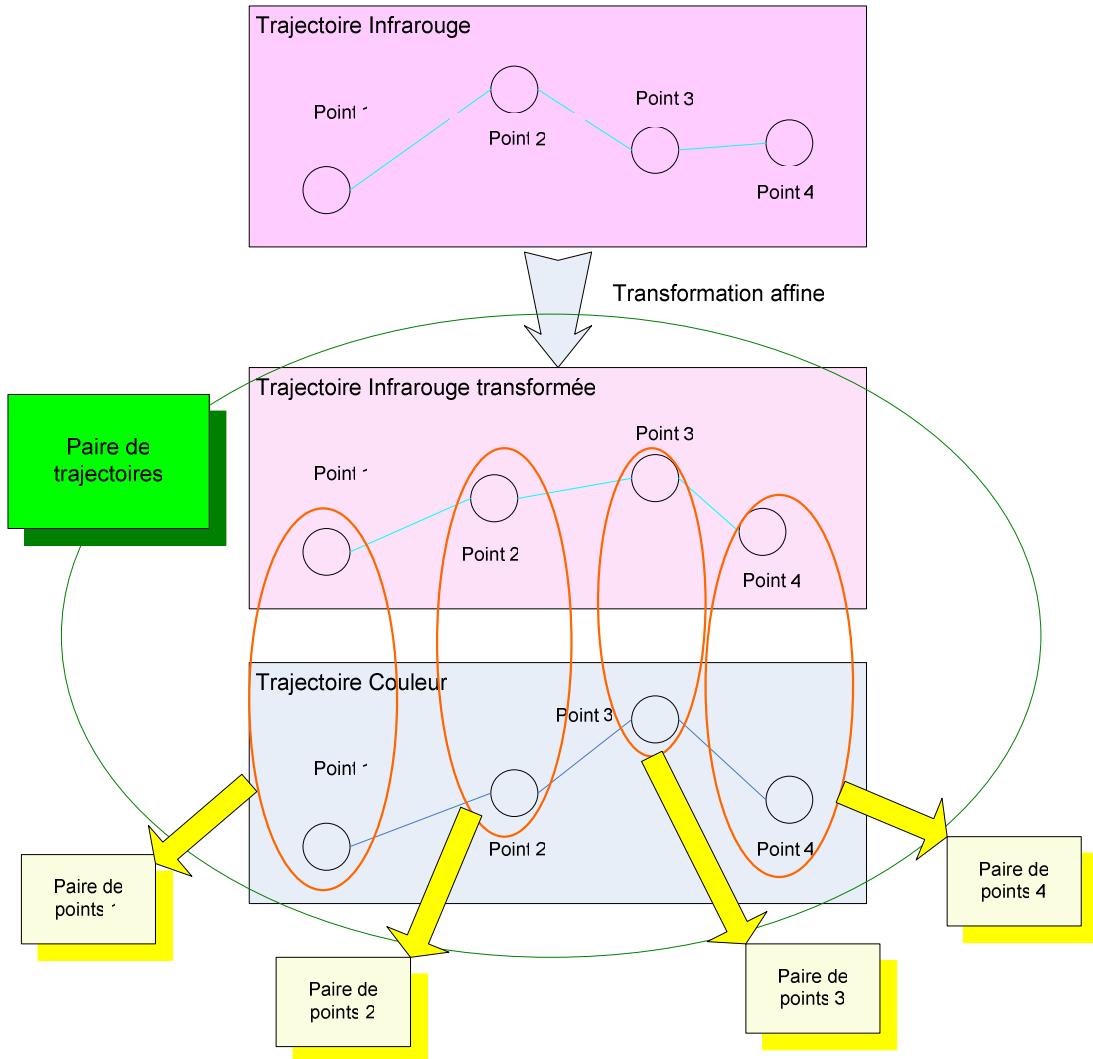


Figure 2.8 Illustrations des définitions des termes paire de trajectoires et paire de points

Une correspondance est ici définie comme une paire de points. Le premier point composant une paire provient de la trajectoire de gauche et l'autre de la trajectoire de droite. Pour réduire le nombre de correspondances possible, l'hypothèse de la synchronisation des séquences vidéo a été faite. Ainsi, chacun des points de la trajectoire de gauche étudiée est apparié avec celui ayant le même numéro de trame dans la trajectoire de droite étudiée. Si des points dans une des deux trajectoires n'ont pas de correspondant (point ayant le même numéro de trame) dans l'autre trajectoire, ils sont ignorés. En effet, il se peut que le « blob » ne soit tout simplement plus dans le champ

de vision d'une des deux caméras. Cette situation n'a pas à être pénalisée. En résumé, une correspondance est composée d'un point dans la trajectoire de gauche et du point de la trajectoire de droite qui a été observée à la même trame.

### **2.3.2.3 Troisième étape : estimation de type RANSAC**

Cette étape est le cœur de l'algorithme et contient trois sous-étapes. Ces sous-étapes sont répétées un certain nombre de fois. Ce nombre d'itérations varie selon une procédure qui est expliquée à la section 2.3.2.5. Aussi, un processus de sélection est appliqué entre chaque itération de cette étape. Après chaque itération, si la nouvelle matrice de transformation a un nombre de correspondances participantes plus grand que précédemment, cette nouvelle matrice et ses correspondances participantes sont sauvegardées. En cas d'égalité, la matrice avec les correspondances participantes ayant le plus petit écart-type est choisie.

La première des sous-étapes de l'étape 3 est de calculer une estimation de la matrice de transformation affine. Pour y arriver, il faut donc piger aléatoirement 3 correspondances (3 paires de points) puisqu'il s'agit du nombre minimal de correspondances nécessaire pour calculer une matrice de transformation affine. Il ne faut pas s'inquiéter d'un mauvais choix aléatoire car dans ce cas, le pointage de la matrice ne devrait pas être bon et cette combinaison de correspondances ne sera pas retenue.

La deuxième sous-étape est de calculer la distance entre cette estimation obtenue et l'observation. Ceci signifie que la transformation est appliquée à tous les points qui composent la trajectoire de gauche étudiée. Ensuite, les distances entre ces nouvelles coordonnées et les coordonnées des points qui composent la trajectoire de droite sont calculées. La mesure de distance utilisée est simplement la distance euclidienne.

La troisième sous-étape consiste à trouver les correspondances participantes à partir de ces distances calculées. Une correspondance est participante si la distance qui lui est associée est plus petite qu'un certain seuil. Après plusieurs essais, la valeur optimale de ce seuil pour l'algorithme développé a été établie à 5. C'est donc dire qu'une correspondance est participante si la distance entre le point de la trajectoire de droite et

celui transformé de la trajectoire de gauche est à moins de 5 pixels. Le nombre de correspondances participantes est conservé comme mesure de qualité de la matrice de transformation affine. Les correspondances participantes sont aussi conservées et viennent s'ajouter aux 3 correspondances pigées aléatoirement à la première sous-étape puisque ces dernières sont nécessairement aussi des correspondances participantes. C'est aussi à cette sous-étape que l'écart-type des distances des correspondances participantes est calculé pour surmonter les cas d'égalités comme mentionné précédemment.

#### **2.3.2.4 Quatrième étape : raffinement de la matrice de transformation**

Cette dernière étape consiste à utiliser toutes les correspondances participantes et pas seulement 3 pour calculer plus précisément la matrice de transformation affine. Étant donné qu'un seuil fixe est utilisé pour définir une correspondance participante, il se peut que certaines correspondances, ayant été rejetées initialement, se qualifient comme correspondance participante après cette ré-estimation de la matrice. C'est pourquoi les deux sous-étapes qui composent cette étape doivent être répétées jusqu'à ce que le nombre de correspondances participantes soit stable. Cette stabilité est atteinte lorsque 4 itérations consécutives de ces sous-étapes ont donné le même nombre de correspondances participantes. Ce seuil de 4 itérations est tiré de la méthode RANSAC suggéré par Hartley & Zisserman (2002). On utilise le seuil de 4 itérations plutôt que 2 pour s'assurer que l'augmentation de la précision à chaque itération de la matrice par la méthode de Levenberg-Marquardt ne fasse plus apparaître de correspondances participantes. Voici maintenant la description de ces sous-étapes.

La première sous-étape est de calculer la matrice de transformation affine en utilisant toutes les correspondances participantes. Il faut donc utiliser une méthode qui trouvera une solution tenant compte de plus de 3 correspondances. Dans Hartley & Zisserman (2002), il est suggéré d'utiliser la méthode de Levenberg-Marquardt (LM). Cette méthode est une méthode numérique très utilisée pour les problèmes d'ajustement à des courbes non linéaires. La situation présente est donc un cas typique puisque les différentes correspondances participantes représentent des points qui forment une

courbe dont la nature est inconnue et qui a donc de grandes chances d'être non linéaire. Pour ce qui est de l'implantation de cette méthode, une librairie disponible en ligne créée par un institut de recherche en informatique grec a été utilisée (Lourakis, 2007). Il suffit de fournir à la méthode une première estimation pour commencer sa recherche, la fonction à minimiser, le jacobien de cette fonction, les points à insérer dans ces fonctions ainsi que les points de la solution idéale. Pour la première estimation, la dernière matrice de transformation trouvée est fournie à l'algorithme. Pour la fonction à minimiser, l'équation qui donne les points d'arrivée à partir des points de départ est utilisée. En d'autres mots, il faut fournir à l'algorithme les différentes équations pour chacune des coordonnées des points de la trajectoire de gauche auxquels l'équation 2.3 a été appliquée. Le résultat est le suivant :

$$X'_x = X_x a_{11} + X_y a_{12} + t_x \quad (2.6)$$

$$X'_y = X_x a_{21} + X_y a_{22} + t_y \quad (2.7)$$

Dans les dernières équations,  $X'_x$  et  $X'_y$  représentent les deux composantes du point de la trajectoire de gauche qui a été transformé par la matrice. De la même manière,  $X_x$  et  $X_y$  sont les composantes d'un point de la trajectoire de gauche avant la transformation. Finalement, les variables  $a_{ij}$ ,  $t_x$  et  $t_y$  sont les composantes de la matrice de transformation affine de l'équation 2.5. Pour ce qui est du jacobien, il doit être calculé en fonction de la matrice de transformation pour chacune des 2 équations précédentes ce qui donne :

$$J_{X_{affine}} = \begin{bmatrix} X_x & X_y & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.8)$$

$$J_{Y_{affine}} = \begin{bmatrix} 0 & 0 & 0 \\ X_x & X_y & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (2.9)$$

Dans ces équations,  $X_x$  et  $X_y$  sont encore les composantes d'un point de la trajectoire de gauche avant la transformation. Aussi,  $J_{xaffine}$  et  $J_{yaffine}$  représentent respectivement la matrice jacobienne pour la composante en x et y du point recherché. Après avoir défini ces équations, il reste à fournir les points participants de la trajectoire de gauche et ceux correspondants de la trajectoire de droite pour que l'algorithme retourne la nouvelle matrice de transformation tenant compte de toutes les correspondances participantes.

La deuxième sous-étape consiste à repérer de nouveau les correspondances participantes et à les compter. Cette sous-étape s'effectue de la même façon que pour les deuxième et troisième sous-étapes de la section 2.3.2.3. À la fin, les correspondances participantes sont sauvegardées ainsi que le nombre de ces dernières.

Pour finir cette section sur le calcul de la matrice de transformation affine, il faut expliquer comment le nombre d'itérations nécessaires avec l'algorithme RANSAC utilisé a été obtenu. C'est le sujet de la prochaine et dernière sous-section.

### 2.3.2.5 Calcul du nombre d'itérations dans un algorithme RANSAC

Dans beaucoup d'algorithmes RANSAC comme ceux de la section 2.3.2 et 2.4, le nombre de répétitions sont proportionnels au pourcentage de données participantes trouvées dans la solution. En d'autres mots, si le pourcentage de données participantes est assez élevé par rapport au nombre d'itérations effectué, les répétitions sont terminées. La formule qui permet de faire cette comparaison est la suivante tirée de la section 4.7 de (Hartley & Zisserman, 2002) :

$$N = \frac{\log(1-p)}{\log(1 - (1-\varepsilon)^S)} \quad (2.10)$$

Dans cette formule,  $N$  représente le nombre d'itérations qui doivent être complété pour considérer (avec une probabilité égale à  $p$ ) qu'au moins une des itérations composées de  $S$  points est exempte de données aberrantes. Il faut donc aussi tenir compte de la probabilité de tomber sur une donnée aberrante, c'est ce que  $\varepsilon$  représente. Pour l'algorithme développé :  $p$  est égal à 0.99,  $S$  à 3 car il faut au moins trois points pour

calculer une matrice de transformation affine, mais  $\varepsilon$  n'est pas connue précisément. Il faut donc l'estimer. Pour y parvenir, Hartley & Zisserman (2002) propose d'utiliser l'équation suivante :

$$\varepsilon = 1 - \frac{nbP}{nbT} \quad (2.11)$$

Dans cette équation,  $\varepsilon$  représente encore une fois la probabilité de piger une donnée aberrante. Aussi,  $nbP$  est le nombre de données participantes et  $nbT$  le nombre total de données. Dans la solution qui a été implantée,  $nbP$  est égal au nombre de correspondances participantes (étape 3c de l'algorithme RANSAC de la section 2.3.2) de la meilleure solution trouvée actuellement et  $nbT$  est égal au nombre total de correspondances possible. Avec ces définitions établies, il est possible de résumer les étapes permettant d'ajuster le nombre d'itérations comme suit :

- Définir  $N$  à un très grand nombre et le compteur d'itération à 0;
- Tant que  $N$  est plus grand que le compteur d'itération, répétez :
  - ✓ Choisir 3 correspondances et trouver le nombre de correspondances participantes ( $nbP$ ) qui s'y rattachent;
  - ✓ Définir  $\varepsilon$  avec l'équation 2.11;
  - ✓ Définir  $N$  avec l'équation 2.10;
  - ✓ Ajouter 1 au compteur d'itération;
- Les répétitions sont terminées.

Pour terminer, rappelons que la même procédure est utilisée pour connaître le nombre d'itérations nécessaire à l'algorithme RANSAC utilisé pour appairer les trajectoires (section 2.4). Les seules différences sont les valeurs données aux variables suivantes :  $S$  égal à 1 car une seule paire de trajectoires est nécessaire pour calculer une matrice de transformation,  $nbP$  est égal au nombre de trajectoires participantes comme défini à la section 2.4.3 et  $nbT$  au nombre total de paires de trajectoires.

## 2.4 Appariement des trajectoires

Pour calculer une première estimation de la matrice de transformation affine, il faut au moins 3 points dans l'image de gauche et 3 points dans celle de droite comme mentionné précédemment (section 2.3.2). L'idée de base est d'obtenir ces points à partir des trajectoires des objets en mouvement observés dans les séquences vidéo. Cependant, aucune présomption n'est faite sur les correspondances de ces trajectoires. C'est donc le rôle de l'appariement de trajectoires de trouver les correspondances des trajectoires. Pour y arriver, un algorithme de type RANSAC est utilisé tout comme pour le calcul de la matrice de transformation affine (section 2.3.2). En effet, on y retrouve les différentes parties qui forment un algorithme RANSAC (réf. Chapitre 1) :

1. Choix d'une paire de trajectoires (pige de donnée);
2. Calcul de la matrice de transformation affine de cette paire (calcul du modèle);
3. Calcul du pointage de cette paire (évaluer la ressemblance au modèle);
4. Recherche des paires participantes (évaluer la ressemblance au modèle);
5. Évaluation du nombre de bons appariements trouvés (choix ou rejet de la donnée pigée);
6. Si tous les bons appariements sont trouvés, passer à l'étape 7. Sinon, retour à l'étape 1.
7. Recalculer la matrice de transformation affine avec toutes les paires participantes.

La plupart des étapes de l'algorithme d'appariement des trajectoires constituent les subdivisions de cette section. Les seuls qui ne le sont pas sont les étapes 2, 6 et 7. Les étapes 2 et 7 ont été traitées en détail dans la section précédente puisqu'il s'agit de calculer la matrice de transformation affine. Pour l'étape 6, elle fonctionne de pair avec l'étape 5 et sera donc expliquée dans cette section. Voyons donc plus en détail ces étapes.

### 2.4.1 Choix d'une paire de trajectoires

Avant d'expliquer le choix d'une paire de trajectoires, il faut d'abord définir ce qu'est une paire de trajectoires. Une paire est constituée d'une trajectoire tirée de la séquence vidéo infrarouge et d'une autre trajectoire tirée de la séquence vidéo couleur. Pour mieux illustrer ces paires, la figure 2.9 montre une image infrarouge et une image couleur ainsi que les trajectoires qu'on y retrouve. Il est à remarquer que ces trajectoires ont été nommées  $L1$ ,  $L2$  et  $L3$  pour la séquence vidéo de gauche alors qu'elles ont été nommées  $R1$  et  $R2$  pour la séquence vidéo de droite.

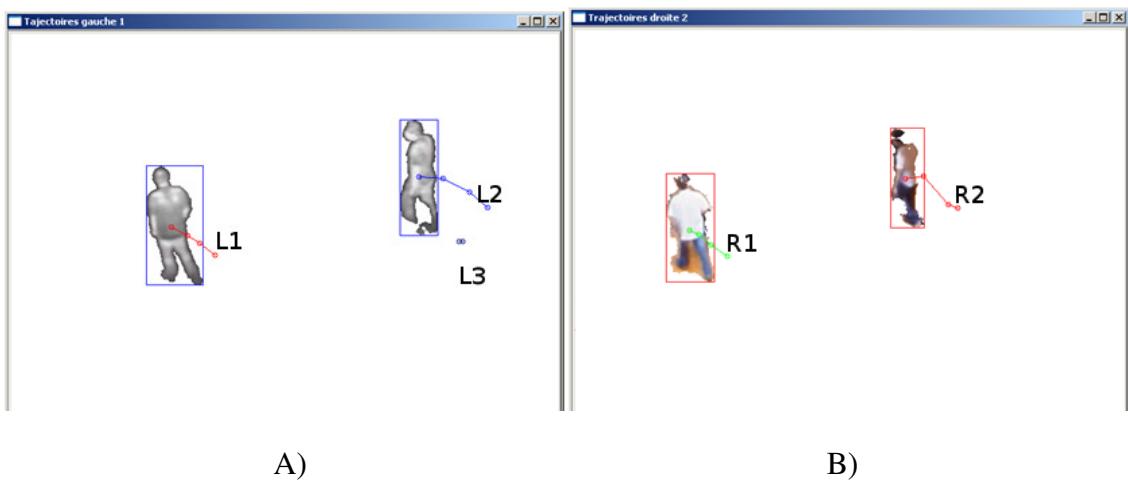


Figure 2.9 Images infrarouge (A) et couleur (B) ainsi que leurs trajectoires

Ainsi, les différentes paires qui peuvent être formées pour ces images sont  $L1-R1$ ,  $L1-R2$ ,  $L2-R1$ ,  $L2-R2$ ,  $L3-R1$  et  $L3-R2$ . Il est à noter que ces paires sont composées de trajectoires qui n'ont pas nécessairement le même nombre de points. Aussi, cette image ne donne aucune indication sur le temps (numéro de trame) où les points ont été observés. Ces données sont importantes puisqu'il serait illogique de comparer des trajectoires observées à des moments différents lorsque l'hypothèse de synchronisation des séquences vidéo est faite. Donc, chaque fois qu'une paire est construite, les trajectoires sont coupées pour n'inclure que les points qui ont été observés au même temps pour les deux trajectoires constituant la paire. Aussi, certaines paires sont rejetées si le nombre de points qui respectent la condition précédente n'est pas égal ou supérieur à 3. Comme mentionné précédemment, si le nombre de points est inférieur à 3, il est

impossible de calculer la matrice de transformation affine. Finalement, le choix de la paire de départ s'effectue de manière aléatoire parmi toutes les paires de points qui respectent les conditions présentées. Ce choix aléatoire s'effectue sans remises puisque les étapes subséquentes sont déterministes pour une paire donnée. Une fois ce choix effectué, la matrice de transformation affine est calculée de la façon décrite dans la section 2.3.2 avec les points qui forment les trajectoires.

#### **2.4.2 Calcul du pointage d'une paire de trajectoires**

Une fois la première estimation de la matrice de transformation affine obtenue, l'algorithme RANSAC dit qu'il faut évaluer la qualité de cette estimation. C'est dans cette partie que se situe la principale contribution de ce projet. En effet, Caspi et al. (2006) proposent de calculer la somme de l'erreur euclidienne pour chaque autre paire de trajectoires et de l'attribuer comme pointage à une matrice donnée. Cette démarche donne des résultats intéressants lorsque le nombre de trajectoires est grand et qu'elle est appliquée hors ligne comme pour les résultats vus dans Caspi et al. (2006). Cependant, les résultats sont beaucoup moins convaincants pour des séquences avec un petit nombre de trajectoires et lorsque le nombre de points est petit comme démontré au CHAPITRE 3.

Pour pallier à ces problèmes, les approches de Bhanu & Ju (2007) qui s'applique pour une seule image et la stratégie tenant compte du temps par les trajectoires de Caspi et al. (2006) ont été combinées. Ainsi, une image combinant jusqu'à 5 trames d'avant-plan observées à des moments différents est créée. Les détails des mécanismes de construction de cette image, qui est nommée image de pointage, sont donnés un peu plus loin (section 2.4.2.1). Aussi, ce ne sont pas ces images de pointages qui sont utilisées directement comme mesure de qualité de la matrice trouvée, mais plutôt le pourcentage des différences de recouvrement de ces images. La figure 2.10 montre les images de pointage des deux caméras ainsi que l'image résultante qui montre les différences et les similitudes de recouvrement entre les deux. Il faut noter que ces images de pointages ne

sont constituées que d'une seule trame comme c'est généralement le cas au début des séquences vidéo utilisées.

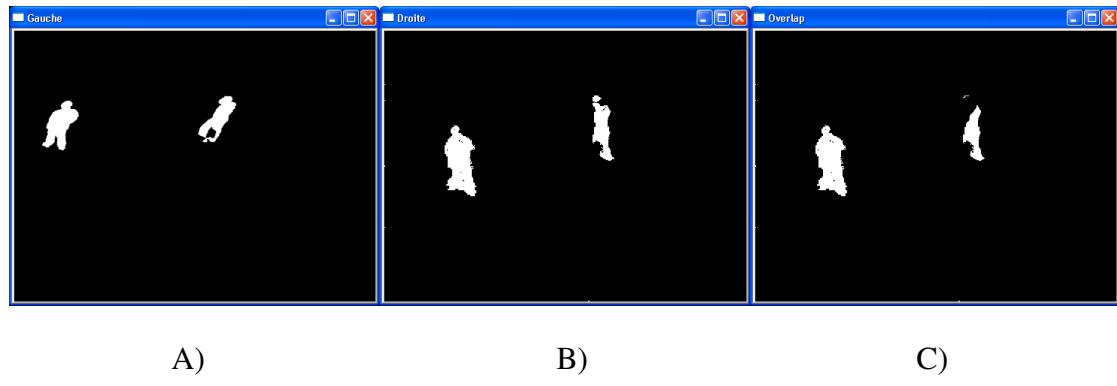


Figure 2.10 Images de pointage infrarouge (A) et couleur (B) ainsi que l'image des différences de recouvrement (C)

Il faut rappeler ici que l'image de pointage de gauche affichée a été transformée par la matrice de transformation selon l'équation 2.3. Cette image est alors inversée logiquement (les pixels blancs deviennent noirs et les pixels noirs deviennent blancs) et mise en recouvrement avec l'image de droite avec l'opérateur logique « et » (AND). Cette combinaison d'une inversion logique suivit d'un « et » logique est appelée une différence à sens unique. On pourrait la résumer par l'expression l'image de droite sans l'image de gauche. Le choix de l'utilisation de cette opération est justifié par la nature d'un recalage. En effet, un recalage consiste essentiellement à transformer une image en une autre. Dans le cas qui nous intéresse il s'agit de transformer l'image de gauche en l'image de droite. C'est donc ce qui est tenté par l'opération de différence à sens unique, retirer l'image de gauche de l'image de droite. On peut donc dire que l'image de droite est vue comme l'image de référence. Il faut donc seulement étudier les différences entre l'image de gauche modifiée et l'image de droite. Par contre, les différences de l'image de droite par rapport à l'image de gauche modifiée ne doivent pas être prise en compte car on ne veut pas calculer les erreurs en double. C'est ce que l'opération de différence à sens unique nous permet de faire. L'image ainsi obtenue contient seulement les différences de la nouvelle image (l'image de gauche) par rapport à l'image de comparaison qui est l'image de droite. Par conséquent, si l'image de gauche était

totalement décalée par rapport à l'image de droite, l'image de recouvrement serait identique à celle de droite. Dans le cas de la figure 2.10, une partie d'une des deux personnes de l'image de gauche est en recouvrement avec une des deux personnes de l'image de droite. Il en résulte que l'image de recouvrement contient l'image de droite moins les parties qui sont en recouvrement. Cette méthode de comparaison a cependant un petit inconvénient, si l'image transformée de gauche est très grande et contient donc un grand nombre de pixels blancs, les différences de recouvrement seront diminuées alors que les deux images sont très différentes. Ce phénomène est visible dans la figure suivante :

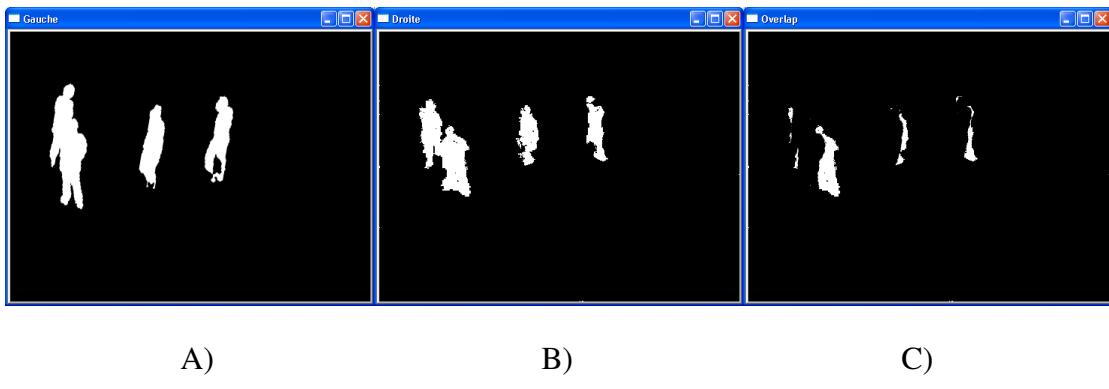


Figure 2.11 Images de pointage infrarouge (A) et couleur (B) ainsi que l'image des différences de recouvrement (C) pour une situation où le grossissement de l'image infrarouge pose problème

Dans ces dernières figures, on voit que même si les personnes sont plus grosses dans l'image infrarouge, leur recouvrement avec l'image de droite donne une erreur relativement petite. Cette erreur ne devrait pas être petite puisque l'homothétie entre l'image infrarouge et l'image couleur est trop prononcée. Pour éviter ce phénomène, un critère de sélection a été ajouté. Ainsi, si le pourcentage de pixels contenu dans l'image de gauche transformée est supérieur à 30 % à celui de l'image de droite, la transformation est refusée en lui attribuant un pointage d'erreur de 100 %.

Pour ce qui est de l'image des différences obtenue à la figure 2.10, elle correspond à un pointage de 92,46 % (ajusté selon l'image utilisée). Ce pointage a été obtenu selon l'équation suivante :

$$E = \frac{Nb_{diff}}{Nb_{droit}} \times 100 \quad (2.12)$$

Dans cette équation,  $E$  correspond à l'erreur en pourcentage et  $Nb_{diff}$  au nombre de pixels blancs dans l'image des différences de recouvrement. Pour ce qui est de  $Nb_{droit}$ , il représente le nombre de pixels blanc dans l'image de pointage de la séquence vidéo couleur qui est celle de comparaison. Le choix d'utiliser un pourcentage plutôt que la valeur absolue est justifié par la nécessité de comparer la valeur d'une matrice de transformation à d'autres obtenue dans le temps avec des tailles de « blobs » différentes. Cette nécessité de comparer les matrices dans le temps découle du fait qu'une architecture temps réel fait en sorte qu'une nouvelle matrice de transformation est calculée à chaque trame. Cependant, des erreurs dans les trajectoires et dans la soustraction d'arrière-plan peuvent entraîner une détérioration de la qualité du recalage issu de cette matrice de transformation. Il est donc erroné de conclure qu'une matrice de transformation obtenue plus tard dans la séquence vidéo sera nécessairement de meilleure qualité. C'est donc pour s'assurer de cette qualité qu'une méthode de comparaison, basée sur un pointage, des matrices est nécessaire. Les implications et le fonctionnement de l'optique temps réel de la méthode de recalage automatique créée sont décrits plus en détail à la section 2.5. Avant de passer à la prochaine étape qui consiste à trouver les paires de trajectoires participantes, il faut maintenant expliquer comment ces images de pointage sont construites.

#### **2.4.2.1 Construction de l'image de pointage**

Comme mentionnée précédemment, l'image de pointage vient de l'idée de combiner les idées d'images instantanées de Bhanu & Ju (2007) et des trajectoires de Caspi et al. (2006). L'idée est simple, au lieu d'utiliser seulement un point par « blob » qui compose les trajectoires, pourquoi ne pas utiliser les « blobs » entiers de toutes les trames précédentes. Le problème est que si l'on utilise tous les « blobs » des trames précédentes superposées dans une image (au lieu des points des « blobs » formant les trajectoires) l'image serait saturée. Ceci aurait comme résultat que n'importe quelle matrice de

transformation pourrait avoir un pointage élevé. Pour éviter ce problème, il faut limiter le nombre de « blobs » choisis. Dans le programme développé, un choix de 5 trames maximum a été fait. Ceci signifie que l'image pointage qui sera utilisée sera composée des « blobs » de chacune des trajectoires observées à au plus 5 moments différents dans la séquence vidéo. Il faut aussi mentionner qu'une image de pointage par caméra sera créée. Il y aura donc 2 images de pointage utilisées par le programme, une pour l'infrarouge et l'autre pour le spectre visible.

Le prochain point à expliquer est comment ces 5 trames sont choisies. D'abord, il a été décidé de toujours porter une attention particulière au moment présent. Donc, une de ces trames sera toujours la trame qui est actuellement traitée. Il reste donc à choisir 4 trames. Comme la méthode de recalage automatique développée suit une optique temps réel, il n'y aura pas toujours 4 trames disponibles puisqu'il se peut que le moment actuellement traité soit composé de moins de 4 trames. Une première façon de faire qui serait intuitive serait de conserver les 4 dernières trames en plus de trame présentement analysée. Pour savoir si ce choix est bon, il faut avoir expérimenté le recalage manuel et observé ce qui influence sa qualité.

Lors des différents essais effectués pour bien évaluer les performances de la méthode de recalage développée, la nécessité d'utiliser des valeurs témoins basées sur un recalage manuel a été mise en évidence (section 3.1.2). Durant ces essais, il a été constaté que l'emplacement des éléments témoins choisis pour évaluer la performance d'un recalage était très important. En effet, même si l'hypothèse a été faite que tous les objets d'avant-plan se situent dans le même plan, il reste que dans les fait, cette hypothèse n'est pas toujours totalement respectée. Ainsi, si tous les éléments témoins se situent dans le même secteur de l'image, le recalage sera peut-être très bien pour les objets se situant dans ce secteur mais pas pour les autres étant donné leurs légères différences de distance avec les caméras. Cette question est abondamment traitée un peu plus loin (section 3.1.2). Ainsi donc, la conclusion a été faite qu'un recalage peut être mieux évalué lorsque les éléments utilisés pour faire cette évaluation sont distribués uniformément dans l'image. Cette même conclusion est utilisée pour le choix des

trames. Ainsi, si 5 trames consécutives sont utilisées, le risque est très grand que les objets d'avant-plan formant l'image de pointage se situe dans la même région. Il faut donc trouver un moyen de garder les trames qui permet d'avoir une image de pointage composée de « blobs » disposés uniformément dans l'image.

Pour y arriver, un système d'acceptation ou de rejet d'image a été implanté. Il n'est appliqué que si l'intervalle de temps donné par l'usager est respecté. Ainsi, si l'usager demande un intervalle de 2 secondes et que le rythme d'acquisition des trames est fixé à 7,5 trames par secondes, alors le système d'acceptation d'ajout de trames ne s'effectuera qu'à toutes les 15 trames. Le système d'acceptation consiste essentiellement à comparer la dernière image de pointage calculée avec une nouvelle image de pointage laquelle est formée de :

- Si l'image de pointage précédente était composée de moins de 4 trames, alors la nouvelle image de pointage est simplement constituée de la juxtaposition de l'image de pointage précédente et de la trame courante ou;
- Si l'image de pointage était composée de 4 trames, alors la nouvelle image de pointage sera formée de la nouvelle trame juxtaposée aux 3 autres trames les plus récentes.

Il est important de rappeler que cette comparaison est faite pour la caméra infrarouge et la caméra couleur. Pour s'assurer que les 2 images de pointages des 2 spectres continuent à être consistantes afin d'être comparables, il faut que ces comparaisons soient effectuées en tenant en compte l'une de l'autre. Donc, les images de pointage ne seront modifiées que si la modification est positive pour les 2 spectres.

Pour connaître si une modification est positive, on sépare l'image de pointage analysée en 4 cadrants. Ensuite, on calcule le nombre de pixels d'avant-plan dans chacun des cadrants. L'hypothèse directrice derrière cette vérification est qu'il y a amélioration si un plus grand nombre de pixels d'avant-plan se trouve dans un cadran. Par contre, pour diminuer les erreurs dues au non respect de l'hypothèse des objets d'avant-plan situés dans le même plan, il faut que cette augmentation soit observée dans le plus grand

nombre de cadrants possible. On acceptera la modification que si le nombre de pixels d'avant-plan est égal ou augmente dans au moins 3 cadrants. Dans les cas où il y aurait égalité (2 cadrants positifs et 2 cadrants négatifs), on ne refusera pas automatiquement la modification. Si la modification est avantageuse dans un spectre (3 cadrants positifs) et nulle dans l'autre (2 cadrants positifs et 2 cadrants négatifs), alors elle sera acceptée. Si les 2 spectres sont nuls, alors la modification est refusée. Dans certains cas, une nouvelle image de pointage vide se trouvait à donner un plus grand nombre de cadrants positifs qu'une image de pointage précédente non vide (par exemple, une image de pointage avec des pixels d'avant-plan seulement dans un seul cadran). Pour éviter cette erreur, une condition a été ajoutée à l'algorithme qui refuse la modification automatiquement lorsqu'on détecte une nouvelle image de pointage vide avec une différence du nombre de pixels d'avant-plan non nul.

Pour mieux comprendre tous les détails de l'algorithme utilisé pour accepter ou refuser une modification à l'image de pointage, voici un diagramme qui le résume :

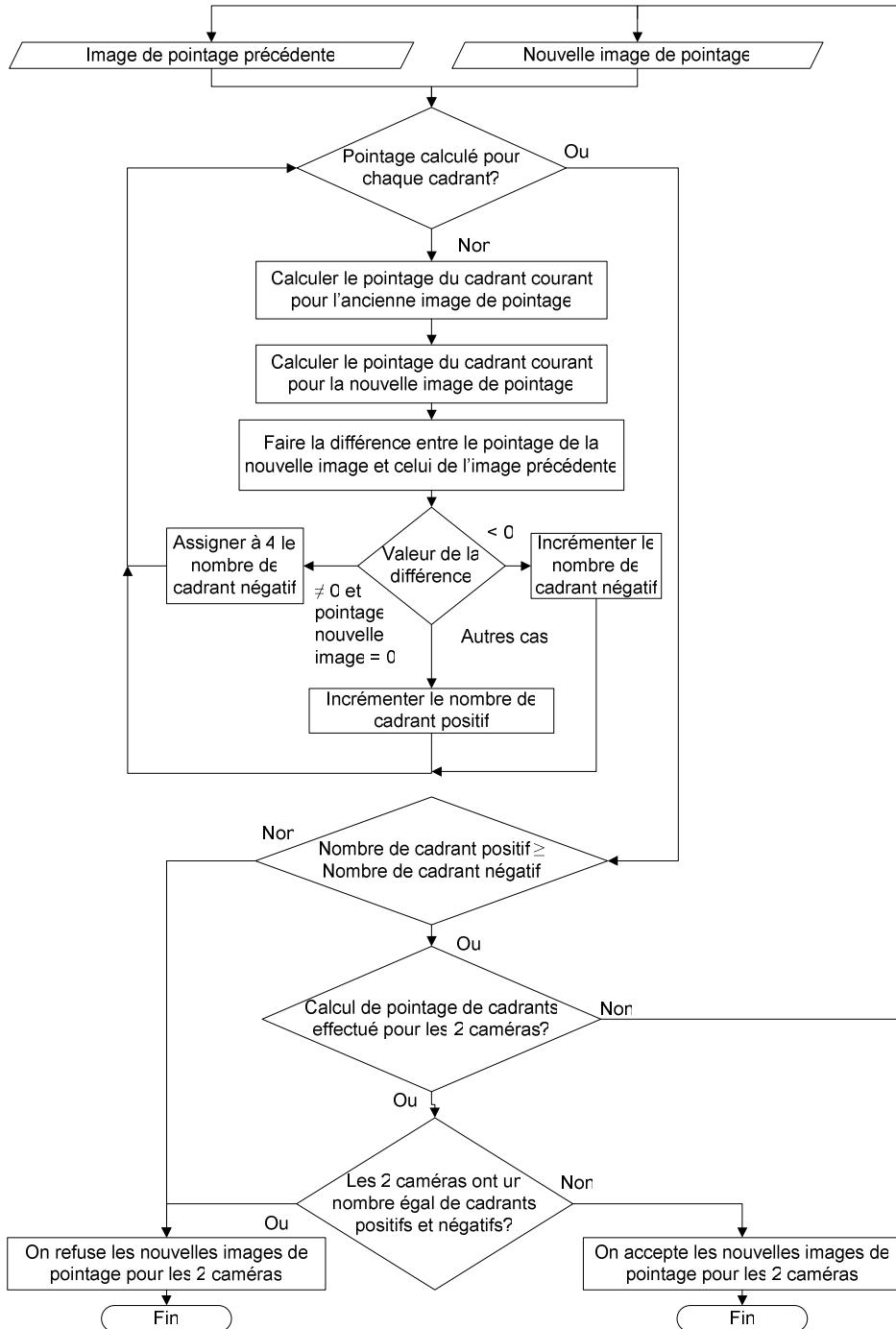


Figure 2.12 Schéma de l'algorithme d'acceptation de modification des images de pointage

### 2.4.3 Recherche des paires de trajectoires participantes

Il a été vu précédemment que la méthode employée pour calculer la matrice de transformation affine peut supporter une infinité de points et que l'estimation de cette matrice tend à s'améliorer en fonction du nombre de points utilisés pour la calculer. De ces observations, il est logique de déduire que d'ajouter les points qui composent les paires de trajectoires qui sont correctement appariées par cette première estimation devrait améliorer la qualité du recalage obtenu. Il faut donc trouver comment identifier les paires de trajectoires qui sont bien appariées. Ces paires de trajectoires seront nommées paires de trajectoires participantes. La méthode pour y arriver est de calculer la somme des erreurs euclidiennes pour chaque paire de trajectoires et d'utiliser un seuil pour définir cette paire comme participante ou non (Caspi et al., 2006). Les désavantages de l'utilisation de seuils sont bien connus. Les seuils sont intimement liés aux séquences vidéo utilisées et doivent souvent être réévalués pour chaque nouvelle séquence vidéo. Le pointage utilisé dans la méthode de recalage automatique présentée dans ce projet permet de contourner ces problèmes.

Afin de connaître si une paire de trajectoires est participante, il suffit d'ajouter les points qui la forment à ceux déjà utilisés lors du calcul de la première estimation de la matrice de transformation et de recalculer cette matrice. Par la même occasion, un nouveau pointage est attribué à cette nouvelle matrice. Si ce nouveau pointage est inférieur à celui obtenu lors de la première estimation, cette paire de trajectoires est une paire de trajectoires participante. Si ce n'est pas le cas, cette paire est une donnée aberrante. Ces opérations sont effectuées pour chaque paire de trajectoires. Toutes les paires de trajectoires participantes sont mémorisées, ainsi que les pointages qui leur sont associés, et sont utilisées lors du calcul final de la matrice de transformation (étape 7 de l'algorithme de la section 2.4).

### 2.4.4 Évaluation du nombre de bons appariements trouvé

La méthode utilisée pour trouver les paires participantes ne contient pas de seuils, mais certaines erreurs peuvent se glisser dans le choix des paires. En effet, il est possible

qu'une paire de trajectoires non participante améliore quand même le pointage d'une matrice. Ce phénomène peut paraître de moindre importance, mais il cause de graves ennuis lors du choix de recalculer une nouvelle matrice de transformation et toutes les étapes subséquentes ou non dans l'algorithme RANSAC (étape 6 de l'algorithme de la section 2.4). Il a été expliqué précédemment (section 2.3.2.5) comment calculer statistiquement le nombre nécessaire d'itérations dans un algorithme RANSAC. Cependant, le cas d'appariement des trajectoires est quelque peu différent du calcul de la matrice de transformation. En effet, lors du calcul de la matrice de transformation, le pourcentage de données aberrantes dans le bassin de données était inconnu. Il fallait donc l'estimer à chaque itération. Pour l'appariement des trajectoires, ce pourcentage peut être calculé à partir du nombre de trajectoires contraint. En effet, ce nombre sera donné par la séquence vidéo qui contient le plus petit nombre de trajectoires. On appelle cette séquence vidéo, la séquence vidéo contraignante. Par exemple, dans l'exemple de la figure 2.9, la séquence vidéo contraignante est celle de droite car c'est elle qui contient le moins de trajectoires (3 pour celle de gauche et 2 pour celle de droite). Il y a donc deux paires de trajectoires participantes à trouver (sur les 5 paires de trajectoires possibles). La première étant choisie lors de la pige initiale, l'algorithme RANSAC se termine lorsqu'une autre paire participante est trouvée puisque le pourcentage de données aberrantes contenues dans cette estimation serait nul. Par contre, si aucune trajectoire participante n'était trouvée, le pourcentage de données aberrantes de l'estimation serait de 50 % et l'algorithme se poursuivrait donc.

Dans cet exemple, les problèmes surviendraient si deux paires ou plus, en plus de la paire pigée, étaient classées comme paires participantes. Dans ce cas, il y aurait clairement un ou des intrus. Pour les trouver et les retirer, le pointage de chacune des paires participantes est analysé. Les  $n$  paires participantes ayant les plus petits pointages sont conservées. Ce nombre  $n$  est égal au nombre de trajectoires de la séquence vidéo contraignante moins un (pour tenir compte de la paire pigée initialement). Une fois toutes les paires participantes trouvées, l'algorithme se termine en recalculant une dernière fois la matrice de transformation affine en utilisant tous les points de toutes les

paires participantes (étape 7 de l'algorithme de la section 2.4). Un pointage final est aussi calculé pour être utilisé lors du choix en temps réel.

## 2.5 Choix à chaque trame de la matrice de transformation

Comme mentionné précédemment, la méthode de recalage développée opère trame par trame. En d'autres termes, elle opère dans une optique de temps réel. Ceci signifie qu'une matrice de transformation affine est calculée pour chaque trame des séquences vidéo synchronisées utilisées. Cependant, il n'existe qu'une seule vraie matrice de transformation par montage. Ceci signifie que si les positions des caméras l'une par rapport à l'autre ne changent pas, il n'existe qu'une seule matrice de transformation qui permet de passer d'une à l'autre. Mentionnons ici que ce constat est vrai seulement si la restriction sur la position des objets de l'avant-plan par rapport à la caméra est respectée (section 2.3.1). Le défi est donc de calculer plusieurs matrices en espérant converger vers la vraie matrice de transformation. Logiquement, il serait normal de constater une amélioration du recalage à mesure que le nombre de points et de trajectoires augmente comme expliqué dans la section 2.3.2. Cependant, les nombreuses erreurs qui s'accumulent lors des opérations de prétraitements ne permettent pas de garantir cette amélioration. C'est en réponse à cette constatation que le calcul à chaque trame d'une matrice de transformation peut être mis à profit. Il faut choisir parmi la multitude de matrices de transformation calculées celle qui est la plus proche de la matrice de transformation parfaite. C'est dans ce but que le pointage basé sur les différences relatives du recouvrement a été introduit à la section 2.4.2. Cependant, il est inutile de sauvegarder chacune des matrices calculées et leur pointage. Il suffit plutôt de comparer, après chaque calcul d'une nouvelle matrice, son pointage avec le meilleur obtenu jusqu'à maintenant. Si ce nouveau pointage est inférieur au meilleur obtenu, la meilleure matrice est remplacée par cette nouvelle. Dans le cas contraire, l'ancienne matrice est conservée. Ainsi, le recalage devrait converger vers le meilleur résultat possible.

Il faut cependant ajouter une précision à ce processus. Comme les images qui sont utilisées pour calculer le pointage varient avec le temps (section 2.4.2.1), les pointages

obtenus au début des séquences vidéo risquent d'être beaucoup plus faibles que ceux obtenus en fin de vidéo. En effet, ces derniers pointages sont obtenus en utilisant des images qui combinent 5 trames. En comparaison, les images de pointages du début des séquences vidéo n'ont généralement qu'une ou deux trames. Pour contourner ce problème, un ajout a été fait à la méthode de comparaison des matrices à chaque trame. Au lieu de regarder directement le pointage précédent avec le pointage courant, il suffit de recalculer le pointage de l'ancienne matrice de transformation avec les nouvelles images de pointage obtenues à chaque trame. Ainsi, il est possible de conclure qu'un pointage plus faible de la nouvelle matrice signifie bel et bien une amélioration du recalage. D'ailleurs, le prochain chapitre (CHAPITRE 3) montre tous les détails des performances qui peuvent être atteintes avec l'approche de recalage automatique qui a été exposée dans ce chapitre.

## CHAPITRE 3 RÉSULTATS ET DISCUSSION

Ce chapitre portera sur les performances de la méthode de recalage automatique développée. Entre autres, elle sera comparée à celle développée par Caspi et al. (2006) grâce à l'implantation de cette méthode dans le code développé. La méthodologie utilisée pour mesurer ses performances sera aussi détaillée dans ce chapitre.

### 3.1 Méthodologie d’expérimentation

Avant de parler des résultats des tests et de l’évaluation des performances qui en découlent, il est important de bien expliquer comment ces tests ont été menés. En premier lieu, il faut rappeler que les algorithmes RANSAC utilisés sont non déterministes puisqu’ils pigent au hasard une trajectoire ou des points selon le cas. Par conséquent, deux exécutions successives ne donneront pas nécessairement les mêmes résultats. C’est pour cette raison que chaque scénario testé a été exécuté 30 fois. Il est ainsi possible de conclure que les résultats sont valides statistiquement comme le mentionne Ostle (1996) à la page 185.

Il faut aussi mentionner que les résultats de la nouvelle méthode de recalage automatique seront toujours comparés à la méthode de base ((Caspi et al., 2006)) ainsi qu’à un recalage manuel. Par conséquent, les mêmes opérations doivent être appliquées par chacune des méthodes pour que la comparaison entre chacune soit valide. En résumé, chacune des méthodes doit effectuer les opérations suivantes :

1. Calculer la matrice de transformation pour la trame courante;
2. Avec cette matrice de transformation, calculer le pourcentage de recouvrement sur les éléments témoins pour la trame courante;

La nature et la constructions de ces éléments témoins sont présentés plus loin dans ce chapitre (3.1.2). Mentionnons simplement que ces éléments témoins sont choisis en fonction du fait qu’ils se trouvent dans chacune des trames d’une séquence vidéo donnée. Ceci permet non seulement de pouvoir comparer les 3 méthodes les une par rapport aux autres, mais aussi de pouvoir évaluer les performances dans le temps

puisque le pointage (aussi nommé pourcentage de recouvrement) est toujours calculé à partir de la même image.

### 3.1.1 Montage utilisé

Une donnée qui ne varie pas beaucoup entre les différents scénarios testés est le type de caméra utilisée. Pour tous les scénarios, une caméra infrarouge et une caméra couleur ont été utilisées simultanément. Pour la caméra infrarouge, un seul modèle a été utilisé. Il s'agit de la FLIR Thermovision A40. Cette caméra donne un angle de vue de 24 degré. Par exemple, à une distance de 1,2 m, on voit 0,51 m de large et 0,38 m de haut. Aussi, cette caméra donne des images de dimensions 320 lignes par 240 colonnes. Cependant, ces images ont été agrandies en des images de dimensions 640 lignes par 480 colonnes pour être égales aux dimensions des images des deux caméras couleurs utilisées. Pour ce qui est de ces deux caméras couleurs, les deux sont fabriquées par Sony : La DFW-SX910 et la XCD-710. La différence entre les deux est que la DFW-SX910 ne fournit que 7,5 images par seconde, mais fournit des images couleur qui n'ont pas besoin de traitement supplémentaire. Par contre, la XCD-710 fonctionne à 30 images par seconde, mais un filtre de Bayer (Wikipedia, 2009) doit être appliqué sur les images pour obtenir des images couleur. Durant les différents tests, la décision a été prise de ramener la vitesse de prise d'image des scénarios filmés avec la XCD-710 à environ 7,5 images par seconde pour réduire le temps de traitement et aussi parce que les performances ne semblaient pas meilleures avec une vitesse d'acquisition supérieure.

Toujours en ce qui a trait au montage, ces caméras sont supportées par des trépieds. Exception faite d'un scénario précis, les caméras étaient supportées par le même trépied grâce à une plaque qui les sépare de 19 cm. Aussi, tous les scénarios ont été filmés en plongée pour éviter que les trajectoires soient trop linéaires ainsi que pour respecter la restriction sur les objets de l'avant-plan qui doivent être dans le même plan (section 2.3.1.) La distance séparant les caméras des objets de l'avant-plan était aussi relativement constante soit environ un peu plus que la hauteur de 2 étages. (approximativement 8 mètres)

### 3.1.2 Crédation des éléments témoins (ground truth)

Étant donné l'utilisation de deux caméras de spectres différents, la soustraction d'arrière-plan sera différente pour les deux séquences vidéo. Aussi, même si l'hypothèse de la synchronisation des deux séquences vidéo a été faite, dans la pratique, il subsiste des erreurs. De plus, la soustraction d'arrière-plan contient, elle aussi, des erreurs qui varient aléatoirement avec le temps dans les séquences vidéo. Pour toutes ces raisons, l'utilisation du recouvrement de l'avant-plan comme mesure exacte des performances de la méthode de recalage automatique développée est biaisée. C'est pour obtenir une mesure de performances plus objective que des éléments témoins ont été utilisés. Cependant, pour que ces éléments témoins puissent être utilisés, ils doivent répondre à certains critères.

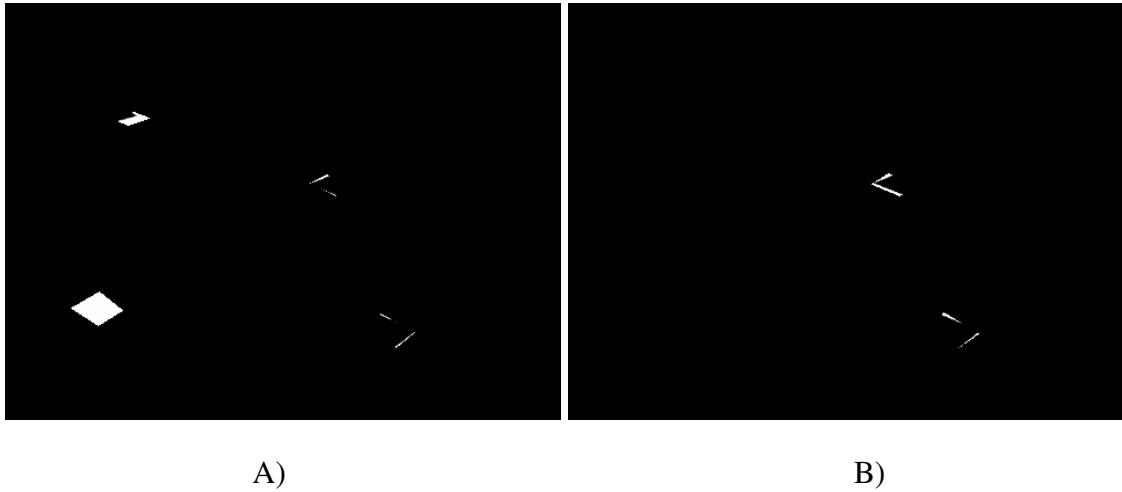
Un premier critère est que ces éléments doivent être visibles dans les deux séquences vidéo et donc dans les deux spectres utilisés, la lumière visible et l'infrarouge. Un autre critère est que ces éléments doivent avoir une surface assez grande pour qu'une matrice de transformation inexacte ait un impact significatif sur le nombre de pixels en recouvrement. Aussi, la forme de ces éléments doit être simple pour qu'ils puissent être facilement séparés du reste de l'image. Finalement, ces éléments doivent être répartis uniformément sur la surface de recouvrement pour diminuer l'impact des différences de distances des objets aux caméras lors de l'évaluation de la matrice de transformation. Ce phénomène est bien illustré par les images suivantes.



A)

B)

Figure 3.1 Images en recouvrement à la trame 63 (A) et 88 (B)



A)

B)

Figure 3.2 Images des erreurs des éléments témoins en recouvrement lorsqu'ils sont bien (A) et mal répartis (B)

Aux figures 3.1 et 3.2, des cartons refroidis sont placés dans la scène comme éléments témoins. À la figure 3.1, on peut observer un parallélogramme gris à l'intérieur des images. Il s'agit du résultat de l'image infrarouge transformée et donc du recalage. Si ce recalage est bon, l'image de la personne en infrarouge devrait être superposée à celle de l'image couleur. C'est le cas dans la figure 3.1 B mais pas dans la figure 3.1 A. Ceci signifie que la matrice de transformation utilisée n'était pas la bonne et qu'elle est n'est adapté que pour un secteur. Il faut donc que les éléments témoins utilisés comme métrique de qualité montrent que cette transformation n'est pas bonne. Si les cartons témoins sont tous disposés dans le même secteur (3.2 B), une matrice de transformation adaptée seulement à ce secteur pourrait résulter en un nombre très petit de pixels en erreur de recouvrement comme c'est le cas pour la figure 3.2 B. Il faut donc éparpiller les cartons témoins pour qu'ils couvrent plus uniformément la zone en recouvrement afin que la mesure de qualité tienne compte des différences des distances entre les objets et les caméras comme c'est le cas dans la figure 3.2 A. En effet, la figure 3.2 A est associée à un pourcentage d'erreur de 49,04 % alors que la figure 3.2 B donne plutôt un pourcentage d'erreur égal à 16,32 %. Il faut ici préciser que ces pourcentages d'erreurs

sont calculés comme ceux des images de pointage (section 2.4.2) et qu'ils sont calculés à chaque trame. Il sera maintenant question de comment la matrice de transformation témoin est obtenue ainsi que de la provenance des pourcentages d'erreur résiduels qui sont associés à cette matrice témoin.

Pour toutes les séquences vidéo utilisées, sauf une, des cartons de couleur ayant une température différente de celle de l'arrière-plan ont été utilisés (figure 3.3) puisqu'ils répondent à tous les critères énoncés précédemment. Dans la séquence pour laquelle des cartons n'ont pas été utilisés, ce sont plutôt des tables présentes dans l'arrière-plan, qui répondaient à tous ces critères, qui ont été utilisés (figure 3.4).

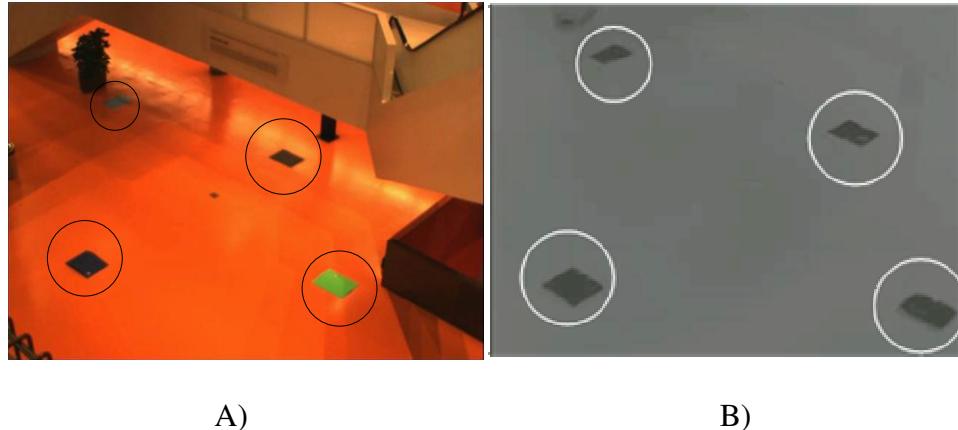


Figure 3.3 Images couleur (A) et infrarouge (B) des éléments témoins utilisés dans la majorité des scénarios

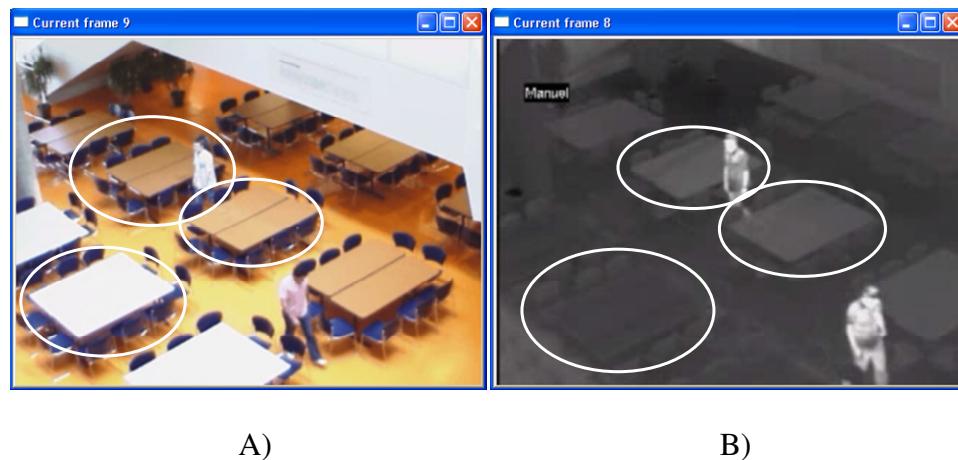


Figure 3.4 Images couleur (A) et infrarouge (B) des éléments témoins utilisés dans le scénario n'utilisant pas des cartons

Une fois des éléments témoins choisis, il faut trouver la matrice de transformation qui leur est associée. Cette matrice sera nommée matrice de transformation témoin. Rappelons que le choix a été fait d'utiliser une matrice de transformation affine et que ce choix est expliqué à la section 2.3.1. Par conséquent, trois paires de points suffisent pour déterminer la matrice. Ces points ont été sélectionnés manuellement avec MatLab. Ce sont quelques coins des quadrilatères utilisés comme éléments témoins.

Finalement, avec cette matrice de transformation témoin et ces images d'éléments témoins, il devient possible de quantifier la qualité d'un recalage. Pour ce faire, le même

procédé que pour le calcul du pointage (équation 2.14, section 2.4.2) lors de l'appariement des trajectoires est utilisé. Rappelons qu'il s'agit essentiellement de mettre en recouvrement l'image de gauche transformée par la matrice et l'image de droite et de calculer le pourcentage de pixels qui ne se trouvent pas en recouvrement. Si les éléments témoins ont été bien découpés et les points pour calculer la matrice bien choisis, ce pourcentage devrait être nul. Cependant, il est presque impossible de réaliser ces deux tâches à la perfection les sources d'erreurs étant : l'imprécision de sélection d'un pixel donné pour le choix des 3 points et les contours flous des objets, surtout dans le cas des images infrarouge. C'est pourquoi le pourcentage obtenu est rarement nul. Néanmoins, il est suffisamment petit et généralement inférieur à celui obtenu par les différentes méthodes de recalage automatique existantes. C'est pourquoi il sert ici de comparaison pour évaluer de manière objective les performances de l'algorithme développé. Il faut toutefois mentionner qu'il est possible que l'algorithme proposé surpassé quelquefois le pourcentage témoin obtenu puisqu'il est très difficile de sélectionner avec précision des points manuellement. Une analyse subjective des résultats semble d'ailleurs confirmer que le recalage obtenu par la méthode automatique est quelque fois meilleur que celui obtenu manuellement.

### **3.1.3 Explication des scénarios**

Pour bien tester les performances de l'algorithme de recalage automatique développé, 9 séquences vidéo différentes ont été tournées. Les différents scénarios présentés ont pour objectif d'évaluer les performances de la nouvelle méthode en faisant varier : le nombre de trajectoires, le degré de recouvrement des caméras, la longueur des séquences vidéo ainsi que l'orientation relative des caméras. Toutes les séquences vidéo utilisées commencent avec un certain nombre de trames n'ayant qu'un arrière-plan immobile pour permettre l'utilisation de l'algorithme de soustraction d'arrière-plan de moyenne temporelle (Shoushtarian & Bez, 2005) afin de détecter les humains en déplacement. La longueur des séquences vidéo utilisée varie entre 12 secondes et 3 minutes 59 secondes. Le nombre de personnes présentes en même temps dans la zone visible par les deux

caméras varie entre 0 et 5 personnes. Le scénario suivi par les acteurs est essentiellement toujours le même, des gens entrent dans la scène, marchent et sortent de la scène. Ces gens peuvent ensuite revenir dans le champ de vision des deux caméras et parfois sortir du champ d'une seule des deux. Généralement, le champ de vision de la caméra infrarouge est complètement inclus dans celui de la caméra couleur sauf pour quatre séquences vidéo. Dans ces séquences particulières, le champ de vision en recouvrement est diminué jusqu'à environ 25 % de celui de la caméra couleur. Finalement, elles ont toutes un arrière-plan plutôt simple étant composé du plancher de l'atrium Lorne-Trottier de l'école Polytechnique de Montréal. Seule exception, la séquence la plus courte dans laquelle les deux personnes présentes contournent des tables qui sont d'ailleurs utilisées comme éléments témoins.

### **3.1.4 Explication des graphiques obtenus**

Dans les sections qui vont suivre, de nombreux graphiques seront présentés. Ils seront divisés en trois types : les graphiques de performances, les graphiques des gains de la nouvelle méthode et les graphiques du nombre de trajectoires. La forme de chacun de ces types de graphiques sera expliquée dans les prochaines sous-sections. Aussi, une dernière sous-section expose et explique un tableau des paramètres utilisés pour chaque jeu de tests (30 exécutions de l'algorithme pour chaque méthode et pour chaque séquence vidéo).

#### **3.1.4.1 Graphiques de performance**

Pour commencer, voici un exemple de ce type de graphique :

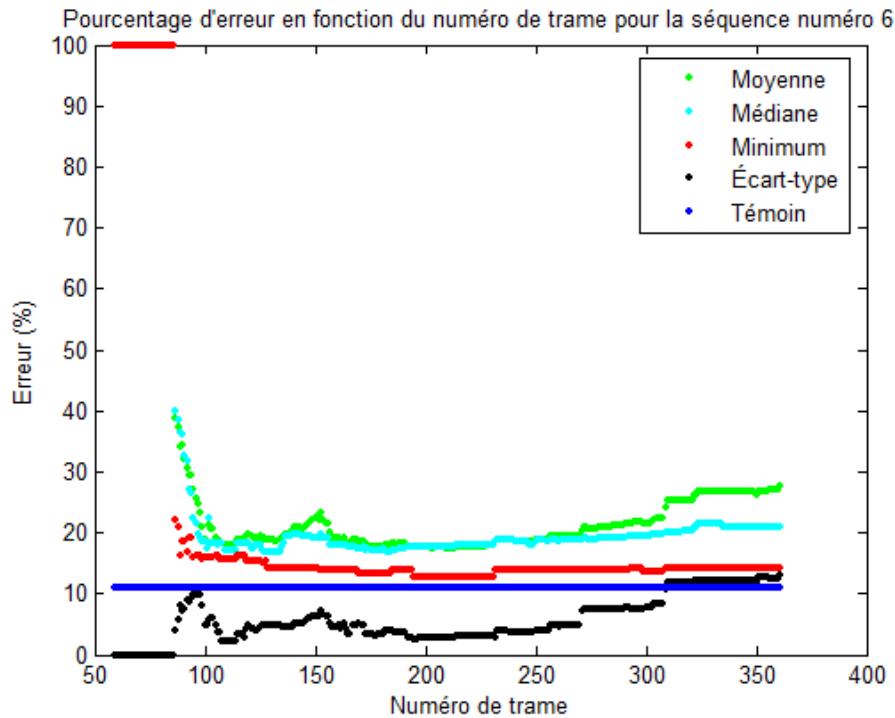


Figure 3.5 Exemple de graphique de performance de la nouvelle méthode de recalage

Tout d'abord, on remarque que l'axe des abscisses indique le numéro de trame dans la séquence vidéo et l'axe des ordonnées, le pourcentage d'erreur du recouvrement. Ensuite, la légende nous indique les différentes données qui sont représentées dans le graphique. Il en existe 5 par graphique : la moyenne, la médiane, le minimum, l'écart-type et le témoin. Ces 5 types de données sont l'objet des prochains paragraphes.

La première donnée représentée par les points verts dans le graphique est la moyenne de l'erreur de recouvrement pour chaque trame. Cette moyenne est obtenue en calculant, pour chaque trame, la moyenne des 30 exécutions de la méthode testée pour une séquence vidéo donnée. Dans la figure 3.5, il s'agit donc de la moyenne des erreurs de recouvrement de la nouvelle méthode sur la séquence vidéo numéro 6.

La deuxième donnée représentée par les points turquoise dans le graphique est la médiane de l'erreur de recouvrement pour chaque trame. Cette médiane est obtenue en calculant, pour chaque trame, la médiane des 30 exécutions de la méthode testée pour

une séquence vidéo donnée. Dans la figure 3.5, il s'agit donc de la médiane des erreurs de recouvrement de la nouvelle méthode sur la séquence vidéo numéro 6. On remarque que la médiane est souvent en dessous de la moyenne. Ceci signifie simplement que certaines exécutions ont donné de très mauvais résultats. Ces mauvais résultats sont engendrées par le choix aléatoire utilisé dans les algorithmes RANSAC.

La troisième donnée représentée par les points rouges dans le graphique est la valeur minimum de l'erreur de recouvrement pour chaque trame. Il s'agit de la plus petite valeur d'erreur de recouvrement observée, pour chaque trame, durant les 30 exécutions de la méthode testée pour une séquence vidéo donnée. Dans la figure 3.5, il s'agit donc des valeurs minimums des erreurs de recouvrement de la nouvelle méthode sur la séquence vidéo numéro 6.

La quatrième donnée représentée par les points noirs dans le graphique est l'écart-type de l'erreur de recouvrement pour chaque trame. L'écart-type est obtenu en calculant, pour chaque trame, l'écart-type des 30 exécutions de la méthode testée pour une séquence vidéo donnée. Dans la figure 3.5, il s'agit donc de l'écart-type des erreurs de recouvrement de la nouvelle méthode sur la séquence vidéo numéro 6.

La dernière donnée représentée par les points bleus dans le graphique est la valeur témoin de l'erreur de recouvrement pour chaque trame. Cette valeur témoin a été calculée en suivant le protocole exposé plus haut (section 3.1.2). Comme cette valeur témoin n'est calculée qu'une seule fois par séquence vidéo, elle est la même pour chaque trame. Néanmoins, cette valeur est illustrée dans le graphique pour chaque trame afin de faciliter la comparaison des résultats. Dans la figure 3.5, il s'agit donc des valeurs témoin des erreurs de recouvrement sur la séquence vidéo numéro 6.

### 3.1.4.2 Graphiques des gains de la nouvelle méthode

Pour commencer, voici un exemple de ce type de graphique :

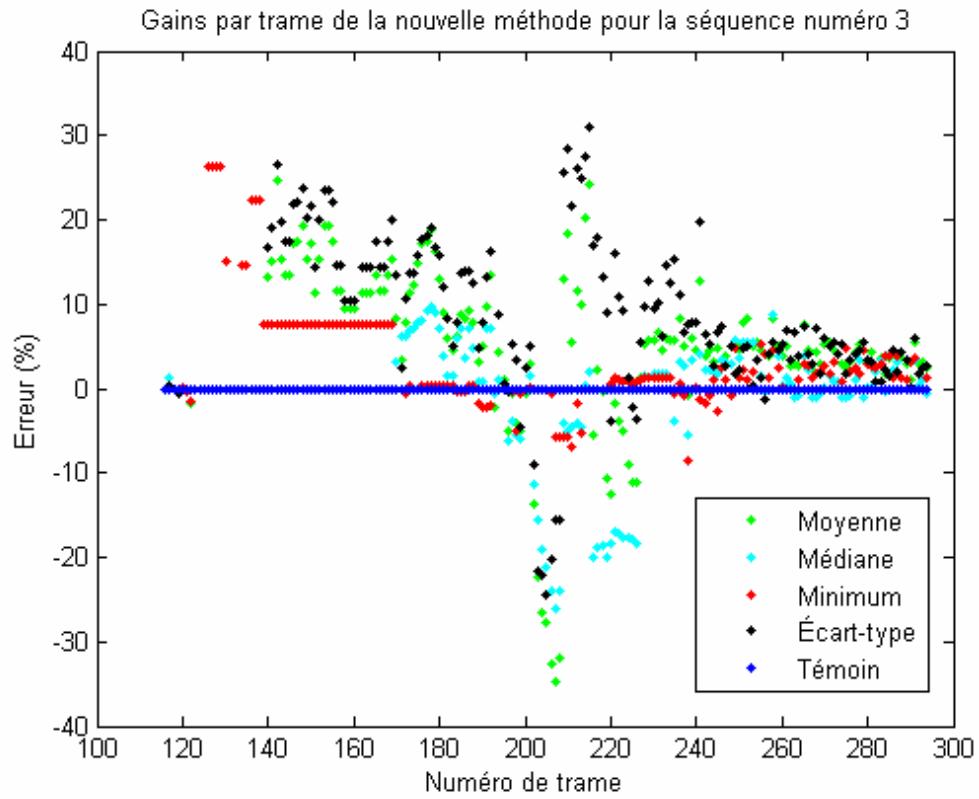


Figure 3.6 Exemple de graphique de gains de la nouvelle méthode de recalage

Tout d'abord, on remarque que l'axe des abscisses indique le numéro de trame dans la séquence vidéo et l'axe des ordonnées, le gain en pourcentage d'erreur du recouvrement. Ce gain en pourcentage d'erreur a été obtenu en soustrayant, pour une séquence vidéo donnée, le pourcentage d'erreur de la nouvelle méthode par le pourcentage d'erreur de la méthode de base. Cette soustraction a été faite pour chaque trame de la séquence vidéo donnée ainsi que pour chaque type de données récoltées. En résumé, un graphique de gains est obtenu en soustrayant le graphique de performance de la nouvelle méthode par celui de la méthode de base pour une séquence vidéo donnée. Notons aussi que la ligne bleu foncé, nommée témoin, représente la différence entre la méthode de base et elle-même ce qui explique pourquoi la valeur témoin est toujours nulle pour ce genre de graphique. Dans le graphique précédent, par exemple, une valeur de -30% signifie la nouvelle méthode a un pourcentage d'erreur 30% supérieur à celui de la méthode de

base. Ce graphique simplifie grandement la comparaison des deux méthodes. En effet, si le gain est positif, c'est donc que la nouvelle méthode surpassé la méthode de base à cette trame. L'inverse signifie que la méthode de base surpassé la nouvelle méthode.

### 3.1.4.3 Graphiques des nombres de trajectoires

Pour commencer, voici un exemple de ce type de graphique :

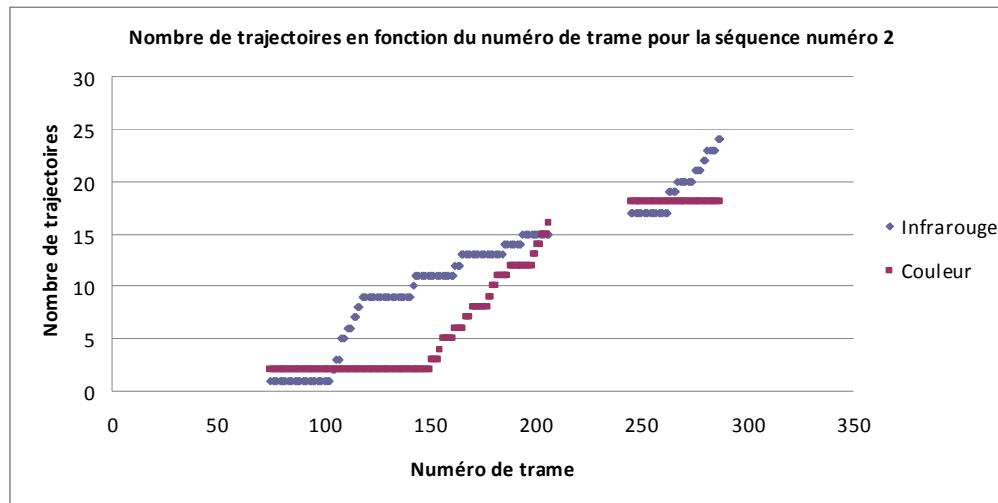


Figure 3.7 Exemple de graphique de nombres de trajectoires

Tout d'abord, on remarque que l'axe des abscisses indique le numéro de trame dans la séquence vidéo et l'axe des ordonnées, le nombre de trajectoires. Ensuite, la légende nous indique la couleur et la nature des deux courbes qui peuvent être observées, la progression du nombre de trajectoires pour le spectre infrarouge et la progression du nombre de trajectoires pour le spectre couleur. Ces nombres sont obtenus en comptant le nombre de trajectoires générées par l'algorithme de génération de trajectoires pour chaque caméra. Ce type de graphique sera utilisé pour établir des liens entre les performances des méthodes et la qualité des trajectoires.

### 3.1.4.4 Paramètres utilisés pour les différents jeux de tests

Voici trois tableaux qui résument les différents paramètres utilisés pour générer les différentes exécutions des différentes séquences :

Tableau 3.1 Paramètres généraux utilisés pour les différents tests

Numéro de séquence	1	2	3	4	5	6	7	8	9
Seuil données participantes	5	5	5	5	5	5	5	5	5
Intervalle entre 2 images de pointage	2	2	2	2	3	1	3	3	1

Le premier paramètre dans le tableau précédent est le seuil des données participantes. Ce seuil est utilisé par l'algorithme d'appariement des trajectoires pour connaître les points qui peuvent être considérés comme données participantes dans la trajectoire. Si la distance entre un point donné et son correspondant transformé est plus petite que ce seuil, le point est alors considéré comme participant (section 2.3.2.3). Ce seuil a été fixé à 5 pixels pour toutes les séquences vidéo testées.

Le deuxième paramètre donné dans le tableau précédent est l'intervalle de temps, en secondes, entre deux prises d'images pour la construction de l'image de pointage (section 2.4.2.1). Il faut ici noter que cet intervalle n'est utilisé que pour le calcul de l'image de pointage et non pas pour le calcul de la matrice de transformation. Rappelons que cette dernière est calculée à chaque trame. Dans certains cas, le calcul de l'image de pointage surcharge un peu trop le processeur. Pour ne pas ralentir inutilement l'algorithme, il a été décidé de faire varier le temps qui s'écoule entre l'acquisition d'images. Cet intervalle de temps ne doit cependant pas nuire à la qualité de l'image de pointage résultante. Ceci signifie qu'il faut quand même qu'un bon nombre d'images soient analysées pour bénéficier des avantages de l'utilisation d'une image de pointage. Les tests ont montré que 10 images sont généralement satisfaisantes pour construire une bonne image de pointage. Dans les situations où les séquences vidéo sont longues, une plus longue période entre deux images n'a que peu d'impact sur la qualité de l'image de

pointage résultante puisque la longueur de la séquence fera en sorte que grand nombre d'images seront quand même analysées et incorporées au besoin. Par contre, l'impact sur la rapidité d'exécution de l'algorithme est grand (presque 2 fois plus rapide entre un intervalle de 1 et 3 secondes mais probablement aussi due à des problèmes de fuites de mémoire) ce qui justifie l'utilisation d'un plus grand intervalle. Notons toutefois que ce critère a été ajouté uniquement pour accélérer les tests puisque la rapidité de l'algorithme n'a aucunement été analysée dans ce projet.

Les deux prochains tableaux contiennent les paramètres de l'algorithme de soustraction d'arrière-plan utilisé pour les deux types de séquences vidéo; couleur et infrarouge. Le premier paramètre que l'on y retrouve concerne le seuil à considérer pour décider si un pixel fait partie de l'avant-plan ou non. Rappelons que l'algorithme de moyenne temporelle utilisé (section 2.2.1) trouve les pixels d'avant-plan en effectuant la différence entre un l'intensité de chaque canal de couleur du pixel courant et la moyenne des intensités de chaque canal de couleur des pixels observés au même endroit précédemment. Si cette différence d'intensité est plus grande que le seuil donné dans ces tableaux sur au moins un canal de couleur, c'est que ce pixel fait partie de l'avant-plan. Notons que si aucune valeur n'est spécifiée pour ce paramètre, l'algorithme utilisera la valeur par défaut qui est fixée à 3 pixels.

Le deuxième paramètre donné dans ces tableaux est relié au filtre de prétraitement de l'algorithme de soustraction d'arrière-plan. Ce filtre est un filtre linéaire gaussien. Ce genre de filtre est très utilisé pour réduire le bruit induit par le capteur de la caméra. Cependant, l'utilisation de ce filtre a tendance à rendre l'image plus floue. C'est pourquoi ce filtre peut être activé ou désactivé et qu'il est possible de lui fournir une valeur entière impaire variant entre 1 et 63. Cette valeur représente la taille de la matrice (matrice carrée) qui est appliquée pour calculer la nouvelle valeur du pixel. Plus la valeur est grande, plus l'image sera floue car le filtre tiendra compte d'un plus grand voisinage. Si aucune valeur n'est donnée dans les tableaux, c'est que le filtre est désactivé pour cette séquence.

Le troisième et dernier paramètre des tableaux de paramètres de soustraction d’arrière-plan concerne la taille des « blobs » qui font parties de l’avant-plan. En effet, rappelons que suite à la soustraction d’arrière-plan, une méthode d’extraction des « blobs » est utilisée pour enlever les pixels d’avant-plan qui pourraient constituer du bruit (section 2.2.2). Ce troisième paramètre donne donc le nombre minimal de pixels qui doivent être connectés ensemble pour être considérés comme un « blob » et constituant donc l’avant-plan.

Tableau 3.2 Paramètres de l'algorithme de soustraction d'arrière-plan utilisés pour les différents tests avec la caméra infrarouge

Tableau 3.3 Paramètres de l'algorithme de soustraction d'arrière-plan utilisés pour les différents tests avec la caméra couleur

Numéro de séquence	1	2	3	4	5	6	7	8	9
Seuil d'avant-plan	-	-	10	10	-	-	-	-	-
Filtre prétraitement	-	-	-	-	-	-	-	-	-
Filtre taille « blob »	150	150	150	150	150	150	150	150	150

### 3.1.5 Explication de la méthode de recalage de base

Cette méthode est inspirée de Caspi et al. (2006). Rappelons que les auteurs n'utilisent que les trajectoires des objets en mouvement pour obtenir leur matrice de transformation. La méthode de base utilise donc la même méthodologie que la nouvelle méthode de recalage présentée dans le chapitre 2 sauf dans la façon de savoir si une matrice est meilleure que l'autre. La méthode de base se différencie de l'article de Caspi et al. (2006) sur deux points principaux. Ces différences étaient nécessaires afin d'être capable de comparer efficacement les deux méthodes. Les impacts de ces changements sont minimes puisqu'on obtiendrait les mêmes résultats à la fin d'une séquence vidéo donné pour la méthode de base que pour la méthode de l'article. Tout d'abord, le fait qu'elle est appliquée à chaque trame est la première différence entre la méthode de base et l'article. C'est une différence majeure puisque dans l'article, la matrice de transformation n'était recherchée qu'une fois la séquence vidéo complétée. Ce changement était nécessaire afin de pouvoir comparer les deux méthodes à chaque trame. Dérivée du fait qu'une matrice est calculée à chaque, on trouve la deuxième différence entre la méthode de base et l'article de Caspi et al. (2006); le choix, après chaque calcul d'une matrice de transformation, de garder la nouvelle matrice calculée ou

de conserver l'ancienne. Ce choix est effectué de la même manière que pour la nouvelle méthode (section 2.5) soit en gardant la matrice de transformation pour laquelle le pointage est le plus élevé. Étant donné que la méthode de base est basée uniquement sur les trajectoires, ceci signifie que l'on choisit la matrice avec le plus grand nombre de trajectoires bien appariées ou, en cas d'égalité, celle avec la plus petite somme d'erreurs euclidienne. Les détails du calcul de ces erreurs sont donnés ici :

1. La méthode de base calcule l'erreur moyenne pour chaque paire de trajectoires. Ceci signifie qu'elle fait d'abord la somme de la distance entre les points correspondants des deux trajectoires. Bien sûr, la trajectoire de gauche a été préalablement transformée par la matrice de transformation que l'on tente d'évaluer. Ensuite, elle divise cette somme d'erreurs par le nombre de points dans une trajectoire ce qui donne la moyenne d'erreur dans la paire de trajectoires.
2. La méthode de base définit une paire de trajectoires comme bien appariée (participante) si sa moyenne d'erreurs est plus petite qu'un certain seuil (30 pixels dans l'implantation utilisée pour les tests).

La méthode de base n'utilise que le nombre de trajectoires qui sont bien appariées pour choisir la meilleure matrice de transformation. Par conséquent, elle choisit la matrice qui a le plus grand nombre de paires de trajectoires participantes. Si ce nombre est égal, elle choisit la matrice pour laquelle la somme des erreurs moyennes des trajectoires participante est la plus basse.

### **3.2 Variables testées et résultats**

C'est dans cette section que se situent les différents scénarios testés, leurs résultats ainsi qu'une analyse de ces résultats. Chaque sous-section sera découpée de la même façon. Dans un premier temps, une description détaillée du scénario de séquence testée sera présentée. Il y sera question des différents événements qui peuvent influencer les résultats ainsi que certaines spécificités des séquences vidéo filmées. Ensuite, les graphiques et tableaux illustrant les résultats pour les 2 méthodes de recalage testées seront présentés. Ces figures et tableaux seront accompagnés d'une courte explication

sur leur nature. Finalement, l'analyse qui peut être faite de ces différents résultats est présentée. Commençons donc avec la première séquence vidéo testée.

### **3.2.1 Séquence vidéo numéro 1**

La première séquence vidéo a été tournée dans le but de vérifier les performances du recalage lorsqu'une seule personne est présente dans la scène. Malheureusement, un deuxième groupe de personnes s'est introduit dans la scène durant un court instant. Néanmoins, leur présence permet de montrer à l'intérieur d'une même séquence, la supériorité des résultats lors de la présence de plus d'un « blob ». Mais avant de s'attarder aux résultats, voici le scénario observé dans cette séquence.

#### **3.2.1.1 Description du scénario**

La séquence commence avec un arrière-plan vide. Une femme entre dans la scène approximativement à la trame 50 et marche dans la zone d'inclusion des champs de vision pour le reste de la séquence. À partir d'approximativement la trame 150, deux personnes marchant en ligne droite côté à côté entrent et sortent de la scène. Leur sortie s'effectue approximativement à la trame 200. Il faut aussi noter que quelques trames sont perdues à 2 reprises dans la séquence soit approximativement aux trames 70 et 230. Ces pertes sont causées par des ajustements automatiques de la caméra infrarouge.

#### **3.2.1.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.8 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 1 et la figure 3.9 donne les résultats pour la méthode de base.

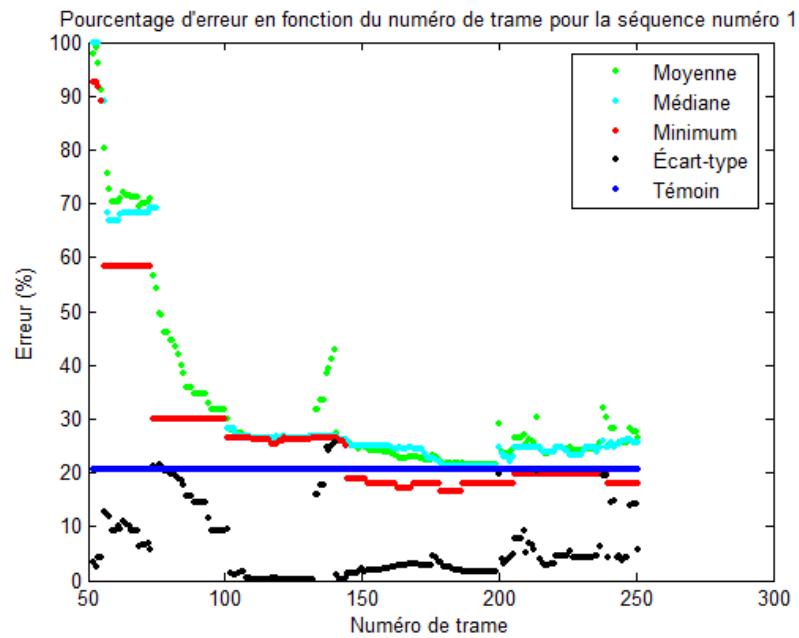


Figure 3.8 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 1

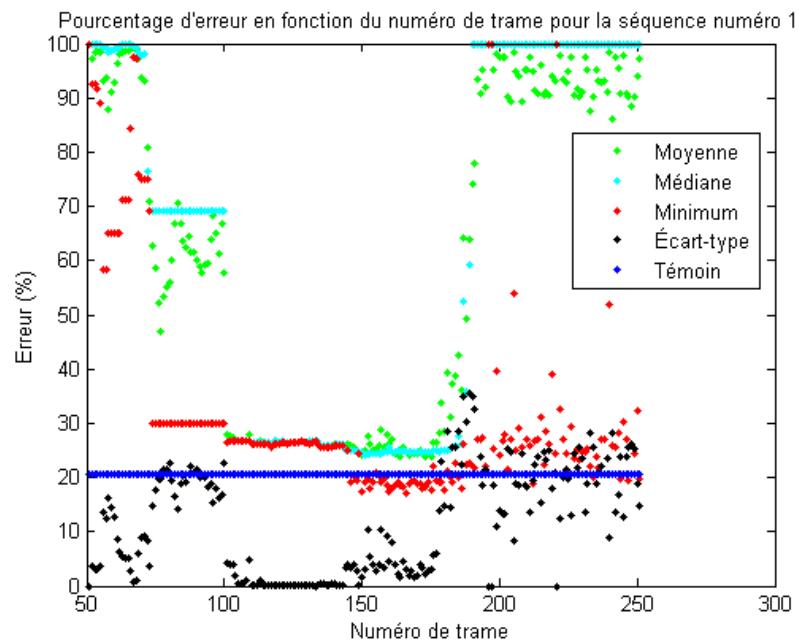


Figure 3.9 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 1

### 3.2.1.3 Analyse des résultats

La première différence qui ressort entre ces deux graphiques, c'est bien sûr la stabilité beaucoup plus grande d'une trame à l'autre de la nouvelle méthode développée comparativement à la méthode de base. L'instabilité est caractérisée par la présence de nuage de points autant pour la moyenne, la médiane, la valeur minimum et pour l'écart-type. Comparativement, pour un résultat plus stable, ces mêmes valeurs forment plutôt des points alignés dans la figure 3.8. Cette différence s'explique par choix de la matrice de transformation qui s'effectue à chaque trame pour le nouvel algorithme développé dans ce travail. Le fonctionnement de ce choix a été décrit à la section 2.5. Donc contrairement à l'algorithme de base, le nouvel algorithme va conserver l'ancienne matrice si celle qui a été nouvellement trouvée n'est pas meilleure que la précédente. Dans la méthode de base, aucune comparaison n'est faite d'où les fréquentes variations entre les trames.

Une autre différence très marquante qui peut être observée se situe à la fin de la séquence entre les trames 200 et 250. Alors que pour la nouvelle méthode, des résultats très proches de la valeur témoin sont observés, la méthode de base est pratiquement incapable d'accomplir le recalage. L'explication vient du scénario. Ces résultats coïncident avec la sortie du champ de vision du deuxième groupe de personnes. C'est aussi durant ce laps de temps que survient une des coupures reliées à la caméra infrarouge. Pour toutes ces raisons, les trajectoires observées sont courtes et nombreuses. Cette situation donne donc de nombreuses solutions possibles. Il est donc difficile de trouver la meilleure solution en n'utilisant que les trajectoires puisque plusieurs se ressemblent. Par contre, l'information additionnelle fournie par les images d'avant-plan permet de mieux évaluer la qualité ce qui aboutira en un meilleur recalage.

Une autre remarque pertinente pour les deux graphiques est que les meilleurs résultats sont observés entre les trames 150 et 200. L'explication à cette observation vient du scénario. En effet, c'est à ce moment qu'entre le deuxième groupe de personnes.

Comme il était possible de s'y attendre en théorie, il y a moins de solutions possibles avec deux groupes de « blobs » (personnes) qui forment des trajectoires qu'avec un seul.

Un autre élément qui ressort dans le graphique relié à la nouvelle méthode est la détérioration de la moyenne dans les environs de la trame 130. Cette détérioration est reliée à la soustraction d'arrière-plan qui a quelques difficultés à ce moment dans la séquence infrarouge. Le recouvrement devient donc moins bon d'où la difficulté de l'algorithme de bien effectuer le recalage. Ce phénomène illustre quelque peu les limites du nouvel algorithme développé.

Une autre particularité qui peut être observée pour les deux méthodes est la présence de plateaux pour les valeurs minimums des pourcentages d'erreurs. On peut associer ces plateaux aux performances optimales, dans une situation données, que peuvent atteindre les méthodes. Par situation donnée, on entend ici un intervalle de temps pour lequel la soustraction d'arrière-plan et la génération de trajectoire sont constants : pas de croisement de trajectoires, pas d'arrivées de nouvelles personnes, source d'erreur de soustraction d'arrière-plan constante ou pas d'erreur de soustraction d'arrière-plan. Pour ces situations données, ces minimums sont donc absous pour les méthodes.

Pour finir, il est aisé de conclure que, pour cette séquence, la nouvelle méthode de recalage surpassé l'ancienne qui ne se basait que sur les trajectoires. Le graphique de la figure 3.10 montrant les gains de la nouvelle méthode illustre encore mieux cette affirmation.

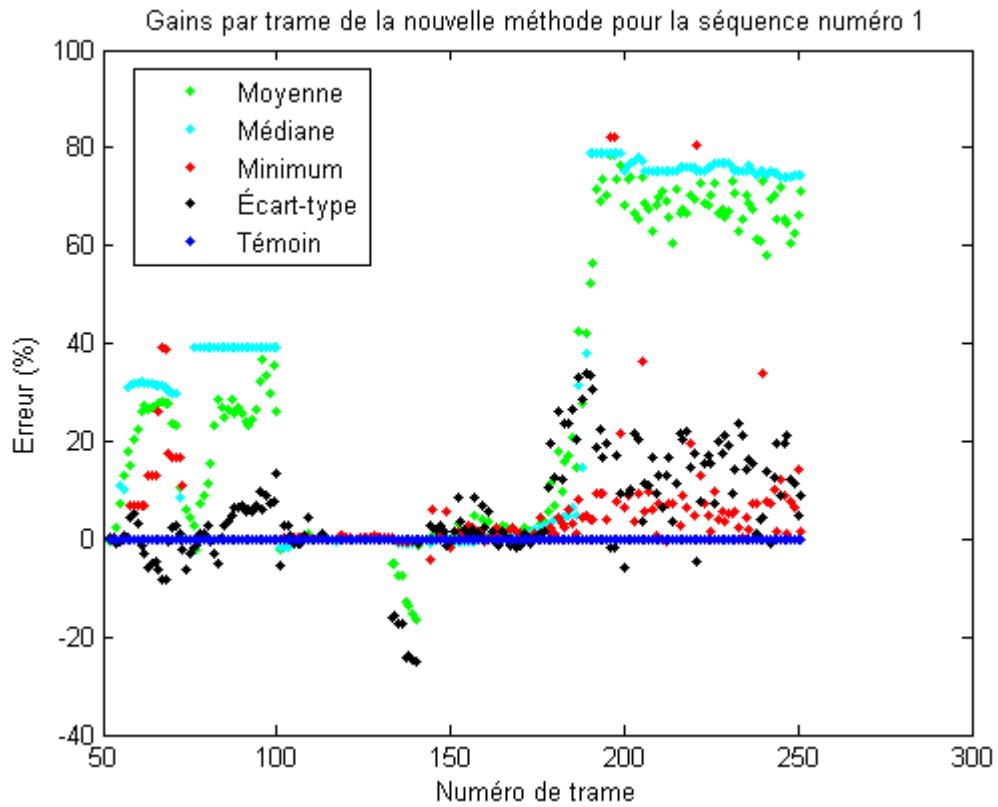


Figure 3.10 Graphique des gains de la nouvelle méthode pour la séquence numéro 1

Dans la figure précédente, on remarque d'ailleurs la supériorité de la nouvelle méthode au début de la séquence vidéo. Le début correspond au moment où le nombre de trajectoires est encore petit. Nous verrons que cette tendance de supériorité de la nouvelle méthode au début des séquences vidéo semble se confirmer dans plusieurs des autres séquences qui sont traitées dans les sections suivantes.

### 3.2.2 Séquence vidéo numéro 2

La seconde séquence vidéo a été tournée dans le but de vérifier les performances du recalage lorsqu'une seule personne est présente dans la scène. Cette séquence démontre que la méthode de base peut parfois surpasser la nouvelle méthode ce qui expose les limitations de cette dernière. Mais avant de s'attarder aux résultats, voici le scénario observé dans cette séquence.

### **3.2.2.1 Description du scénario**

La séquence commence avec un arrière-plan vide. Une femme entre dans la scène approximativement à la trame 70 et marche dans la zone d'inclusion des champs de vision jusqu'à approximativement la trame 200. À ce point, elle sort de la scène pour entrer de nouveau approximativement à la trame 240. Elle continue alors à marcher dans la zone d'inclusion des champs de vision jusqu'à la fin de la séquence. Approximativement à la trame 160, le reflet de la femme apparaît en infrarouge dans une plaque de l'arrière-plan.

### **3.2.2.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.11 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 2 et la figure 3.12 montre les résultats de recalage pour la méthode de base.

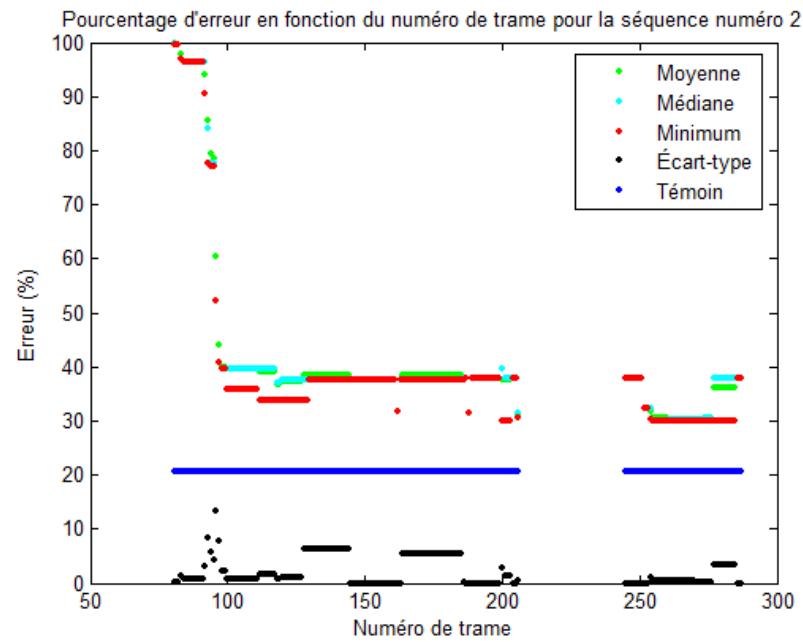


Figure 3.11 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 2

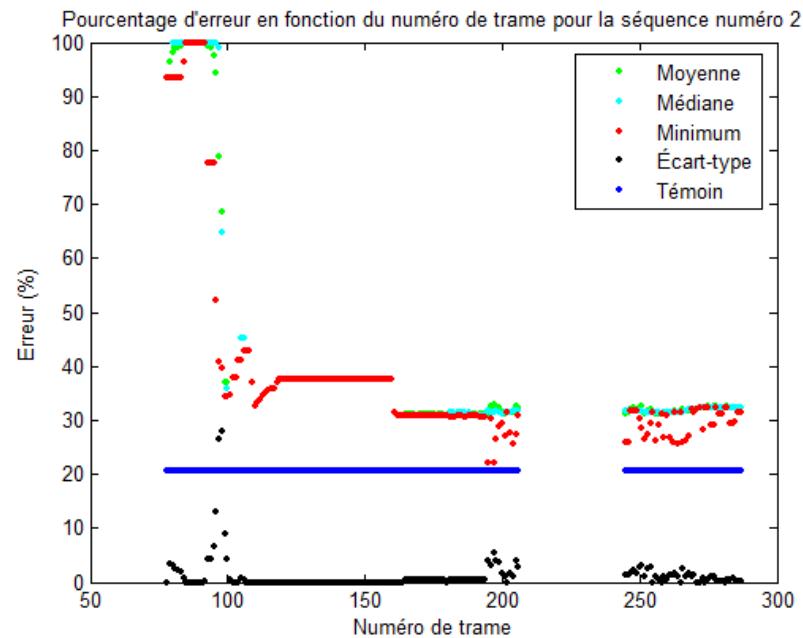


Figure 3.12 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 2

### 3.2.2.3 Analyse des résultats de la séquence

La première différence notable entre ces deux graphiques est, comme pour la première séquence vidéo, la plus grande stabilité entre 2 trames de la nouvelle méthode développée comparativement à la méthode de base. Rappelons que cette instabilité est caractérisée par la présence de nuage de points autant pour la moyenne, la médiane, la valeur minimum que l'écart-type. Dans la figure 3.12, ces nuages de points sont particulièrement visibles pour minimums pour les environs de la trame 200 et entre les trames 250 et 280. Comme mentionné précédemment, cette différence s'explique par choix de la matrice de transformation qui s'effectue à chaque trame pour le nouvel algorithme. La plus grande stabilité entre 2 trames est donc encore une fois observée pour cette séquence vidéo.

Une autre particularité qui peut être observée dans les graphiques précédents est la présence d'un trou dans les données entre les trames 200 et 250. Ce trou montre simplement qu'aucune matrice de transformation n'est calculée si aucun objet d'avant-plan n'est visible pour l'une ou l'autre des caméras. Dans ces situations, l'algorithme conserve la dernière matrice choisie mais aucune donnée n'est calculée pour ne pas fausser les statistiques globales qui seront données et analysées dans la section 3.3.

On remarque des différences entre les moments où les performances sont optimales pour les deux méthodes. Pour la nouvelle méthode, les meilleures performances sont obtenues dans les environs de la trame 100, dans les environs de la trame 200 et entre les trames 255 et 280. À ces moments, la nouvelle méthode surpassé la méthode de base. Par contre, la méthode de base surpassé la nouvelle méthode entre les trames 150 et 200. Pour mieux illustrer ces constatations, la figure 3.13 montre les différences des pourcentages d'erreur entre les deux méthodes. Notons ici la confirmation de la supériorité de la nouvelle méthode en début de séquence comme pour la séquence précédente.

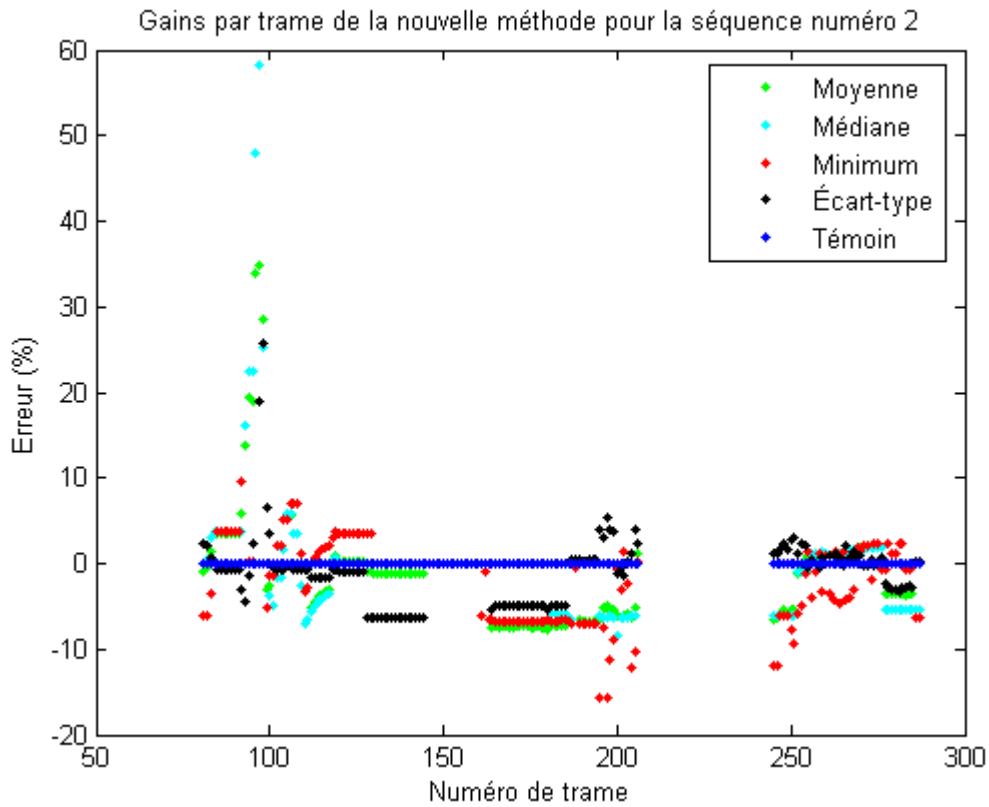


Figure 3.13 Graphique des gains de la nouvelle méthode pour la séquence numéro 2

L’explication de ces observations prend sa source dans l’utilisation, pour la nouvelle méthode, des images en recouvrement. En effet, l’obtention de ces dernières est reliée à la soustraction d’arrière-plan. La génération des trajectoires est aussi dépendante de cette soustraction d’arrière-plan. Par conséquent, les deux méthodes, qui utilisent tout deux les trajectoires, sont dépendantes de cette soustraction d’arrière-plan. Par contre, la nouvelle méthode est plus dépendante que la méthode de base sur cette soustraction d’arrière-plan, car elle l’utilise doublement. Malheureusement, il s’agit d’une arme à deux tranchants. Si cette soustraction est bonne, le fait de l’utiliser davantage augmente les performances du recalage comme c’est le cas pour les moments où la nouvelle méthode surpassé la méthode de base. Par contre, si cette soustraction est déficiente, la nouvelle méthode voit ses performances se dégrader plus rapidement. La méthode de base prend donc le dessus dans ces moments. Une preuve de cette explication peut être

vue dans le prochain graphique. Ce dernier montre le nombre de trajectoires qui est obtenu dans les deux séquences vidéo infrarouge et visible :

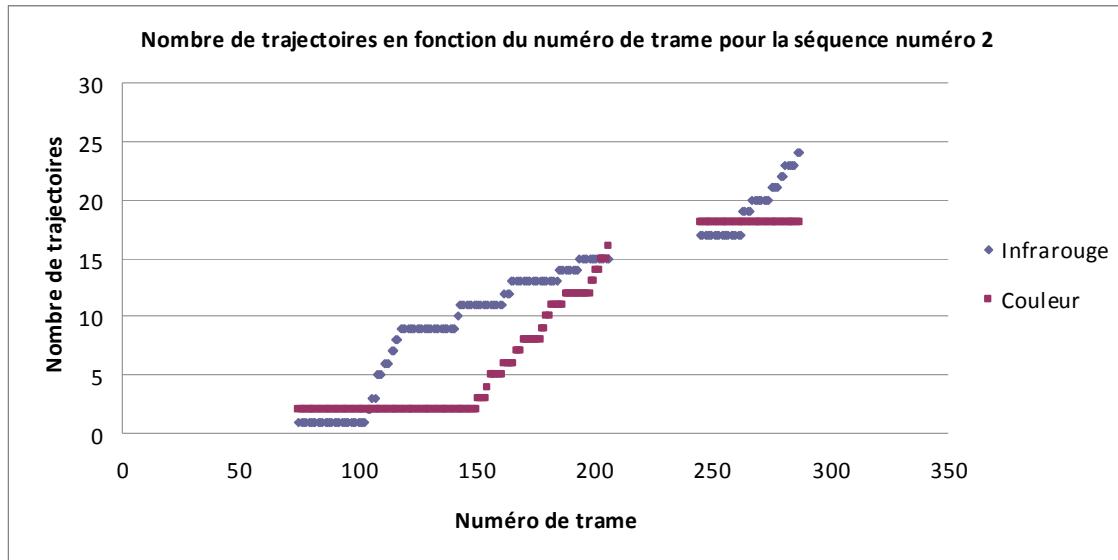


Figure 3.14 Graphique du nombre de trajectoires par trame pour les spectres infrarouge et couleur sur la séquence vidéo numéro 2

Il est possible d'observer les moments où la soustraction d'arrière-plan est bonne sur ce graphique. Comme cette séquence vidéo ne contient qu'un seul acteur, une bonne soustraction d'arrière-plan correspond à un nombre stable de trajectoires. En d'autres mots, le nombre de trajectoires demeure constant pour quelques trames consécutives. De plus, le recouvrement des images devrait être optimal lorsque la soustraction d'arrière-plan est bonne dans les deux spectres. Dans le graphique, ceci correspond aux moments où le nombre de trajectoires est stable autant pour la séquence vidéo infrarouge que pour la séquence vidéo couleur. Ces moments surviennent entre les trames 75 et 100, 125 et 150 et entre les trames 250 et 275. Les périodes 75-100 et 250-275 correspondent justement aux périodes où la nouvelle méthode obtient ses meilleurs résultats (environ de la trame 100 et trames 255-280). De plus, on remarque aussi que les meilleurs résultats de la nouvelle méthode de recalage coïncident avec les moments où le nombre de trajectoires de la caméra couleur et infrarouge sont presque les mêmes. Par opposition, on constate aussi que le moment où la méthode de base obtient ses meilleurs

résultats est le même que celui où le nombre de trajectoires des deux caméras est en augmentation presque constante (trames 150 à 200). Il est donc possible de conclure que la méthode de base est meilleure lorsque la soustraction d'arrière-plan est défaillante, et que la nouvelle méthode est supérieure lorsque cette même soustraction d'arrière-plan est efficace.

### **3.2.3 Séquence vidéo numéro 3**

Cette autre séquence vidéo a été tournée dans le but de vérifier les performances du recalage lorsqu'une seule personne est présente dans la scène et que la zone d'inclusion des champs de vision des caméras est réduite. De plus, une occlusion partielle survient ce qui rend la soustraction d'arrière-plan moins efficace. Sous ces conditions plus difficiles, la nouvelle méthode semble supérieure à celle de base. Mais avant de s'attarder aux résultats, voici le scénario observé dans cette séquence.

#### **3.2.3.1 Description du scénario**

La séquence commence avec un arrière-plan vide. Une ombre devient visible par la caméra infrarouge entre les trames 15 et 50. Une femme entre dans la scène dans les environs de la trame 115 et marche dans la zone d'inclusion des champs de vision jusqu'à la fin de la séquence. Approximativement à la trame 190, la femme s'approche d'un pot de fleurs qui a sensiblement la même couleur que ses vêtements. La surface d'inclusion des champs de vision de la caméra infrarouge dans celui de la caméra couleur est d'environ 23 % contrairement aux 100 % habituels.

#### **3.2.3.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.15 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 3 et la figure 3.16 montre les résultats obtenus avec la méthode de base.

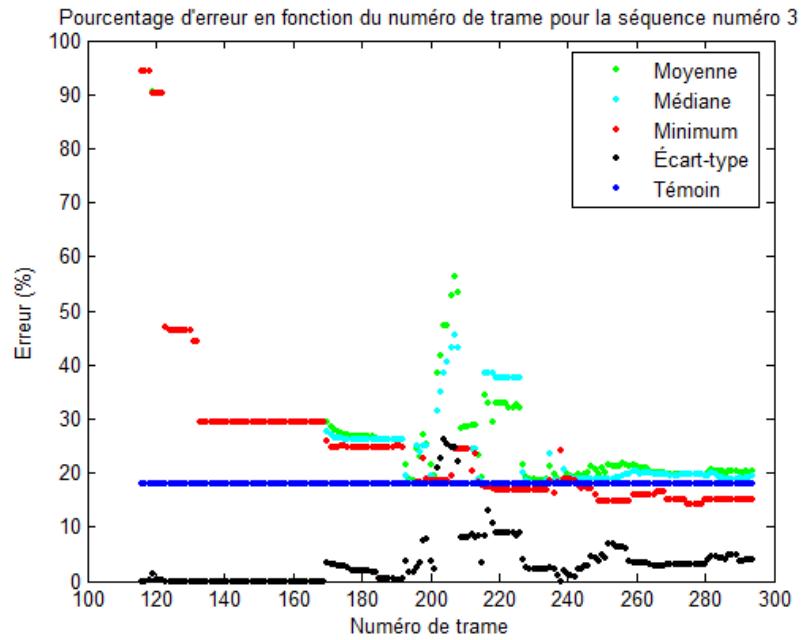


Figure 3.15 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 3

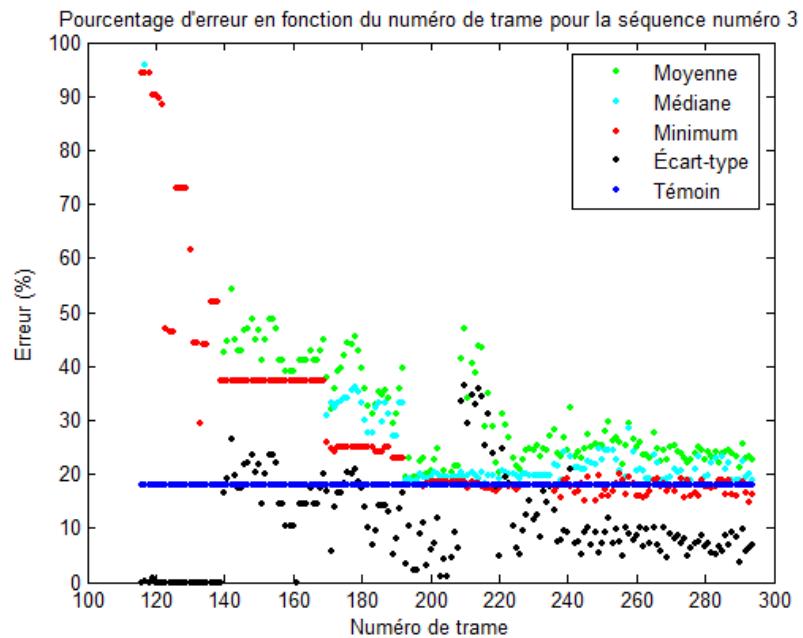


Figure 3.16 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 3

### 3.2.3.3 Analyse des résultats de la séquence

Encore une fois, on remarque des nuages de points dans le graphique de la méthode de base. Ces nuages confirment les conclusions faites sur la plus grande stabilité intertrame de la nouvelle méthode de recalage. Cela est confirmé par toutes les expériences réalisées.

Une autre différence observable entre les 2 figures reliée à la stabilité est que l'écart-type varie énormément dans la deuxième figure, mais pas dans la première. Le seul moment où l'écart-type est supérieur à 10 % dans le graphique de la nouvelle méthode est lorsque l'occlusion partielle survient. Comme cette augmentation est aussi observée dans la méthode de base, on peut conclure que cette anomalie est causée par le scénario et non par la méthode de recalage. La plus grande stabilité de la nouvelle méthode de recalage signifie que peu importe le nombre d'essais, le résultat obtenu est presque toujours le même. Dans un premier temps, cette faible variance montre que l'utilisation des images en recouvrement pour sélectionner la meilleure matrice rend les résultats plus indépendants du choix aléatoire des points et des trajectoires dans les différentes exécutions des algorithmes RANSAC utilisés. Dans un deuxième temps, la plus petite variance de la nouvelle méthode comparativement à la méthode de base montre que l'utilisation des images en recouvrement diminue l'impact des choix aléatoires des exécutions des algorithmes RANSAC utilisés dans les 2 méthodes. Pour s'en convaincre et pour mieux illustrer les prochaines observations, la figure 3.17 montre les gains de la nouvelle méthode sur celle de base. Notons encore ici la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes.

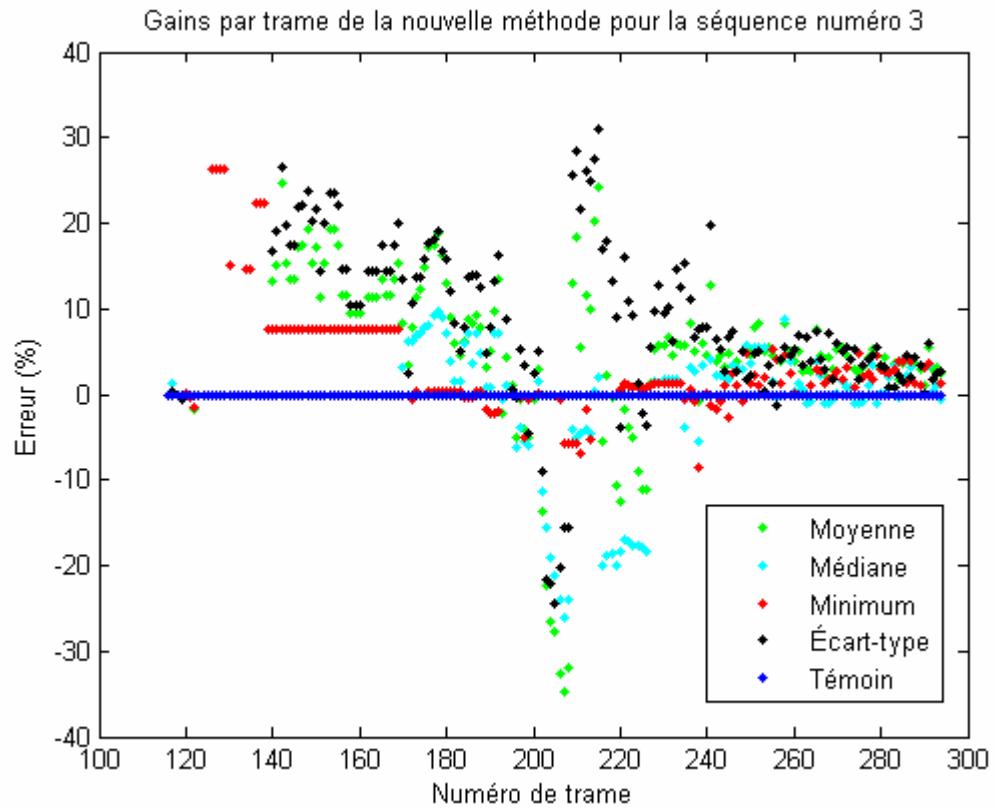


Figure 3.17 Graphique des gains de la nouvelle méthode pour la séquence numéro 3

L'autre tendance forte qui peut être déduite des deux graphiques de performance et du graphique des gains est que la moyenne, la médiane et la valeur minimum des résultats de la nouvelle méthode sont toujours égales ou plus basses que celles de la méthode de base. Le seul moment où cette observation est fausse est durant l'occlusion. Cependant, les performances des deux méthodes sont plutôt médiocres durant cet événement. De plus, l'écart-type plus grand de la méthode de base durant l'occlusion porte à croire que les meilleurs résultats de la méthode de base sont plus attribuables à la chance qu'à une réelle supériorité de la méthode dans les situations d'occlusions. Il est donc légitime de conclure que la nouvelle méthode de recalage est supérieure à celle de base pour cette séquence vidéo.

En ce qui a trait à l'impact du scénario sur les performances, il est intéressant de constater que les meilleurs résultats pour les deux types de méthode sont observés après l'occlusion. Cette observation s'explique par l'algorithme de génération des trajectoires utilisé. En effet, l'algorithme n'étant pas robuste aux occlusions, il génère une deuxième trajectoire après que l'occlusion soit terminée. Comme les deux méthodes sont basées sur l'appariement des trajectoires, on peut en conclure que l'ajout d'une deuxième trajectoire a un effet bénéfique sur les performances de recalage. Les prochaines séquences vidéo confirmeront cette tendance.

Une dernière observation intéressante pouvant être tirée de ces graphiques est reliée à la valeur minimum du recalage après l'occlusion qui se situe en dessous de la valeur témoin surtout pour le nouvel algorithme. Cette observation est attribuée à la zone d'inclusion des champs de vision plus restreinte de cette séquence vidéo. Comme cette zone est plus restreinte, les différences de distance entre les objets qui se trouvent dans cette zone et la caméra sont plus petites. Par conséquent, l'hypothèse fixée dans les restrictions (section 2.3.1) sur les distances égales entre les objets de la scène et la caméra est encore mieux respectée. Le recalage s'en trouve donc d'autant plus précis, même plus précis que le recalage manuel. La conclusion, que la présence de valeurs minimum en dessous des valeurs témoins démontre l'importance de cette restriction est confirmée par les 2 autres séquences vidéo à zone d'inclusion des champs de vision réduite (séquences 4 et 8) comme il en sera question un peu plus loin.

### 3.2.4 Séquence vidéo numéro 4

Cette séquence vidéo a été tournée dans le but de vérifier si les performances du recalage avec zone d'inclusion des champs de vision réduite sont influencées par le nombre de personnes qui marchent dans la scène. Une augmentation du nombre de personnes entraîne généralement une augmentation du nombre de trajectoires et devrait donc fournir de meilleures performances. Voici d'abord le scénario de la séquence.

### 3.2.4.1 Description du scénario

La séquence commence avec un arrière-plan vide. Une femme entre dans la scène approximativement à la trame 250 et marche dans la zone d'inclusion des champs de vision jusqu'à la fin de la séquence. Approximativement à la trame 300, un homme entre dans la scène. Cet homme sort de la zone d'inclusion des champs de vision environ à la trame 345 pour revenir environ à la trame 400 et finalement sortir définitivement environ à la trame 520. La surface d'inclusion des champs de vision des deux caméras est d'environ 23 %. La figure suivante permet de se faire une meilleure idée de ce que représente 23 % d'inclusion de champs de vision.



Figure 3.18 Images fusionnées des caméras couleur et infrarouge montrant une inclusion des champs de vision de 23 % (A) et de 100 % (B)

### 3.2.4.2 Présentation des résultats de la séquence

Le graphique de la Figure 3.19 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 4 et la Figure 3.20 donne le résultat obtenu avec la méthode de base.

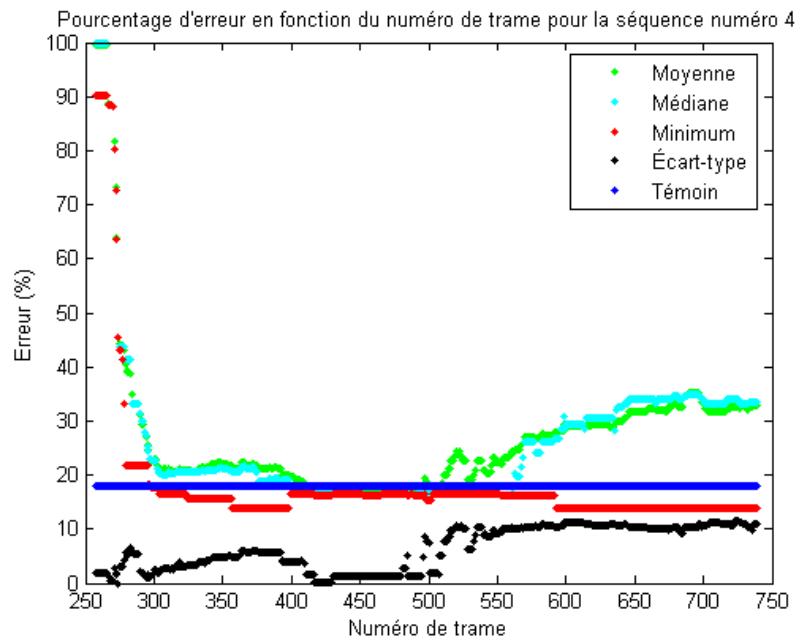


Figure 3.19 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 4

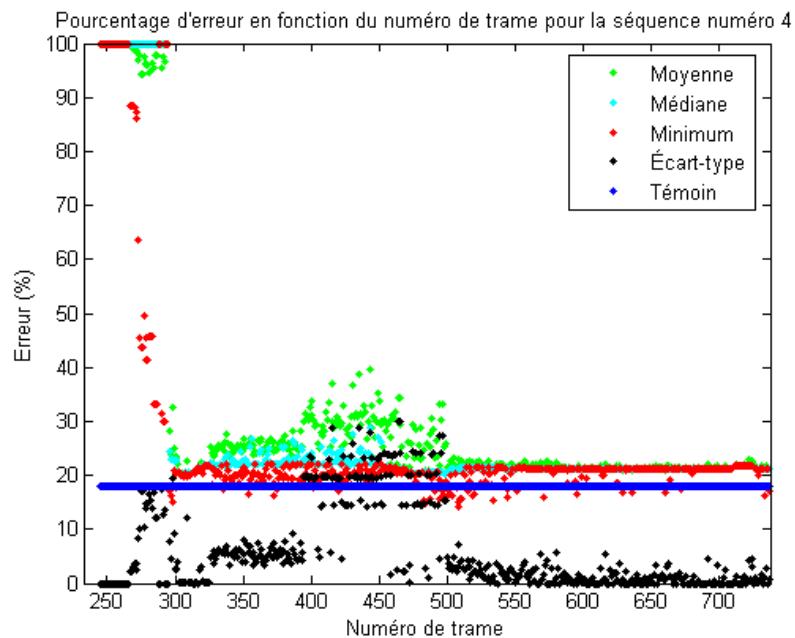


Figure 3.20 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 4

### 3.2.4.3 Analyse des résultats de la séquence

On remarque des nuages de points dans le graphique de la méthode de base. Une observation intéressante confirme l'hypothèse que l'augmentation du nombre de trajectoires augmente grandement les performances des méthodes de recalage automatique basées sur les trajectoires. En effet, les deux méthodes obtiennent des résultats très près des valeurs témoins presque aussitôt que la deuxième personne entre en scène (approximativement à la trame 300). Remarquons aussi qu'avant l'arrivée de cette deuxième personne, la méthode de base a une moyenne et une médiane d'erreur très élevée alors que la nouvelle méthode montre une tendance à la baisse de ces erreurs. Ceci est due au fait que l'image de pointage se complexifie (plus de trames la compose) avec le temps ce qui permet de mieux choisir la matrice de transformation et donc d'obtenir de meilleurs résultats de recalage. Par contre, la méthode de base n'a pas cet avantage ce qui explique pourquoi ses performances sont encore médiocres et ne s'améliorent que lorsqu'une nouvelle trajectoire est disponible.

Une autre différence marquante entre les deux graphiques est visible entre les trames 400 et 500. Alors qu'il s'agit du moment de performances optimal pour le nouvel algorithme, il s'agit plutôt d'un moment de grande instabilité pour la méthode de base. Pour mieux illustrer cette observation, la figure 3.21 montre les gains de la nouvelle méthode pour cette séquence vidéo. Notons encore ici la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes.

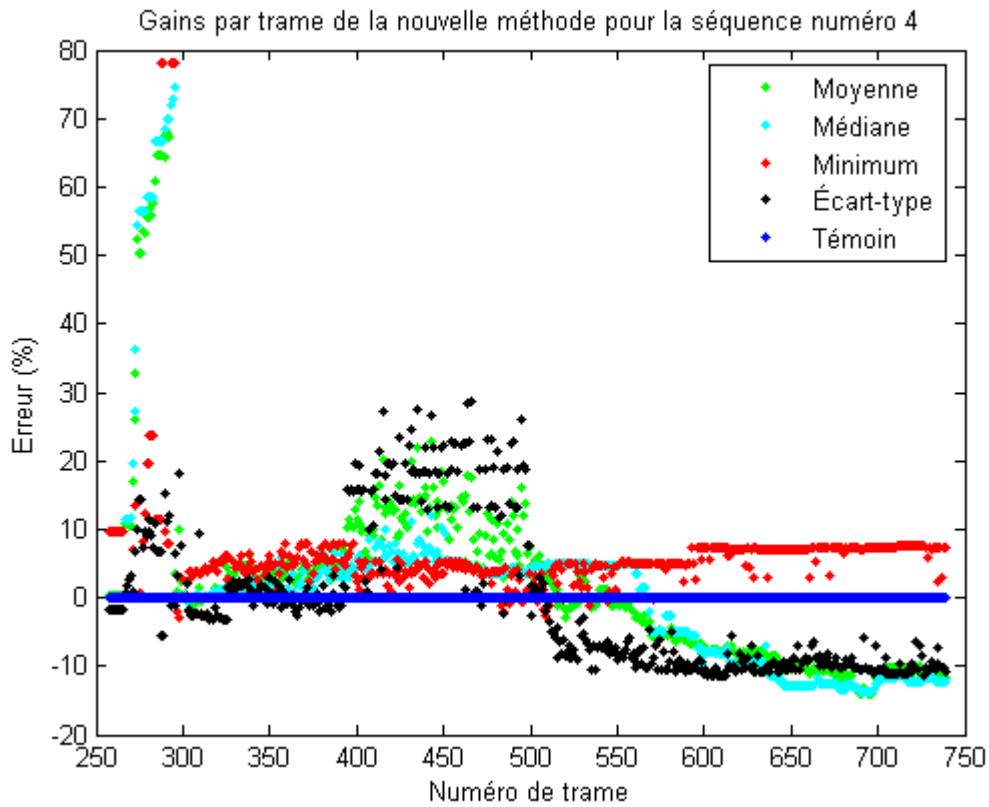


Figure 3.21 Graphique des gains de la nouvelle méthode pour la séquence numéro 4

En observant le scénario, on remarque que ces résultats coïncident avec le moment du retour de la deuxième personne dans la zone d'inclusion des champs de vision. Cette observation semble pointée vers une augmentation des performances de la nouvelle méthode lorsque plusieurs personnes sont présentes dans la scène alors que la méthode de base semble plutôt déstabilisée par cet ajout de trajectoire. La conclusion inverse semble aussi confirmée par les observations tirées des trois graphiques précédents entre les trames 500 et la fin de la séquence. En effet, lorsque la deuxième personne sort de la zone d'inclusion des champs de vision pour ne plus y revenir, la méthode de base obtient ses meilleurs résultats alors que la nouvelle méthode diverge. L'analyse qui peut en être faite est que l'utilisation de l'image en recouvrement instantanée améliore les performances lorsque les objets en mouvement sont nombreux et visibles au même moment. Par contre, si ceux-ci sont moins nombreux, l'utilisation unique des trajectoires

est avantagee car le nombre de points qui composent ces trajectoires continue d'augmenter avec le temps ce qui augmente ainsi la quantité de données disponibles pour la méthode de base. Cette même constatation est faite dans l'analyse des séquences ultérieures.

Une dernière observation intéressante pouvant être tirée de ces graphiques est reliée à la valeur minimum qui se situe parfois en dessous de la valeur témoin. Comme mentionné précédemment (séquence 3), ceci indique que la restriction sur la coplanarité des objets de la scène est mieux respectée dans les séquences vidéo à zone réduite d'inclusion des champs de vision.

### **3.2.5 Séquence vidéo numéro 5**

Cette séquence vidéo a été tournée dans le but de vérifier si les performances du recalage de la nouvelle méthode développée pouvaient aussi fournir une bonne estimation dans des situations où les caméras ne sont pas fixées au même trépied. Voici d'abord le scénario de la séquence.

#### **3.2.5.1 Description du scénario**

La séquence commence avec un arrière-plan vide. Cette séquence met en scène trois personnes qui marchent de manière aléatoire dans la scène en se croisant à multiple reprise. Le premier homme entre dans la zone d'inclusion des champs de vision vers la trame 70. Cette séquence vidéo est la seule qui a été tournée avec des caméras sur 2 trépieds différents ce qui fait en sorte que les caméras ne sont pas à la même hauteur et n'ont pas le même angle. Dans ces conditions, les conditions préalables pour l'utilisation de la matrice de transformation affine ne sont pas respectées.

#### **3.2.5.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.22 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 5 et la figure 3.23 donne les résultats obtenus pour la méthode de base.

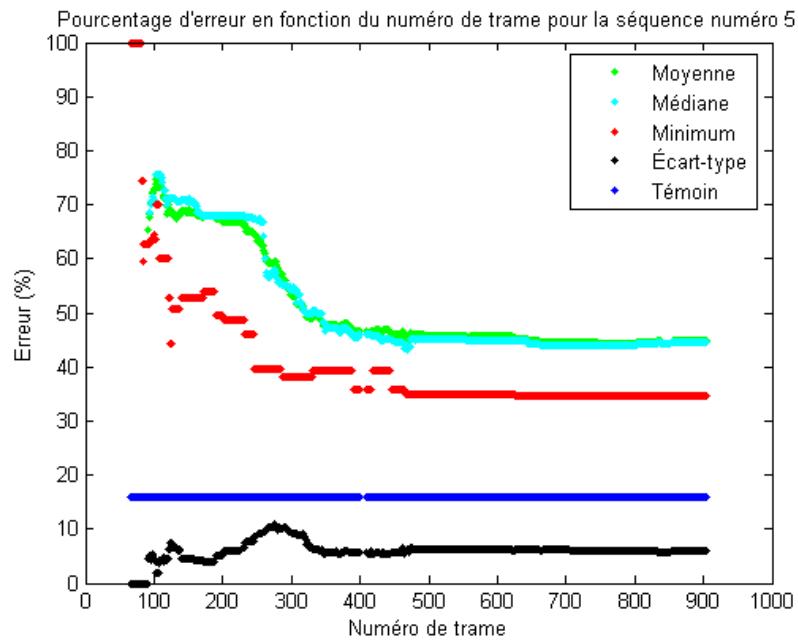


Figure 3.22 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 5

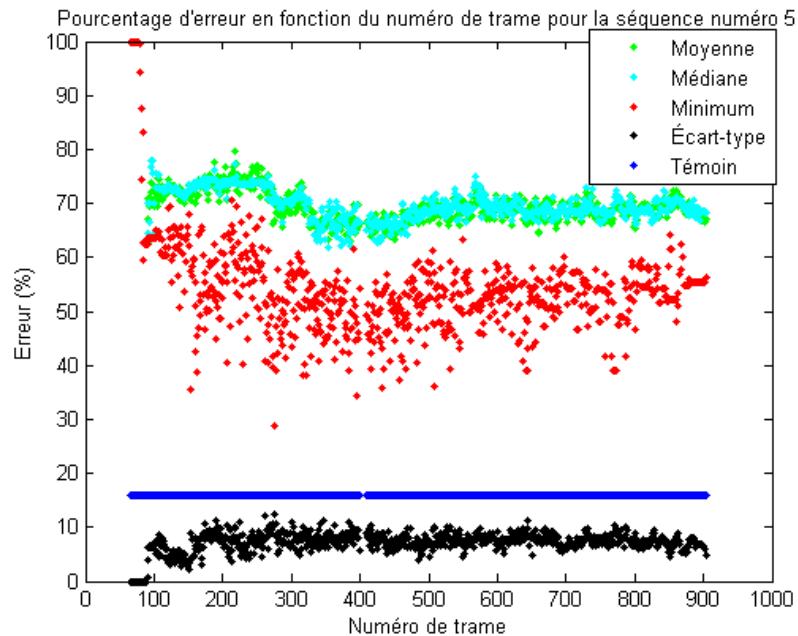


Figure 3.23 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 5

### 3.2.5.3 Analyse des résultats de la séquence

Une caractéristique qui fait en sorte que cet essai se démarque est le haut pourcentage d'erreur des 2 techniques de recalage automatique. En effet, alors que ce pourcentage d'erreur atteint les 10 ou 20 %, il se maintient plutôt vers les 45 % pour la nouvelle méthode de recalage alors que la méthode de base se maintient plutôt près de 65 %. Cependant, on remarque aussi que le pourcentage témoin est plus élevé que dans les autres essais. Ces pourcentages d'erreurs plus élevés peuvent être attribués au fait que l'utilisation de deux trépieds rend plus difficile l'ajustement des caméras qui doivent être à la même hauteur. Des caméras qui ne sont plus à la même hauteur fait en sorte que les conditions d'utilisation de la matrice de transformation affine ne sont plus tout à fait respectées. Par contre, comme cette différence de hauteur est minime, on peut assumer que le recalage demeure quand même possible. C'est ce que les prochaines observations semblent démontrer.

Le scénario est intéressant car il simule des situations réelles avec plusieurs personnes qui entrent et sortent de la zone d'inclusion des champs de vision. Par conséquent, le fait que l'on remarque de meilleurs résultats et une stabilisation des résultats dans la nouvelle méthode laisse croire que la nouvelle méthode de recalage surpassé la méthode de base lorsque le nombre de personnes visibles dans la scène croît. La figure 3.24 illustrant les gains de la nouvelle méthode confirme la supériorité de la nouvelle méthode pour cette séquence vidéo. Notons encore ici la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes.

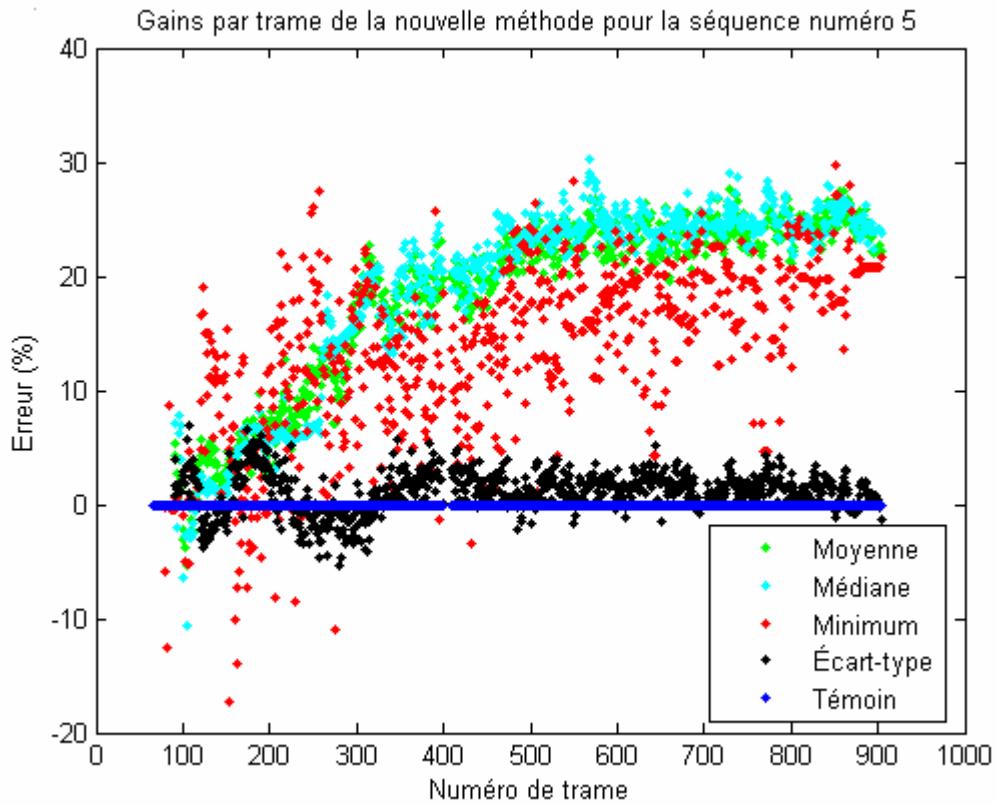


Figure 3.24 Graphique des gains de la nouvelle méthode pour la séquence numéro 5

Cette conclusion a déjà été énoncée précédemment et les prochains essais semblent aussi la confirmer. Somme toute, les résultats permettent de confirmer que même si les conditions d'utilisation de la matrice affine ne sont pas tout à fait respectées, la méthode donne quand même une bonne approximation du recalage.

### 3.2.6 Séquence vidéo numéro 6

Cette séquence vidéo a été tournée dans le but de vérifier les performances du nouvel algorithme dans une situation se rapprochant de la réalité, mais en limitant à 3 personnes le nombre d'acteurs présent dans la scène. Voici d'abord le scénario de la séquence.

### **3.2.6.1 Description du scénario**

La séquence commence avec un arrière-plan vide. Une femme entre dans la scène dans les environs de la trame 50 et marche dans la zone d'inclusion des champs de vision pour le reste de la séquence. À partir d'environ la trame 150, un homme entre dans la scène et marche dans la zone d'inclusion des champs de vision jusqu'à la fin de la séquence. Approximativement à la trame 225, un troisième homme entre dans la scène et marche dans la zone d'inclusion des champs de vision jusqu'à la fin de la séquence. À partir de ce point, 3 personnes sont en tout temps dans la zone et se croisent à quelques reprises.

### **3.2.6.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.25 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 6 et la figure 3.26 donne les résultats obtenus avec la méthode de base.

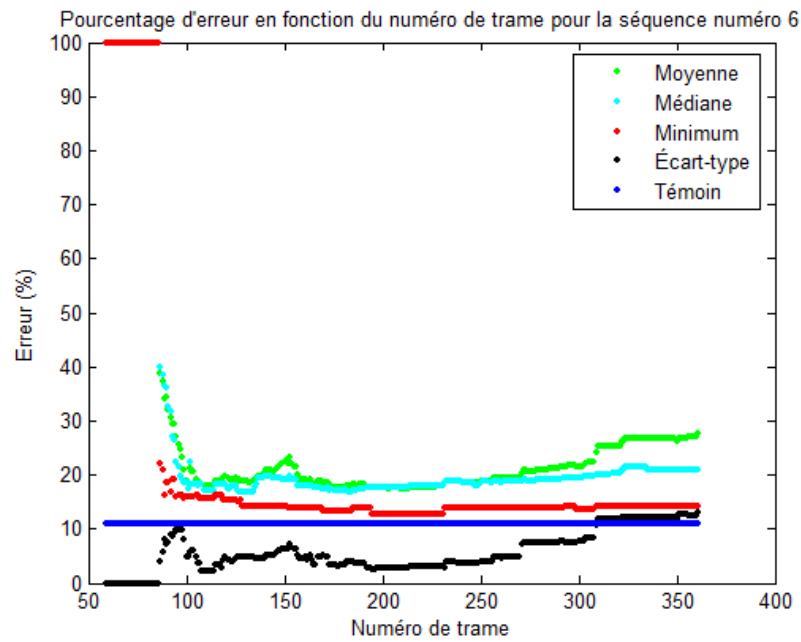


Figure 3.25 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 6

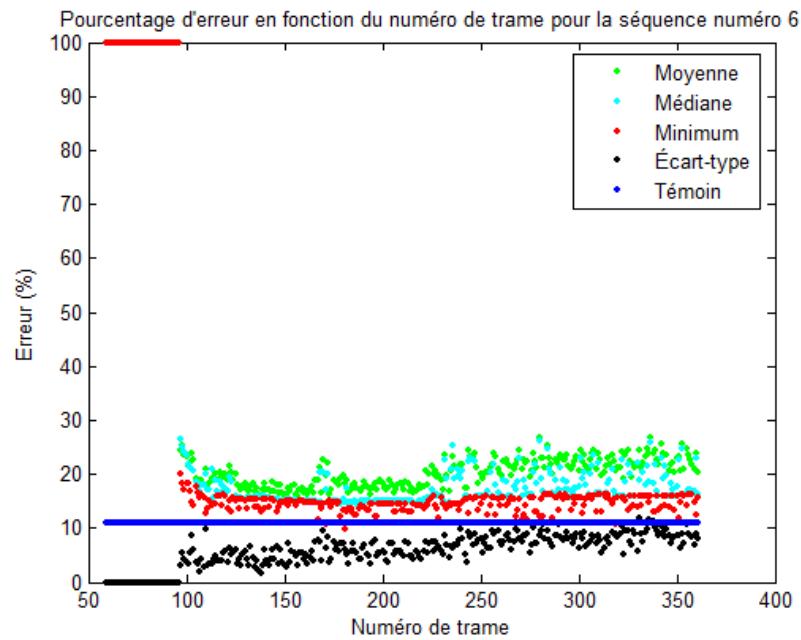


Figure 3.26 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 6

### 3.2.6.3 Analyse des résultats de la séquence

Une première constatation est que les deux graphiques comportent plusieurs similarités. D'abord, ils ont globalement des performances semblables. La figure 3.27 des gains de la nouvelle méthode le montre d'ailleurs très clairement. Notons aussi la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes.

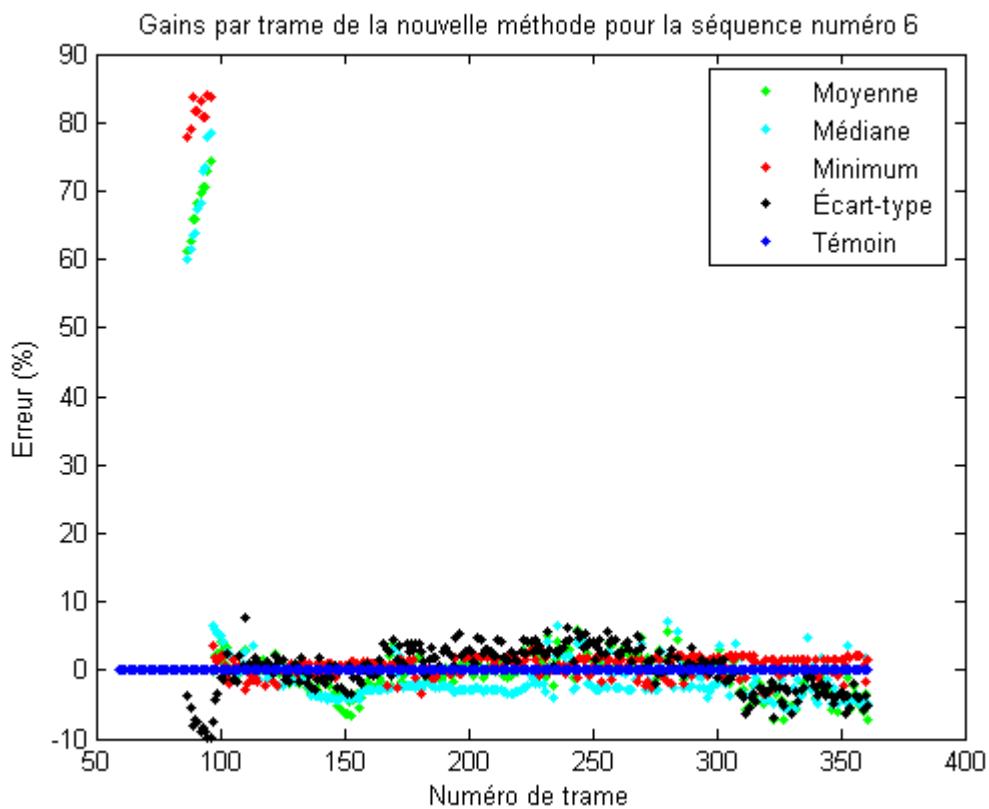


Figure 3.27 Graphique des gains de la nouvelle méthode pour la séquence numéro 6

Ensuite, tous les deux obtiennent des bons résultats (près de 15 %) aux mêmes moments soit lorsque la deuxième personne entre en scène et quelques trames après l'apparition de la troisième personne. Cette constatation confirme l'hypothèse qu'il faut plus d'un objet en mouvement pour que les performances soient intéressantes. Aussi, le pic d'erreur dans les environs de la trame 150 qui correspond à une détérioration des

performances est causé par des erreurs de l'algorithme de génération des trajectoires dues au fait que la troisième personne croise une des 2 personnes déjà présente. On remarque que ce pic survient plus rapidement pour la nouvelle méthode que pour la méthode de base. Comme la nouvelle méthode se base sur les images en recouvrement, elle est plus affectée au début du croisement car il y a fusion des deux blobs ce qui rend plus difficile le recouvrement car la forme est plus irrégulière et différente entre les 2 spectres. Par contre, lorsque le croisement sera terminé, les silhouettes pourront de nouveau être appariées plus efficacement ce qui entraîne une augmentation de performances. Par contre, la méthode de base est plus affectée au moment où les croisements sont finis car c'est à ce moment qu'une nouvelle trajectoire est créée. Lors de sa création, cette nouvelle trajectoire n'aura qu'un seul point ce qui la rend bien appariée pour beaucoup de matrices différentes ce qui augmente les chances de mal évaluer quelle matrice est la plus performante. Cette ambiguïté diminue bien sûr avec le temps puisque le nombre de points composant la nouvelle trajectoire augmente ce qui la rend plus significative. Enfin, le dernier élément semblable pour les 2 méthodes est le fait que les performances semblent moins stables et moins bonnes à la fin de la séquence. Cette détérioration s'explique par le fait que les trois personnes demeurent toujours dans la zone d'inclusion des champs de vision et se croisent ainsi beaucoup plus souvent. Comme l'algorithme de génération des trajectoires n'est pas robuste aux situations de fusions et de séparation des « blobs », il génère beaucoup d'erreurs. Il est beaucoup plus difficile de bien apparier des trajectoires non valides que de bonnes trajectoires.

Le seul élément qui semble confirmer la supériorité du nouvel algorithme se situe vers la fin de la séquence. En effet, on remarque que la divergence de l'algorithme, causée par le trop grand nombre d'erreurs de trajectoires, survient plus tard (trame 260) que dans la méthode de base (trame 225). On peut donc conclure que l'utilisation des images en recouvrement rend le recalage automatique un peu plus robuste aux erreurs de trajectoires.

### 3.2.7 Séquence vidéo numéro 7

Cette séquence vidéo a été tournée dans le but de vérifier les performances à plus long terme. Il s'agit de la plus longue séquence tournée pour tester le nouvel algorithme de recalage. Avec l'algorithme de génération de trajectoires utilisé, une plus longue séquence signifie une très grande quantité de trajectoires et aussi d'erreurs de trajectoires car cet algorithme n'est pas le plus performant. Les résultats pourraient donc s'apparenter à une séquence réelle tournée dans un lieu public avec un très grand nombre de personnes. Voici d'abord le scénario de la séquence.

#### 3.2.7.1 Description du scénario

Il contient jusqu'à 5 personnes qui se promènent dans la zone d'inclusion des champs de vision, en sortent et y entrent. Voici quelques événements qui sont visibles dans les graphiques. La séquence commence avec un arrière-plan vide. Deux personnes entrent presque côté à côté approximativement à la trame 150. Deux autres personnes entrent approximativement à la trame 165. Environ à la trame 200, il ne reste que 2 personnes dans la scène, dont une qui se trouve dans une zone d'occlusion avec un pot de fleurs. De la trame 450 jusqu'à la trame 500, aucune personne n'est visible par les 2 caméras. Approximativement à la trame 550, 5 personnes en mouvement sont visibles par la caméra couleur. Dans les environs de la trame 800, 2 personnes restent presque immobiles dans la zone d'inclusion des champs de vision. Dans les environs de la trame 1200, plus personne n'est en mouvement dans chacune des 2 caméras.

#### 3.2.7.2 Présentation des résultats de la séquence

Le graphique de la figure 3.28 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 7 et la figure 3.29 donne les résultats pour la méthode de base.

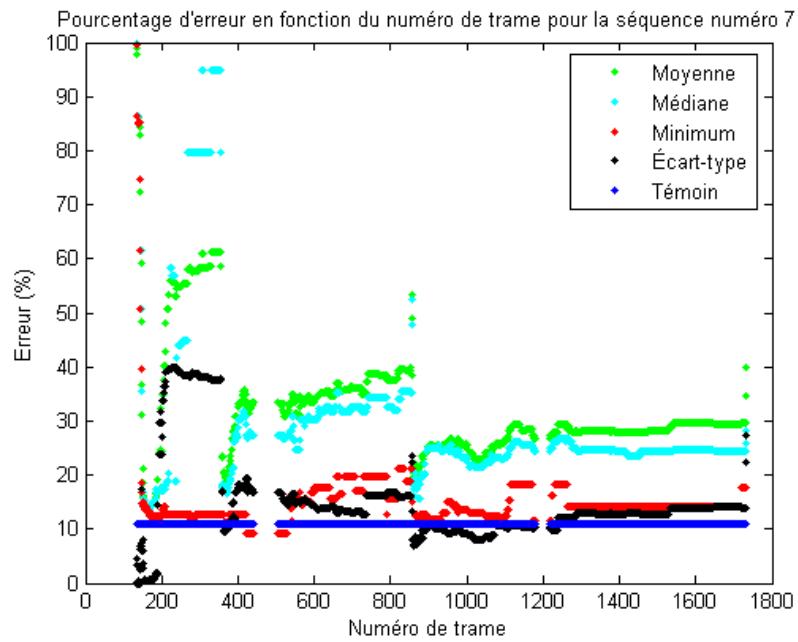


Figure 3.28 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 7

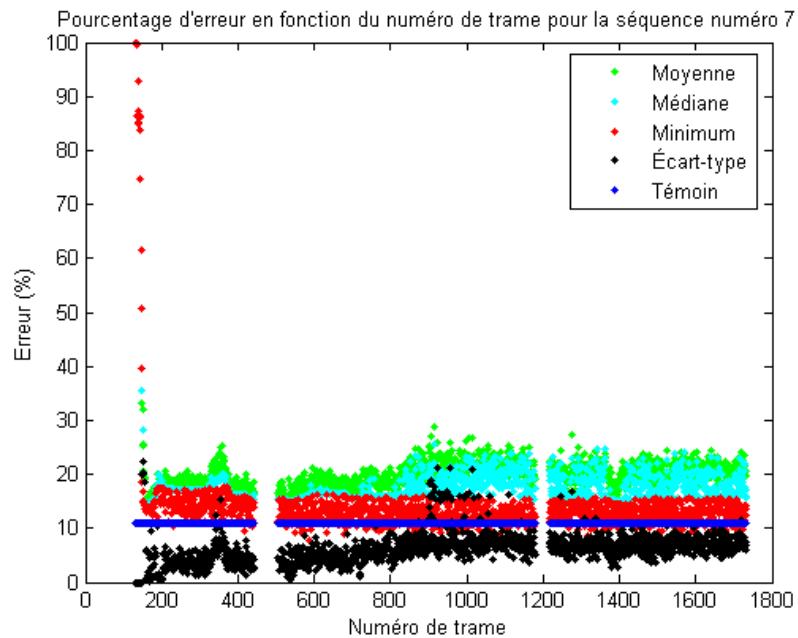


Figure 3.29 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 7

### 3.2.7.3 Analyse des résultats de la séquence

À première vue, on remarque que la méthode de base est plus performante que la nouvelle méthode. En effet, si l'on ne tient pas compte des premières trames où l'algorithme n'est pas stabilisé, la moyenne et la médiane de la méthode de base ne dépassent pas les 30 % d'erreur et se maintiennent la plupart du temps autour de 20 %. Par comparaison, la nouvelle méthode est presque toujours au dessus des 30 % lors des 800 premières trames et la moyenne se stabilise à 30 % à la fin de la séquence. Le graphique des gains de la nouvelle méthode (figure 3.30) le démontre sans équivoque. De plus, notons la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes.

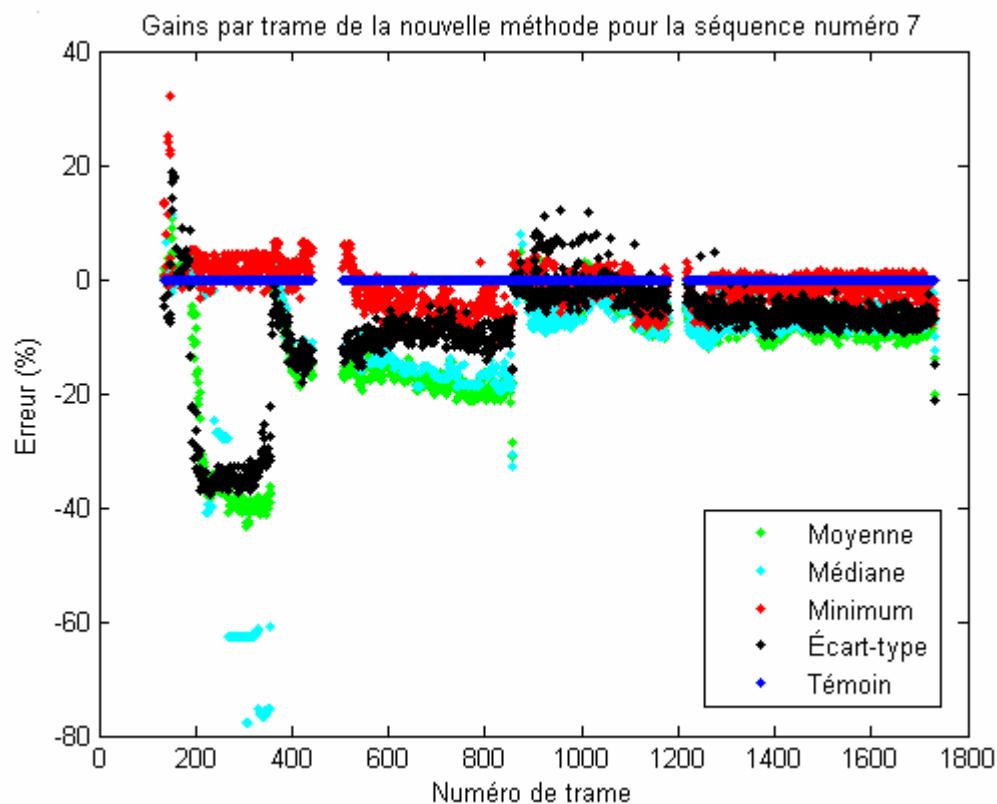


Figure 3.30 Graphique des gains de la nouvelle méthode pour la séquence numéro 7

Cette constatation était prévisible. Rappelons ici que l'utilisation des images en recouvrement avait pour but d'améliorer les performances de l'algorithme de base dans les situations où il n'y avait pas beaucoup de trajectoires et donc que la quantité d'information obtenue de ces trajectoires était restreinte. Dans le scénario présent, la longueur de la séquence et le grand nombre d'acteurs qui entrent et sortent de la scène font en sorte qu'une grande quantité de trajectoires est générée. Ceci fait en sorte que l'algorithme de base n'est plus désavantage. De plus, l'ajout des images de pointage en recouvrement fait en sorte que le nouvel algorithme est plus sensible aux changements instantanés et aux erreurs dans la soustraction d'arrière-plan. Aussi, ces images de pointages deviennent saturées dans les cas où de nombreuses personnes entrent et sortent de la scène. Cette saturation a comme conséquence d'augmenter le degré de liberté ce qui fait en sorte que le recouvrement semble toujours bon, même si la matrice de transformation est moins optimale. Tous ces éléments mis ensemble et les résultats affichés dans les graphiques nous forcent à conclure que le nouvel algorithme est moins bien adapté aux longues séquences très complexes.

### **3.2.8 Séquence vidéo numéro 8**

Cette séquence vidéo a été tournée dans le but de vérifier les performances dans une zone d'inclusion des champs de vision réduite mais un peu plus grande que précédemment. Aussi, le nombre de personnes présentes dans la scène a été augmenté et varié jusqu'à 3. De plus, les acteurs entrent et sortent régulièrement de la scène ce qui fait en sorte que l'on peut simuler ici aussi une séquence plus réelle qui aurait pu être tournée dans un lieu public. Voici d'abord le scénario de la séquence.

#### **3.2.8.1 Description du scénario**

Cette séquence contient jusqu'à trois personnes qui entrent et sortent plusieurs fois aléatoirement dans les champs de vision des 2 caméras. Il serait lourd et fastidieux de décrire le mouvement de chacun. C'est pourquoi il n'est mention ici que des événements reliés à des particularités des graphiques. La séquence commence avec un arrière-plan vide. Une première personne entre dans la zone d'inclusion des champs de vision

environ à la trame 50. Quelques trames plus tard, le réglage automatique de la caméra infrarouge fait en sorte que quelques trames sont perdues ce qui donne l'impression que deux autres personnes font leur apparition dans la scène en même temps approximativement à la trame 70. Le premier homme sort de la zone d'inclusion des champs de vision environ à la trame 90. Il entre de nouveau dans la zone approximativement à la trame 110. Il quitte de nouveau la zone d'inclusion des champs de vision à approximativement la trame 180. Une entrée et une sortie presque simultanée sont observables environ à la trame 250. Une autre sortie et entrée simultanées survient approximativement à la trame 260. Entrée et sortie de 2 personnes environ à la trame 290. Trois personnes sont dans la zone d'inclusion des champs de vision entre les trames 350 et 400. Une fusion de deux « blobs » (occlusion entre deux personnes) se produit approximativement à la trame 450. Une autre fusion de « blobs » survient environ à la trame 520. Un autre problème relié au réglage automatique de la caméra est visible approximativement à la trame 550. Une seule personne est dans la zone d'inclusion des champs de vision à la fin de la séquence. La zone d'inclusion des champs de vision est de 33 % pour cette séquence.

### **3.2.8.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.31 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 8 et la figure 3.32 montre le résultat pour la méthode de base.

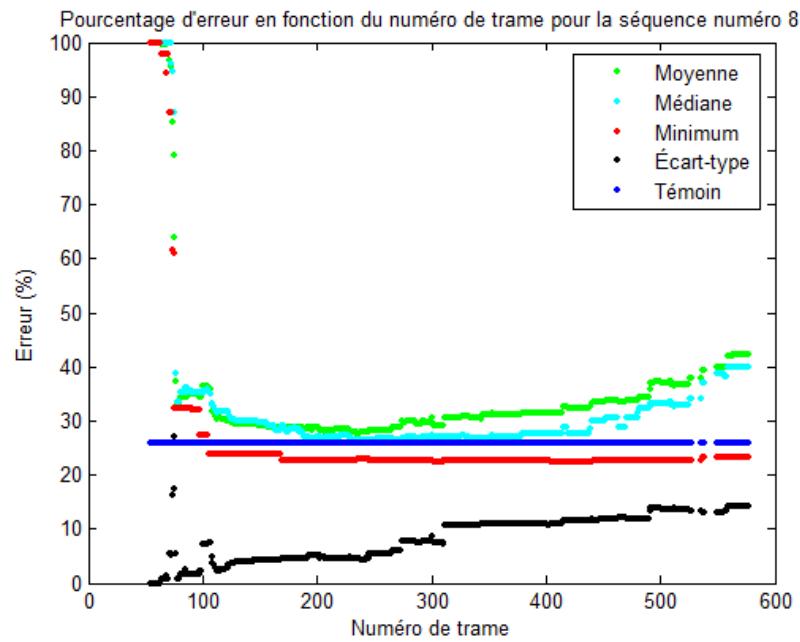


Figure 3.31 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 8

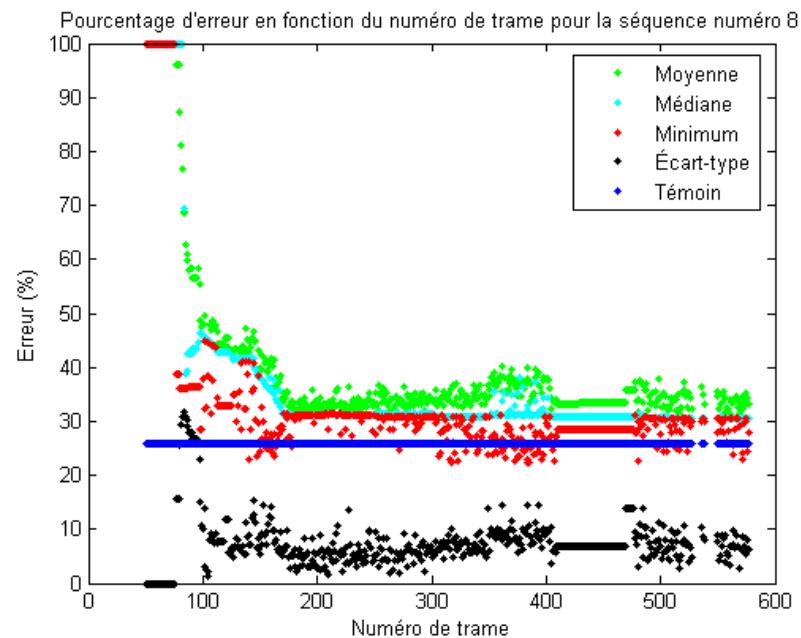


Figure 3.32 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 8

### 3.2.8.3 Analyse des résultats de la séquence

Encore une fois, on remarque des nuages de points dans le graphique de la méthode de base. Une observation intéressante dans ces graphiques est que le nouvel algorithme performe mieux que celui de base dans la première portion de la séquence (entre les trames 75 et 400) alors que la méthode de base surpassé la nouvelle méthode dans la dernière portion (après la trame 400). Le graphique des gains de la nouvelle méthode de la figure 3.33 permet de mieux voir le phénomène. Par la même occasion, notons la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes.

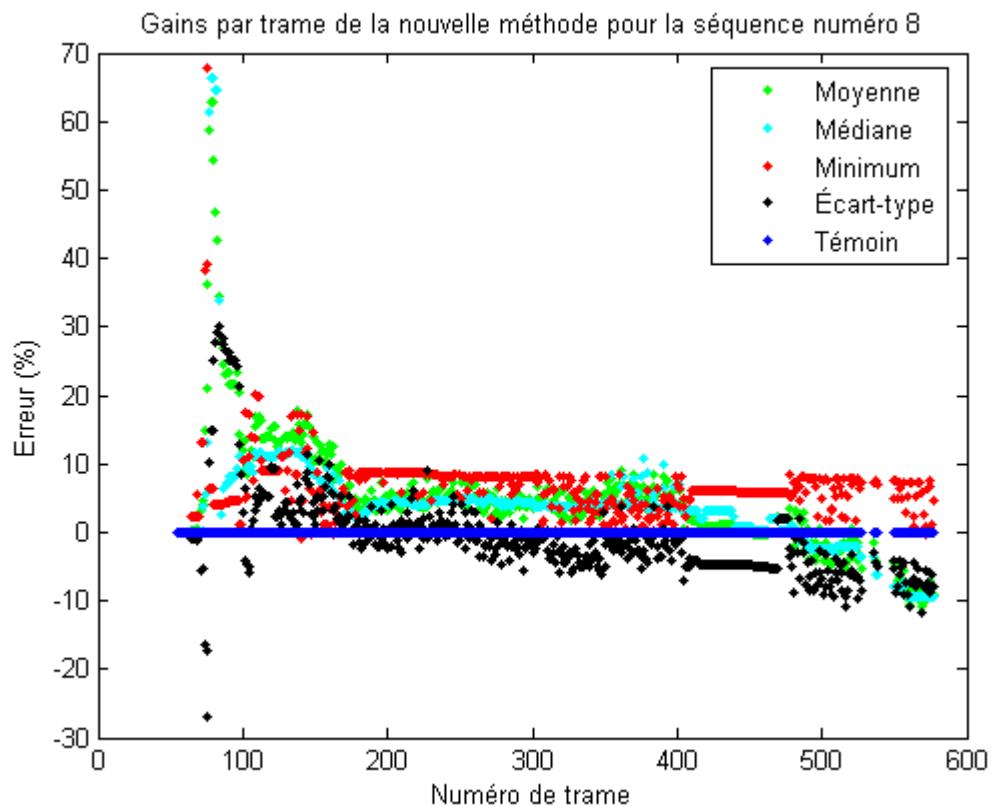


Figure 3.33 Graphique des gains de la nouvelle méthode pour la séquence numéro 8

La cause de ce phénomène a déjà été mentionnée pour la 7e séquence de test. En effet, il a déjà été mentionné que le nouvel algorithme performe moins bien que celui de base

lorsque le nombre de trajectoires croît. Rappelons que la méthode de base utilise toute les trajectoires observées de tous les objets qui ont fait parti de l'avant-plan depuis le début de la séquence. Donc même si le nombre de personnes visibles dans la séquence diminue, le nombre de trajectoire continue quand même d'augmenter ou de stagner. Aussi, des erreurs dans les trajectoires peuvent nuire aux performances de l'algorithme de base. Ceci peut faire en sorte que les performances peuvent être inférieures à certains moments même si le nombre de trajectoire augmente. C'est ce qui se produit entre la trame 300 et 400. Néanmoins, on peut voir que lorsque la séquence atteint une certaine longueur, la méthode de base performe mieux que la nouvelle méthode car la saturation de l'image de pointage augmente ce qui fait en sorte que les performances de la nouvelle méthode se détériorent plus rapidement que celles de la méthode de base et ce malgré les erreurs dans les trajectoires qui s'accumulent.

En observant le nombre de trajectoires qui correspond au début de l'augmentation du pourcentage d'erreur (trame 225) dans le graphique de la nouvelle méthode, on obtient ce qui semble être un moment de performances optimales. Le graphique de la figure 3.34 donne le nombre de trajectoires pour chacun des spectres en fonction du temps ce qui nous permettra d'identifier ce moment.

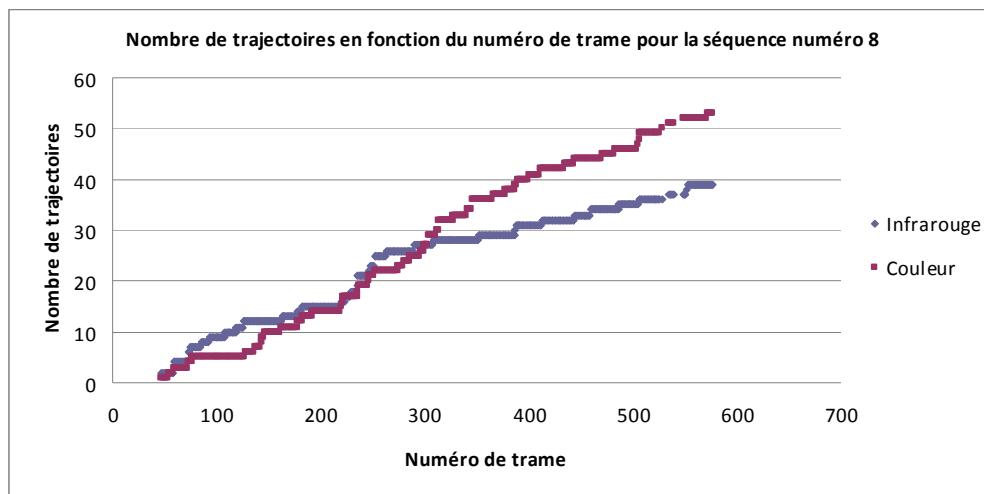


Figure 3.34 Graphique du nombre de trajectoires par trame pour les spectres infrarouge et couleur sur la séquence vidéo numéro 8

On remarque dans ce graphique que ce moment de performances optimales (trame 225) correspond encore une fois à une stabilité du nombre de trajectoires dans les 2 spectres ainsi qu'à un nombre de trajectoires pour les deux spectres presque identiques. Dans l'analyse de la séquence 2, ces deux critères avaient déjà été identifiés comme correspondant à des moments où la soustraction d'arrière-plan est bonne et par conséquent propice à la nouvelle méthode.

Une dernière observation intéressante pouvant être tirée de ces graphiques est reliée à la valeur minimum qui se situe toujours en dessous de la valeur témoin après la trame 100 pour la nouvelle méthode. Comme mentionné précédemment (séquence 3), ceci indique que la restriction sur la coplanarité des objets de la scène est mieux respectée dans les séquences vidéo à zone réduite d'inclusion des champs de vision.

### **3.2.9 Séquence vidéo numéro 9**

Cette séquence vidéo est la plus courte qui est été tourné. C'est aussi la première tournée et c'est principalement elle qui a servi à développer la méthode. C'est probablement une des raisons qui expliquent les bons résultats obtenus. Elle a été incluse car elle démontre une tendance à l'amélioration des performances lorsque le nombre de trajectoires augmente. Voici d'abord le scénario de la séquence.

#### **3.2.9.1 Description du scénario**

La séquence commence avec un arrière-plan vide. Deux personnes apparaissent vers la trame 10. Ils marchent autour des tables présentent dans l'arrière-plan jusqu'à la fin de la séquence. Ils se croisent, ce qui entraîne des fusions de « blobs » (occlusion partielle de deux personnes), approximativement aux trames suivantes : 30, 50, 75 et 80.

#### **3.2.9.2 Présentation des résultats de la séquence**

Le graphique de la figure 3.35 montre les résultats obtenus pour le recalage utilisant la nouvelle méthode développée dans ce travail pour la séquence vidéo numéro 9 et la figure 3.36 montre les résultats avec la méthode de base.

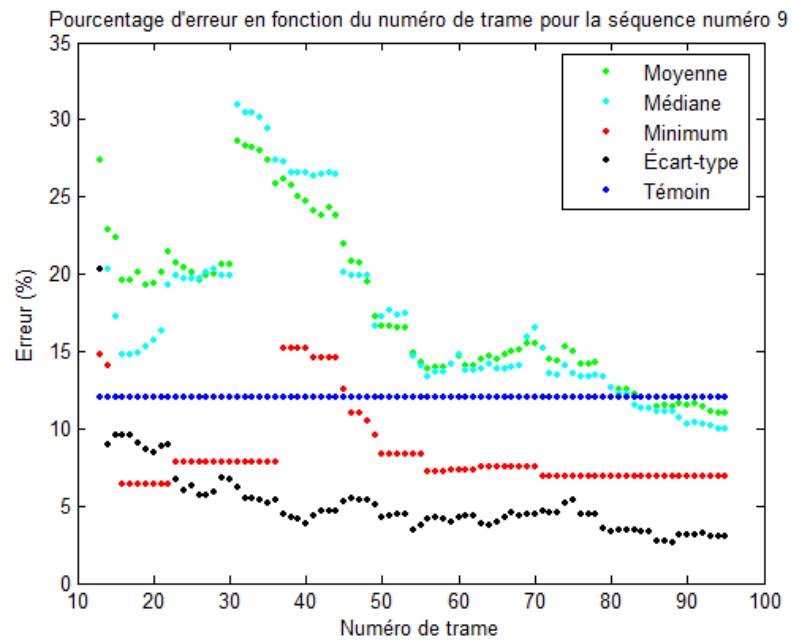


Figure 3.35 Graphique des performances en pourcentage d'erreur de la nouvelle méthode sur la séquence vidéo numéro 9

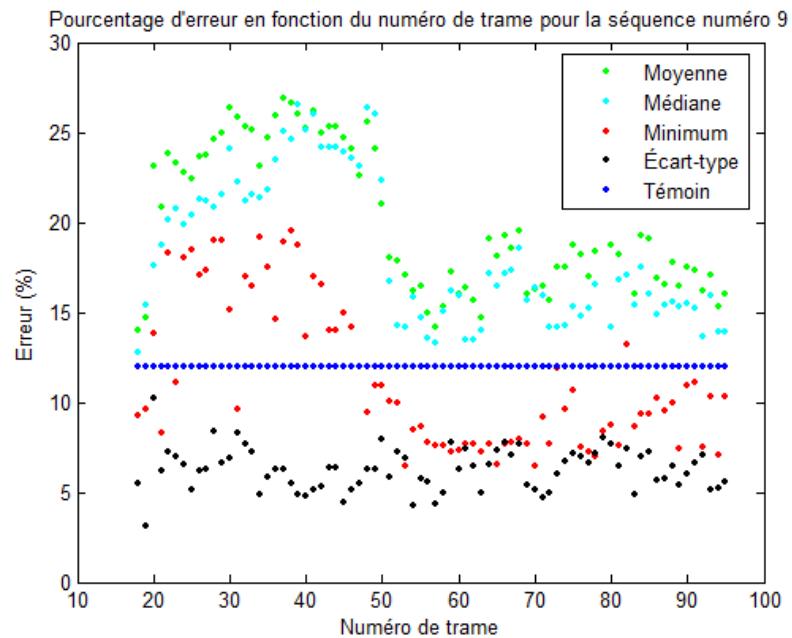


Figure 3.36 Graphique des performances en pourcentage d'erreur de la méthode de base sur la séquence vidéo numéro 9

### 3.2.9.3 Analyse des résultats de la séquence

On remarque encore les nuages de points dans le graphique de la méthode de base. Une autre remarque qui a été faite dans les précédentes analyses est que la valeur minimum du recalage est en dessous de la valeur témoin pour beaucoup de trames dans la séquence. Cependant, il ne faut pas conclure ici qu'il s'agit d'un phénomène relié aux différences de distances entre les objets d'avant-plan et des caméras puisque la zone d'inclusion des champs de vision n'est pas réduite. Cette fois, il faut plutôt regarder du côté de la façon dont les valeurs témoins ont été calculées. Cette séquence est la seule pour laquelle ce n'est pas des cartons refroidis qui ont été utilisés pour calculer les valeurs témoins. Ce sont plutôt les tables présentent dans l'arrière-plan qui ont été utilisée. Malheureusement, ces tables ne sont pas aussi bien visibles que les cartons en infrarouge. L'erreur humaine de sélection des éléments témoins est donc beaucoup plus élevée pour cette séquence. Probablement que dans des conditions idéales, la valeur témoin serait plutôt située près de 5 %.

Pour analyser les performances des deux méthodes, regardons d'abord le graphique des gains de la nouvelle méthode (figure 3.37).

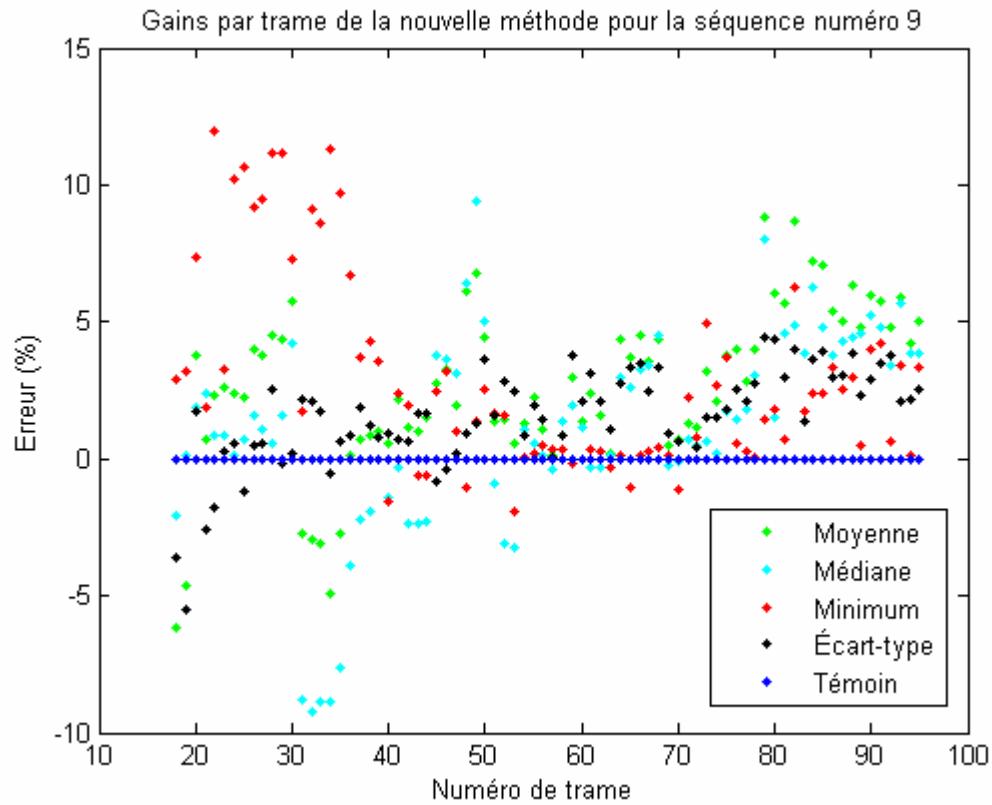


Figure 3.37 Graphique des gains de la nouvelle méthode pour la séquence numéro 9

Pour commencer, notons la confirmation de la supériorité de la nouvelle méthode en début de séquence observée dans les séquences précédentes. Ensuite, il faut rappeler qu'une fusion de « blobs » entraîne l'apparition d'une nouvelle trajectoire. En gardant ce fait en mémoire, on peut conclure que le seul moment où la nouvelle méthode ne surpassé pas la méthode de base, en dehors du début de la séquence où les performances ne sont pas stabilisées, est lorsque le nombre de trajectoires vient de passer à 6 (trame 29) pour la caméra infrarouge. En plus, ce moment ne dure que quelques trames. En observant le graphique du nombre de trajectoires (figure 3.38), il est possible de conclure sur les causes de ces bonnes performances.

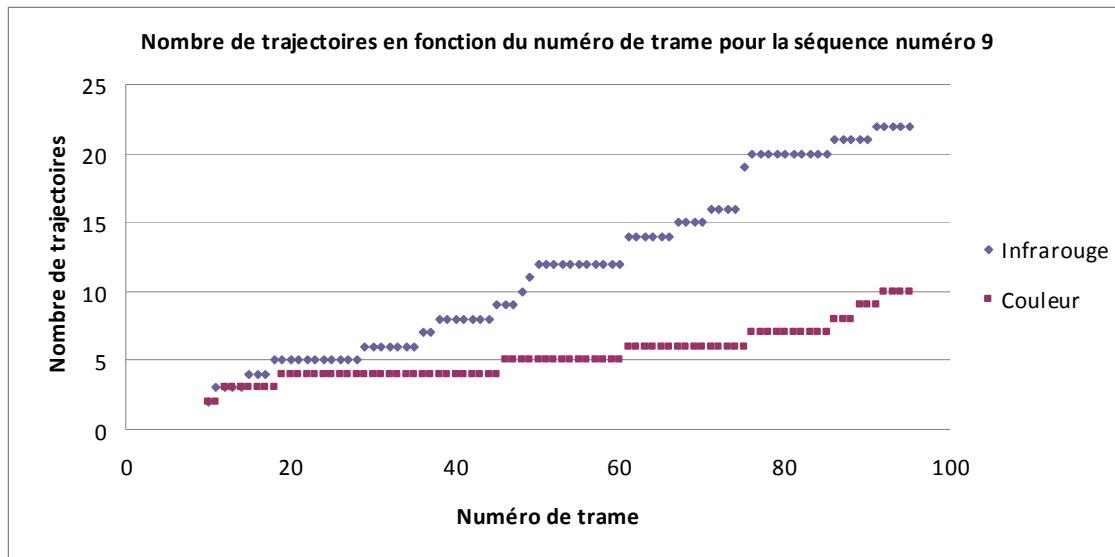


Figure 3.38 Graphique du nombre de trajectoires par trame pour les spectres infrarouge et couleur sur la séquence vidéo numéro 9

Dans ce graphique, on remarque de nombreux plateaux caractéristiques d'une bonne soustraction d'arrière-plan. Aussi, on remarque que ces plateaux sont souvent simultanés pour les deux spectres. Ceci signifie que la soustraction d'arrière-plan est de bonne qualité pour les deux spectres au même moment. Comme conclu précédemment, ces conditions sont idéales pour la nouvelle méthode ce qui explique sa supériorité dans la majeure partie de la séquence vidéo. Pour ce qui est du maximum de performance que l'on observe à la fin de la séquence, il peut être expliqué par le fait que le nombre de trajectoires est plus grand. Ceci pourrait être vue comme une contradiction avec l'observation faite plus tôt qui disait que la méthode de base est meilleure lorsque le nombre de trajectoire est plus grand mais ce n'est pas le cas. En effet, comme la séquence vidéo est courte et peu complexe, le nombre d'erreurs dans les trajectoires n'est pas très grand tout comme la saturation de l'image de pointage. Dans cette situation, une augmentation du nombre de trajectoire est bénéfique pour la nouvelle méthode aussi comme nous l'avons vu dans les séquences précédentes. Par conséquent, la combinaison d'une bonne soustraction d'arrière-plan et d'un assez grand nombre de

trajectoires rend le nouvel algorithme très efficace. Cette conclusion est appuyée par les autres analyses des séquences précédentes.

### **3.3 Analyse des résultats globaux**

Dans les sections précédentes, les résultats ont été analysés trame par trame, séquence par séquence. Cependant, aucune analyse n'a été faite du comportement global de la nouvelle méthode. Même si cette méthode n'est pas parfaite, est-elle, en général, meilleure que la méthode de base? Les prochains tableaux permettent de conclure que oui en compilant différentes statistiques sur les performances globales des 2 méthodes dans chaque séquence vidéo testée. Il sera donc question, dans les prochains paragraphes des données qui contiennent ces tableaux et de leur analyse.

Le premier tableau présenté ci-dessous rassemble différentes statistiques sur les 4 types de données dont il était question dans les graphiques de performance des sections précédentes. Le tableau est découpé en 9 sections horizontales, une pour chaque séquence vidéo testée. Chaque section contient 4 lignes, une pour chaque courbe illustrée dans les graphiques de performances. Pour chacune de ces lignes, on calcule 4 statistiques : la moyenne, la médiane, l'écart-type et la valeur minimum. Ces statistiques sont calculées sur toutes les trames où une donnée est disponible. Par exemple, la moyenne de la courbe moyenne de la méthode de base pour la séquence 1 a été calculée en additionnant toutes les valeurs de la courbe moyenne entre les trames 50 et 250 et en divisant cette somme par le nombre de trames; 200 pour cet exemple. On obtient ainsi la performance moyenne de la moyenne des 30 exécutions pour la méthode de base dans la séquence vidéo numéro 1. Des opérations similaires sont effectuées pour obtenir la médiane, l'écart-type et l'erreur minimum. Ceci nous permet donc de comparer globalement, pour une séquence vidéo donnée, si la nouvelle méthode est supérieure à la méthode de base.

Tableau 3.4 Tableau des données statistiques des performances des 2 méthodes sur la durée totale de chaque séquence vidéo

Numéro de la séquence	Courbe	Moyenne (%)		Médiane (%)		Écart-type (%)		Minimum (%)	
		Base	Nouvelle	Base	Nouvelle	Base	Nouvelle	Base	Nouvelle
1	Moyenne	60,16	32,85	61,36	26,35	31,11	16,76	23,80	21,47
	Médiane	62,77	31,20	69,24	25,57	34,32	16,44	24,22	21,18
	Minimum	32,25	26,71	26,29	19,60	19,62	14,62	17,20	16,66
	Écart-type	11,59	6,38	11,04	4,09	9,87	6,48	0,00	0,02
2	Moyenne	41,63	41,92	32,55	37,78	20,85	16,70	30,81	30,34
	Médiane	41,71	41,72	32,44	37,45	21,32	16,61	30,76	30,05
	Minimum	39,45	40,28	32,44	37,45	19,32	16,86	21,99	30,05
	Écart-type	1,14	2,07	0,35	0,77	3,28	2,54	0,00	0,00
3	Moyenne	35,10	29,42	29,02	26,65	16,09	14,94	19,40	18,51
	Médiane	30,57	28,87	22,58	26,19	16,71	14,97	17,53	18,05
	Minimum	28,02	25,14	18,78	18,97	17,63	15,69	14,83	14,11
	Écart-type	10,88	3,49	9,17	2,41	8,27	4,95	0,00	0,00
4	Moyenne	34,99	27,03	22,52	22,74	26,93	14,10	19,79	16,41
	Médiane	33,32	26,39	21,47	20,90	27,82	14,73	19,84	16,13
	Minimum	29,92	18,38	21,18	16,27	25,03	13,49	14,11	14,01
	Écart-type	5,53	6,46	2,04	5,91	7,61	3,96	0,00	0,00
5	Moyenne	69,98	52,20	69,18	45,89	5,07	11,54	59,49	44,30
	Médiane	70,18	52,03	69,19	45,25	5,06	12,15	59,49	43,26
	Minimum	54,07	40,70	53,68	34,91	8,98	11,55	28,68	34,54
	Écart-type	7,18	6,03	7,44	6,10	1,95	1,70	0,00	0,00
6	Moyenne	30,21	28,43	21,00	20,02	26,58	23,16	15,93	17,45
	Médiane	27,74	26,79	16,68	18,95	27,53	23,56	14,57	16,79
	Minimum	25,29	22,08	14,94	14,15	28,36	24,93	9,87	12,67
	Écart-type	5,79	5,66	6,23	4,72	2,96	3,67	0,00	0,00
7	Moyenne	20,64	32,75	19,80	29,37	7,22	10,99	13,39	13,69
	Médiane	18,29	30,60	17,00	24,79	7,44	16,16	13,25	13,44
	Minimum	13,90	15,18	13,09	14,20	7,78	6,80	7,60	9,19
	Écart-type	6,22	15,22	6,18	13,14	2,84	8,60	0,00	0,00
8	Moyenne	40,39	34,86	34,46	31,17	16,19	13,98	31,27	27,68
	Médiane	37,39	32,75	31,15	28,16	17,11	14,65	30,52	26,34
	Minimum	33,42	26,41	30,22	22,69	16,15	14,71	22,35	22,40
	Écart-type	7,54	8,24	6,92	7,86	4,94	4,19	0,00	0,00
9	Moyenne	20,05	17,71	18,73	15,52	3,95	5,09	13,98	10,99

Médiane	18,31	17,20	16,75	14,84	4,04	5,74	12,75	10,02
Minimum	11,27	8,46	9,78	7,54	4,06	2,58	6,44	6,39
Écart-type	6,25	5,12	6,30	4,44	1,17	2,42	3,16	2,62

D'abord, il est intéressant de constater que les meilleurs résultats des moyennes coïncidentes avec ceux de la moyenne. Ceci signifie que les valeurs extrêmes ne faussent pas les conclusions qui peuvent être tirées des tendances visibles dans le tableau. Une première de ces tendances est que la nouvelle méthode obtient ses meilleures performances dans les séquences courtes et contenant au moins 2 acteurs. C'est le cas pour les séquences numéro 9, 4 et 6 qui se classent respectivement du premier au troisième rang pour la plus basse moyenne des différentes courbes. Ces trois séquences ont des moyennes sous la barre des 30 % pour les courbes moyennes, médianes et minimums. La seule autre séquence qui obtient ces courbes sous la barre des 30 % est la séquence numéro 3. Par contre, il faut mentionner que cette séquence a une zone d'inclusion des champs de vision réduite ce qui fait en sorte que l'hypothèse des objets de l'avant-plan tous situés dans le même plan est beaucoup mieux respecté. Ceci explique fortement les bons résultats obtenus. Néanmoins, on remarque que les résultats de la séquence numéro 4 sont meilleurs ce qui confirme que même avec une zone d'inclusion des champs de vision réduite, l'ajout d'une deuxième personne améliore les performances de la nouvelle méthode.

Pour ce qui est de la méthode de base, elle obtient aussi ses meilleurs résultats de moyenne dans la séquence numéro 9. Cependant, c'est plutôt la séquence 7 qui se pointe au deuxième rang pour ce qui est du classement des séquences avec les meilleurs résultats sur la courbe de la moyenne. La séquence 9 étant une séquence vidéo très courte, il est normal de constater des valeurs moyennes, médianes et minimums plus petites. On le voit d'ailleurs dans le fait que c'est la seule séquence dont les écarts-types sont sous la barre des 10 % pour les 2 méthodes. Ceci est du au fait qu'il y a moins de temps pour générer les valeurs de moins bon recalage souvent observées à la fin des autres séquences plus longues. Rappelons que ces moins bonnes performances sont

engendrées par la saturation de l'image de pointage dans le cas de la nouvelle méthode et pas les erreurs dans les trajectoires pour la méthode de base. Par contre, il est très significatif de constater que de bons résultats de la méthode de base correspondent à la séquence la plus longue et contenant un très grand nombre de trajectoires. En effet, l'article à la base de cette méthode (Caspi et al., 2006) rapporte la méthode comme efficace pour des séquences vidéo relativement longues et contenant un grand nombre des trajectoires. Cette affirmation confirme donc leurs résultats.

On remarque aussi que les pires résultats sont obtenus pour les deux méthodes dans la même séquence vidéo soit la séquence numéro 5. Ceci confirme les observations précédentes de la section 3.2.5.3. Rappelons que ces mauvais résultats avaient été attribués à une différence dans la hauteur des caméras. Comparons maintenant les performances globales des deux méthodes.

Le prochain tableau facilite la comparaison des performances globales des 2 méthodes. Encore une fois, le tableau est divisé en 9 sections qui correspondent aux 9 séquences vidéo testées. Ces sections sont subdivisées en 5 lignes qui correspondent aux différentes courbes des graphiques de performance. Pour ce qui est des valeurs qui se retrouvent dans le tableau, il s'agit de la soustraction des données du tableau précédent entre la méthode de base et la nouvelle méthode. On obtient ainsi une mesure de l'amélioration ou de la détérioration de la nouvelle méthode sur la méthode de base. Une amélioration correspond à un nombre positif et une détérioration à un nombre négatif.

Tableau 3.5 Tableau des gains sur le pourcentage d'erreur de la nouvelle méthode sur la méthode de base

Numéro de la séquence	Courbe	Moyenne	Médiane	Écart-type	Minimum
1	Moyenne	27,31	35,01	14,36	2,32
	Médiane	31,57	43,67	17,89	3,05
	Minimum	5,54	6,70	5,00	0,54
	Écart-type	5,22	6,95	3,39	-0,02
2	Moyenne	-0,29	-5,23	4,15	0,47
	Médiane	-0,02	-5,01	4,71	0,71
	Minimum	-0,83	-5,01	2,46	-8,06
	Écart-type	-0,93	-0,42	0,74	0,00
3	Moyenne	5,68	2,37	1,14	0,88
	Médiane	1,70	-3,61	1,73	-0,53
	Minimum	2,88	-0,19	1,94	0,72
	Écart-type	7,39	6,76	3,32	0,00
4	Moyenne	7,96	-0,22	12,84	3,38
	Médiane	6,93	0,58	13,09	3,71
	Minimum	11,55	4,91	11,54	0,10
	Écart-type	-0,93	-3,87	3,65	0,00
5	Moyenne	17,78	23,29	-6,47	15,19
	Médiane	18,15	23,94	-7,09	16,23
	Minimum	13,37	18,78	-2,57	-5,86
	Écart-type	1,15	1,34	0,26	0,00
6	Moyenne	1,78	0,98	3,42	-1,52
	Médiane	0,95	-2,27	3,96	-2,22
	Minimum	3,22	0,79	3,43	-2,80
	Écart-type	0,13	1,51	-0,71	0,00
7	Moyenne	-12,11	-9,57	-3,77	-0,30
	Médiane	-12,31	-7,79	-8,72	-0,18
	Minimum	-1,28	-1,11	0,98	-1,58
	Écart-type	-9,00	-6,96	-5,76	0,00
8	Moyenne	5,54	3,30	2,21	3,58
	Médiane	4,65	2,99	2,46	4,18
	Minimum	7,01	7,53	1,44	-0,05
	Écart-type	-0,70	-0,94	0,75	0,00
9	Moyenne	2,34	3,22	-1,14	2,99
	Médiane	1,11	1,91	-1,70	2,73
	Minimum	2,81	2,25	1,48	0,05
	Écart-type	1,13	1,87	-1,25	0,54

La première constatation qui peut être faite de ce tableau est que la grande majorité des données sont positives. Par conséquent, il est possible de conclure que la nouvelle méthode est supérieure à celle de base dans la grande majorité des séquences filmées. Cette conclusion est particulièrement supportée par les moyennes, médianes et erreurs minimums obtenues pour les différentes courbes. Cette supériorité est constatée pour les séquences 1, 3, 4, 5, 6, 8 et 9. Dans ces séquences, on remarque cependant quelques exceptions.

Dans la séquence 6, on remarque une légère détérioration de la médiane. Cependant, on remarque aussi que les différents gains des autres courbes ne sont pas très élevés. Aussi, on remarque que les différentes valeurs minimums sont toutes légèrement inférieures à la méthode de base. Toutes ces informations nous permettent de conclure que les performances de la nouvelle méthode sont très comparables à celles de la méthode de base pour la séquence numéro 6. Néanmoins, on constate une légère amélioration des performances, car la médiane donne habituellement la meilleure mesure sur les performances générales.

Par contre, dans les séquences 2 et 7, la plupart des moyennes, médianes et erreurs minimums sont négatives et donc inférieures à la méthode de base. On peut donc en conclure que les scénarios de ces deux séquences sont mieux adaptés à la méthode de base. Les raisons qui expliquent cette conclusion ont déjà été abordées en détail dans les sous-sections précédentes qui analysaient les résultats de ces deux séquences.

Les gains observés dans les écarts-types indiquent plutôt une plus grande stabilité inter-trame des différentes courbes de la nouvelle méthode. Cette conclusion confirme les précédentes observations à ce sujet des analyses des sections précédentes.

Finalement, les résultats globaux confirment les mêmes observations faites dans les graphiques de performances et l'analyse trame par trame qui y a été faite.

## CONCLUSION

Le projet présenté ici consistait à effectuer le recalage de deux séquences vidéo de spectres différents (infrarouge et couleur) et ce, trame par trame. Pour y arriver, les trajectoires des objets en mouvement dans la scène ainsi que les images d'avant-plan ont été utilisées.

Il a d'abord été question des différentes techniques de recalage multispectral dans la littérature dans le chapitre 1. On y montre que ce ne sont pas toutes les techniques qui permettent d'effectuer efficacement un recalage infrarouge-visible. Néanmoins, on y mentionne une technique qui constitue la base de ce projet. Cette technique utilise les trajectoires des objets en mouvement pour trouver la matrice de transformation entre une caméra couleur et une caméra infrarouge. C'est aussi dans ce chapitre qu'on fait mention de différents types de matrice de transformation qu'il est possible d'utiliser pour effectuer un recalage.

Il a ensuite été question des détails d'implantation de cette nouvelle méthode dans le chapitre 2. Rappelons que cette méthode est principalement basée sur deux algorithmes RANSAC. Le premier est utilisé pour trouver la meilleure matrice de transformation affine à partir des points d'une paire de trajectoires. Le deuxième est utilisé pour trouver quelles sont les trajectoires qui sont bien appariées. La meilleure matrice de transformation est celle qui permet d'apparier le plus grand nombre de trajectoires. Le critère utilisé pour déterminer si les trajectoires sont bien appariées et aussi pour éliminer les points aberrants lors du calcul de la matrice de transformation est basé sur les images de pointages. Ces images sont composées d'au plus 5 images d'avant-plan qui tentent de simuler l'effet des trajectoires. On minimise le pourcentage d'erreur du recouvrement de ces images de pointage lorsqu'on veut comparer deux matrices de transformation.

Dans le chapitre 3, une analyse des performances de cet algorithme est faite. On compare le pourcentage d'erreur de recouvrement à chaque trame du nouvel algorithme avec le pourcentage d'erreur de recouvrement d'une méthode dite de base qui

correspond à celle mentionnée dans Caspi et al. (2006), mais aussi à un recalage manuel. Les deux méthodes ont été testées sur 9 séquences vidéo différentes. Les conclusions qui peuvent être tirées sont les suivantes :

- La nouvelle méthode est plus stable entre deux trames que la méthode de base;
- En début de séquence, donc lorsque le nombre de trajectoire est petit, la nouvelle méthode surpassé la méthode de base.
- Lorsque la soustraction d'arrière-plan et la génération de trajectoires sont bonnes, la nouvelle méthode surpassé celle de base;
- Dans les situations où la zone d'inclusion des champs de vision est restreinte, les deux méthodes surpassent à certains moments le recalage manuel;
- La nouvelle méthode à une moyenne et une médiane d'erreur totale plus faible que la méthode de base dans 7 des 9 séquences vidéo testées;
- La méthode de base est quelque peu supérieure à la nouvelle méthode dans les situations générant plusieurs trajectoires et beaucoup d'erreurs de soustraction d'arrière-plan.

De manière générale, on peut résumer ainsi les situations qui avantageant la nouvelle méthode de recalage présentée dans ce texte : lorsque la soustraction d'arrière-plan peut être bien effectuée, que les trajectoires sont claires et que le nombre de trajectoire n'est pas très élevé; le nouvel algorithme de recalage performe mieux que l'algorithme de base ne tenant compte que des trajectoires. Par contre, dans des situations avec un grand nombre de trajectoires (comme des séquences de plus longues durées) et où la soustraction d'arrière-plan est difficile, l'algorithme de base semble un peu mieux adapté. Par contre, il est important de rappeler que notre approche de recalage n'a pas nécessairement besoin d'être appliquée à chaque trame. Le calcul d'une matrice à chaque trame n'a été fait ici que pour vérifier précisément l'évolution des performances. Notons aussi que les deux méthodes de recalage automatique étaient très gourmandes en temps de calcul. Ainsi donc, une implantation de ces méthodes dans un système réel de

surveillance vidéo ne pourrait probablement pas se permettre de calculer une matrice à chaque trame sur une longue période. De plus, la soustraction d'arrière-plan est en général plutôt facile dans le genre de séquence obtenue dans un système de vidéo surveillance. Sachant cela, notre algorithme semble donc mieux adapté à une application dans domaine de la surveillance vidéo.

De ces conclusions, ainsi que des précédents chapitres, il est maintenant possible d'analyser l'atteinte des objectifs qui avaient été fixés dans l'introduction. Pour atteindre le premier objectif, il fallait réussir à développer une méthode de comparaison de qualité de recalage en se servant des images d'arrière-plan. La méthode utilisée et expliquée à la section 3.1.2 a effectivement permis de comparer les recalages de la méthode de base, les recalages manuels et les recalages de la nouvelle méthode. Aussi, cette méthode de comparaison était bel et bien basée sur les images d'arrière-plan. Le deuxième objectif concernait aussi cette méthode de comparaison. Il fallait que les éléments choisis de l'arrière-plan soient visibles dans les deux spectres. Nous avons utilisés des cartons de couleurs avec une température différente du reste de la scène pour la majorité des séquences tests et des tables dans une autre séquence test. Les figures 3.3 et 3.4 de la section 3.1.2 démontrent clairement que cet objectif a été atteint. Le dernier objectif mentionnait qu'il fallait obtenir un plus petit pourcentage d'erreur avec la nouvelle méthode qu'avec la méthode de base dans les situations où le nombre de trajectoires est petit. C'est effectivement le cas comme le montrent tous les graphiques des gains en pourcentage d'erreur de la nouvelle méthode au chapitre précédent. En effet, on observe des gains pour la nouvelle méthode à chaque début de séquence. Le début des séquences correspond au moment où le nombre de trajectoire est encore relativement petit ce qui confirme l'atteinte du dernier objectif fixé.

On dénombre donc deux principales contributions dans ce projet. La première a été de développer une méthode de recalage de séquences vidéo infrarouge et visible qui combine l'utilisation des trajectoires des objets en mouvement avec les images en recouvrement. L'autre contribution de ce projet fut d'utiliser un pourcentage d'erreurs

sur des images en recouvrement obtenues à partir d'un recalage manuel pour mesurer la qualité du recalage obtenu.

### Travaux futurs

La méthode développée réussit dans la plupart des cas à trouver au moins une approximation de la matrice de transformation. Cependant, elle ne converge pas nécessairement vers la meilleure solution avec le temps. Il serait donc intéressant d'affiner la méthode de sélection de la meilleure matrice entre chaque trame.

Une autre faiblesse de la méthode développée est ses performances lorsque la soustraction d'arrière-plan et la génération des trajectoires sont défectueuses. Une façon d'y remédier serait d'utiliser la même approche mais avec des méthodes de soustractions d'arrière-plan et de génération de trajectoires supérieures. La littérature regorge de ce genre de méthode.

Les algorithmes plus avancés de génération de trajectoires sont généralement limités par la nature des caméras qu'elles utilisent. L'utilisation de notre méthode de recalage permettrait de surmonter ces difficultés. Par exemple, différents types d'algorithme de suivis ont des difficultés dans des situations précédant et suivant des fusions de « blobs ». Notre algorithme pourrait être utilisé dans ces situations pour identifier les « blobs » d'avant et d'après la fusion.

## RÉFÉRENCES

- Bhanu, B., & Ju, H. (2007). Fusion of color and infrared video for moving human detection. *Pattern Recognition*, 40(6), 1771-1784.
- Brown, L. G. (1992). A survey of image registration techniques. *Computing Surveys*, 24(4), 325-376.
- Canny, J. (1986). A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-8(6), 679-698.
- Caspi, Y., Simakov, D., & Irani, M. (2006). Feature-Based Sequence-to-Sequence Matching. *International Journal of Computer Vision*, 68(1), 53-64.
- Chen, H.-m., Varshney, P. K., & Slamani, M. A. (2003). *On registration of regions of interest (ROI) in video sequences*, Miami, FL, USA.
- Coiras, E., Santamaria, J., & Miravet, C. (2000). Segment-based registration technique for visual-infrared images. *Optical Engineering*, 39(1), 282-289.
- Cover, T. M., & Thomas, J. A. (2006). *Elements of information theory* (2nd ed.). Hoboken, N.J.: Wiley-Interscience.
- Elbakary, M. I., & Sundareshan, M. K. (2007). Multi-modal image registration using local frequency representation and computer-aided design (CAD) models. *Image and Vision Computing*, 25(5), 663-670.
- Fischler, M. A., & Bolles, R. C. (1981). Random Sample Consensus: A Paradigm for Model Fitting With Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6), 381-395.
- Forsyth, D. A., & Ponce, J. (2003). *Computer Vision A modern approach*. Upper Saddle River: Prentice Hall.
- Fuentes, L. M., & Velastin, S. A. (2006). People tracking in surveillance applications. *Image and Vision Computing*, 24(11), 1165-1171.

- Hartley, R., & Zisserman, A. (2002). *Multiple View Geometry in computer vision* (1st ed.). Cambridge.
- Hild, M., & Umeda, G. (2005). *Image registration in stereo-based multi-modal imaging systems*, Zagreb, Croatia.
- Hu, M.-K. (1962). Visual pattern recognition by moment invariants. *Institute of Radio Engineers Transactions on Information Theory*, IT-8(2), 179-187.
- Huang, X., & Chen, Z. (2002). *A wavelet-based multisensor image registration algorithm*, Beijing, China.
- Huttenlocher, D. P., Huttenlocher, D. P., Klanderman, G. A., & Rucklidge, W. J. (1993). Comparing images using the Hausdorff distance  
Comparing images using the Hausdorff distance. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 15(9), 850-863.
- Intel. (2006). Open Source Computer Vision Library. Retrieved 22 janvier 2007, from <http://www.intel.com/technology/computing/opencv/index.htm>
- Joo, J. W., Choi, J. W., & Cho, D. L. (2003). *Robust registration in two heterogeneous sequence images on moving objects*  
*Robust registration in two heterogeneous sequence images on moving objects*. Paper presented at the Information Fusion, 2003. Proceedings of the Sixth International Conference of.
- Kern, J. P., & Pattichis, M. S. (2007). Robust multispectral image registration using mutual-information models. *IEEE Transactions on Geoscience and Remote Sensing*, 45(5), 1494-1505.
- Kong, S. G., Jingu, H., Bougħorbel, F., Yue, Z., Abidi, B. R., Koschan, A., et al. (2007). Multiscale fusion of visible and thermal IR images for illumination-invariant face recognition. *International Journal of Computer Vision*, 71(2), 215-233.

- Krotosky, S. J., & Trivedi, M. M. (2007). Mutual information based registration of multimodal stereo videos for person tracking. *Computer Vision and Image Understanding*, 106(2-3), 270-287.
- Kyoung Soo, K., Jae Hak, L., & Jong Beom, R. (2005). *Robust multi-sensor image registration by enhancing statistical correlation*, Philadelphia, PA, USA.
- Laganière, R. (2009). *Epipolar geometry, The fundamental matrix and the tensor*. Paper presented at the Canadian Conference on Computer and Robot Vision.
- Lourakis, M. (2007). levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++ (Version 2.1.2). Heraklion: Institute of Computer Science, Foundation for Research and Technology - Hellas.
- McKenna, S. J., Jabri, S., Duric, Z., Rosenfeld, A., & Wechsler, H. (2000). Tracking groups of people. *Computer Vision and Image Understanding*, 80(1), 42-56.
- Ostle, B. (1996). *Engineering statistics : the industrial experience*. Belmont ; Montréal: Duxbury Press.
- Shoushtarian, B., & Bez, H. E. (2005). A practical adaptive approach for dynamic background subtraction using an invariant colour model and object tracking. *Pattern Recognition Letters*, 26(1), 5-26.
- Sobel, I. (1978). Neighborhood coding of binary images for fast contour following and general binary array processing. *Computer Graphics and Image Processing*, 8(1), 127-135.
- St-Onge, P.-L. (2008). *Appariement de points caractéristiques trouvés à même les régions d'avant-plan de vidéos à spectres visible et infrarouge*. Unpublished M.Sc.A., Ecole Polytechnique, Montreal (Canada).
- Stone, H. S., & Wolfov, R. (2002). Blind cross-spectral image registration using prefiltering and Fourier-based translation detection. *IEEE Transactions on Geoscience and Remote Sensing*, 40(3), 637-650.

- Wang, J.-G., Sung, E., & Venkateswarlu, R. (2004). *Registration of infrared and visible-spectrum imagery for face recognition*, Seoul, South Korea.
- Wikipedia (2008). Mutual information. *Journal*. Retrieved from [http://en.wikipedia.org/w/index.php?title=Mutual\\_information&oldid=190932223](http://en.wikipedia.org/w/index.php?title=Mutual_information&oldid=190932223)
- Wikipedia (2009). Bayer filter. *Journal*. Retrieved from [http://en.wikipedia.org/w/index.php?title=Bayer\\_filter&oldid=270560623](http://en.wikipedia.org/w/index.php?title=Bayer_filter&oldid=270560623)
- Zitova, B., & Flusser, J. (2003). Image registration methods: a survey. *Image and Vision Computing*, 21(11), 977-1000.