

Université de Montréal

Construction d'une mosaïque à partir d'une séquence vidéo

Par

Ali Abbas

Département d'informatique et
de recherche opérationnelle

Faculté des arts et sciences

Mémoire présenté à la Faculté des études supérieures
en vue de l'obtention du grade de
Maître ès sciences (M. Sc.)
en informatique

août 2002

©Ali Abbas, 2002

Université de Montréal

Faculté des études supérieures

Ce mémoire intitulé :

Construction d'une mosaïque à partir d'une séquence vidéo

Présenté Par :

Ali Abbas

A été évalué par un jury composé des personnes suivantes :

Président rapporteur :	Max Mignotte
Directeur de recherche :	Sébastien Roy
Membre du jury :	Pierre Poulin

Mémoire accepté le :

Résumé

L'objet d'étude de cette présente thèse consiste aux étapes de la construction d'une mosaïque d'images, problème important du domaine de la vision par ordinateur. Nous présentons une étude des méthodes existantes qui traite de toutes les étapes de construction d'une mosaïque. Nous proposons une nouvelle approche pour la mise en correspondance en utilisant une technique de comparaison par triangles.

En effet, la construction d'une mosaïque se fait en quatre étapes: la détection des points d'intérêt, la mise en correspondance, la calibration de la caméra et la projection des images sur un plan cylindrique.

Pour accomplir la première étape consistant à extraire les points d'intérêt, nous avons adopté l'algorithme de Harris. Le choix d'un seuil pour sélectionner les points d'intérêt est un problème difficile. Pour améliorer la distribution des points extraits dans l'image, nous proposons une technique de « seuil adaptatif ». Durant la deuxième étape, nous proposons notre approche de mise en correspondance, utilisant les points d'intérêt obtenus dans la première étape, pour identifier les paires de la mise en correspondance. Pour atteindre ce but, nous élaborons une méthode appelée « méthode de comparaison par triangles » qui ajoute une dimension « géométrique » au choix des paires de points correspondants. Durant l'étape de calibration, nous implantons l'algorithme de la calibration affine « estimation directe ». Finalement nous combinons plusieurs images auxquelles nous avons appliqué les trois étapes précédentes pour obtenir la mosaïque en utilisant la projection cylindrique.

Mots-clé : points d'intérêt, mise en correspondance, calibration, homographie, mosaïque.

Abstract

In this thesis, we study the stages of the construction of an image mosaic, an important problem in the field of computer vision. First, we evaluate the existing methods, which deal with all the stages of construction of a mosaic. Next, we propose a new approach for the mapping by using the technique of comparison by triangles. The construction of a mosaic is done in four stages: the detection of the points of interest, the correspondence mapping, the calibration of the camera and the projection of the images on a cylindrical projection plane. To achieve the first stage which consists in extracting the points of interest, we adopted the algorithm of Harris. The choice of a threshold value to select the points of interest is a difficult problem. To improve the distribution of the points extracted in the image, we propose a technique of "adaptive thresholding". In the second stage, we propose our approach of mapping that uses the points of interest obtained in the first stage to identify the pairs of the mapping. To achieve this goal, we work out a method called "method of comparison by triangles" which adds a "geometrical" dimension to the choice of the pairs of corresponding points. Finally we combine several images processed through the three preceding stages to obtain the mosaic by using cylindrical projection.

Key words: points of interest, mapping, calibration, homography, mosaic.

Tables des matières

REMERCIEMENTS	6
1 INTRODUCTION	7
1.1 MOTIVATION	7
1.2 CONTRIBUTIONS	9
1.3 PLAN DU MÉMOIRE	10
2 MÉTHODES ACTUELLES	11
2.1 DIFFÉRENTES MÉTHODES DE RECONSTRUCTION D'UNE MOSAÏQUE	11
2.2 RECONSTRUCTION DE MOSAÏQUE.....	12
2.2.1 <i>Détection des points d'intérêt</i>	12
2.2.2 <i>Mise en correspondance</i>	18
2.2.3 <i>Calibration des caméras</i>	22
3 POINTS D'INTÉRÊT	26
3.1 NOTRE CHOIX DE MÉTHODE (ALGORITHME DE HARRIS).....	26
3.2 L'ALGORITHME DE DÉTECTION DES POINTS D'INTÉRÊT.....	27
3.2.1 <i>Le seuil</i>	27
3.2.2 <i>Filtrage et gradient par convolution</i>	30
3.3 RÉSULTATS.....	31
3.4 DISCUSSION DE RÉSULTATS OBTENUS.....	36
4 MISE EN CORRESPONDANCE	38
4.1 MISE EN CORRESPONDANCE PAR CORRÉLATION	38
4.1.1 <i>Mesures de corrélation</i>	39
4.1.2 <i>Difficultés liées à l'utilisation de la corrélation</i>	40
4.2 MISE EN CORRESPONDANCE FINALE	42
4.2.1 <i>La méthode des triangles</i>	43
4.2.2 <i>Extraction des paires de référence</i>	43
4.2.3 <i>Explication avec un exemple</i>	45
4.2.4 <i>Exemples de la recherche des points de référence</i>	48
4.2.5 <i>Comparaison par triangulations</i>	54
4.2.6 <i>Explication par un exemple</i>	55
4.2.7 <i>Exemples de la méthode du triangle</i>	57
4.2.8 <i>Échec de la méthode</i>	62
4.2.9 <i>Avantage de la méthode</i>	63
5 CALIBRATION ET CONSTRUCTION DE MOSAÏQUE	64
5.1 CALIBRATION ET HOMOGRAPHIE	64
5.1.1 <i>Modèle géométrique d'une caméra</i>	64
5.1.2 <i>Paramètres d'une caméra</i>	66
5.1.3 <i>Calibration d'un modèle affine</i>	69
5.2 CONSTRUCTION DE LA MOSAÏQUE.....	71
6 RÉSULTATS ET DISCUSSION	73
7 CONCLUSION	84
BIBLIOGRAPHIE	86

Tables de figures

FIGURE 1.1: UNE VUE PANORAMIQUE (16 IMAGES).....	7
FIGURE 1.2: EXEMPLE POUR DONNER L'IDÉE DE NOTRE TRAVAIL.....	9
FIGURE 2.1: DIFFÉRENTS TYPES DE POINTS D'INTÉRÊT.....	12
FIGURE 2.2: DEUX POINTS DE MISE EN CORRESPONDANCE.....	18
FIGURE 3.1: EXEMPLE DU DÉTECTEUR DE POINTS D'INTÉRÊT EN UTILISANT LES TROIS SEUILS ..	31
FIGURE 3.2: EXEMPLE DU DÉTECTEUR DE POINTS D'INTÉRÊT EN UTILISANT LES TROIS SEUILS ..	33
FIGURE 3.3: EXEMPLE DU DÉTECTEUR DE POINTS D'INTÉRÊT EN UTILISANT LES TROIS SEUILS ..	35
FIGURE 4.1: RECHERCHE D'UN LIEN PAR CORRÉLATION.....	39
FIGURE 4.2: LA VALEUR DES POINTS DÉTECTÉS.....	42
FIGURE 4.3: PREMIÈRE ÉTAPE DE LA MÉTHODE DU TRIANGLE.....	46
FIGURE 4.4: EXEMPLE1-A EN UTILISANT LES DEUX CRITÈRES.....	48
FIGURE 4.5: EXEMPLE1-B DEUX POINTS DE RÉFÉRENCE AVEC UN SEUL CRITÈRE.....	49
FIGURE 4.6: EXEMPLE 2 EN UTILISANT LES DEUX CRITÈRES.....	50
FIGURE 4.7: EXEMPLE 3 DEUX POINTS DE RÉFÉRENCE.....	51
FIGURE 4.8: EXEMPLE 4 DEUX POINTS DE RÉFÉRENCE EN UTILISANT LES DEUX CRITÈRES.....	52
FIGURE 4.9: EXEMPLE 5 DEUX POINTS DE RÉFÉRENCE EN UTILISANT LES DEUX CRITÈRES.....	53
FIGURE 4.10: EXEMPLE 5 DEUX POINTS DE RÉFÉRENCE EN ANNULANT LA COMPARAISON DES ANGLES.....	53
FIGURE 4.11 : ILLUSTRATION DES PAIRES DE CORRESPONDANCE ERRONÉES.....	54
FIGURE 4.12: ILLUSTRATION DE LA MÉTHODE DE LA TRIANGULATION PAR UN EXEMPLE.....	56
FIGURE 4.13: EXEMPLE 1 MISE EN CORRESPONDANCE PAR TRIANGULATION AVEC UN SEUIL FIXE	57
FIGURE 4.14: EXEMPLE 2 MISE EN CORRESPONDANCE PAR TRIANGULATION AVEC UN SEUIL FIXE	58
FIGURE 4.15: EXEMPLE 3 MISE EN CORRESPONDANCE PAR TRIANGULATION AVEC UN SEUIL FIXE	59
FIGURE 4.16: EXEMPLE 4 MISE EN CORRESPONDANCE PAR TRIANGULATION AVEC UN SEUIL FIXE	60
FIGURE 4.17: EXEMPLE MISE EN CORRESPONDANCE AVEC UN SEUIL ADAPTATIF.....	61
FIGURE 4.18: VALEURS DES SCORES.....	63
FIGURE 5.1: MODÈLE GÉOMÉTRIQUE D'UNE CAMÉRA.....	65
FIGURE 6.1: EXEMPLE DE CONSTRUCTION DE LA MOSAÏQUE AVEC UN SEUIL MOYEN.....	74

FIGURE 6.2: EXEMPLE DE CONSTRUCTION DE LA MOSAÏQUE AVEC UN SEUIL MOYEN.....	75
FIGURE 6.3: EXEMPLE DE CONSTRUCTION DE LA MOSAÏQUE AVEC UN SEUIL MOYEN.....	77
FIGURE 6.4: EXEMPLE DE CONSTRUCTION DE LA MOSAÏQUE AVEC UN SEUIL MOYEN.....	80
FIGURE 6.5: EXEMPLE DE LA CONSTRUCTION D'UNE MOSAÏQUE AVEC UN SEUIL ADAPTATIF.....	83

Remerciements

Mes grands remerciements reviennent à Monsieur Sébastien Roy, professeur à l'Université de Montréal, pour m'avoir donné l'occasion de faire partie du groupe *vision et modélisation géométrique*, ainsi que pour avoir supervisé ce travail.

Je remercie également tous les membres de ma famille.

Chapitre 1 :

1 Introduction

1.1 Motivation

La nécessité de combiner des images dans un panorama existe depuis le début de la photographie, car le champ visuel d'une caméra est souvent plus petit que le champ visuel humain. De plus, les objets de grandes tailles ne peuvent pas être capturés dans une image simple.

L'utilisation d'un champ visuel large peut être une solution partielle, mais les images obtenues avec un tel champ visuel ont des distorsions substantielles.

La construction d'une mosaïque consiste à recoller toutes les images de référence dans le même repère, de façon à reconstituer une seule vue d'ensemble de la scène.

Pour reconstituer un panorama à partir d'un grand nombre d'images, étudions tout d'abord, l'association de deux images qui se recouvrent partiellement et ensuite, appliquons les principes en question de plus près pour l'ensemble des images d'origine.



Figure 1.1: Une vue panoramique (16 images).

Pour la construction automatique de la mosaïque, les étapes suivantes sont nécessaires :

- Détection des points d'intérêt
- Corrélation automatique
- La mise en correspondance
- Approximation de l'homographie correspondante
- Formation du panorama.

Les points d'intérêt sont des zones dans l'image où l'intensité lumineuse varie beaucoup. Un point d'intérêt correspond à un changement bidimensionnel du signal. Par exemple, les coins, les jonctions en T et les endroits où la texture varie fortement.

La mise en correspondance est un des problèmes difficiles de base de la vision. Dans le but de déterminer la transformation qui envoie les points de la première image sur ceux de la deuxième, il faut, dans le cas idéal, connaître les coordonnées des points associés. Alors, pour chaque point dans l'image de référence, on cherche sa correspondance avec l'ensemble des points d'intérêt de la deuxième image en utilisant une fonction de corrélation.

De manière générale, la technique de recherche d'une correspondance par corrélation se déroule de la façon suivante. On utilise une fenêtre fixe $F1$ dans l'image A centrée sur un point $p1$, ainsi qu'une fenêtre glissante $F2$ dans l'image B qui va parcourir la zone de recherche. Pour chaque position de $F2$, on calcule une valeur de corrélation entre $F1$ et $F2$. Le point choisi sera le point qui a la valeur le plus élevé.

Les difficultés liées à l'utilisation de la corrélation sont:

- Images très bruitées
- Différences de luminosité
- Occlusions
- Zones uniformes

Deux images provenant de la vue panoramique d'une scène à partir d'un même point ou provenant d'une scène prise depuis plusieurs points différents, sont liées par une transformation appelée homographie. Les points de mise en correspondance seront utilisés pour le calcul de l'homographie.

Il y a plusieurs techniques pour traiter l'homographie, et elles seront présentées à la section 2.2.3.

Chaque point du panorama final est une projection d'un point de l'ensemble des images originales en tenant compte que celui-ci peut être projeté sur plusieurs images en même temps.

1.2 Contributions

La contribution principale de cette thèse est d'avoir développé une nouvelle méthode de mise en correspondance qui est basée sur la notion de triangles.

On dit que trois points de la première image (triangle1) correspondent aux trois points de la deuxième image (triangle2), si les deux triangles sont idéalement identiques et si leurs angles ont les mêmes signes par rapport à l'axe des x (voir figure 1.2).

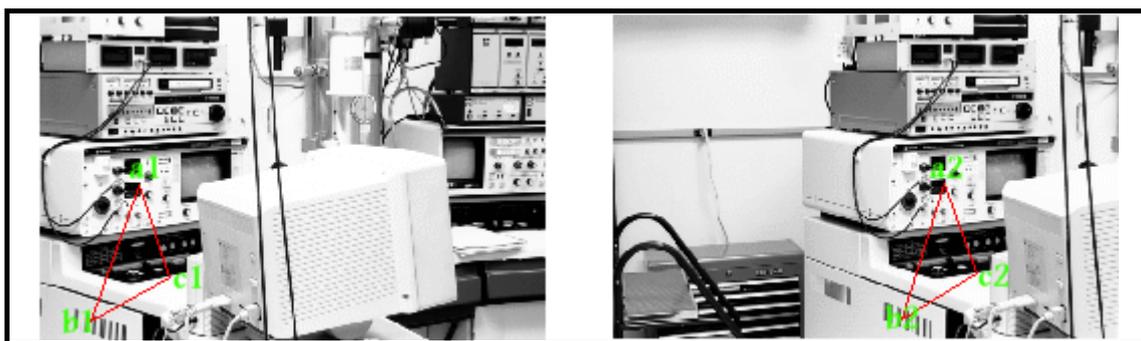


Figure 1.2: Exemple pour donner l'idée de notre travail

Les points a_1 , b_1 et c_1 de l'image A ont comme correspondance a_2 , b_2 et c_2 dans l'image B si et seulement si les deux triangles sont égaux et leurs angles ont le même signe par rapport à l'axe des x. Ainsi, les couples $\{a_1, a_2\}$, $\{b_1, b_2\}$ et $\{c_1, c_2\}$ sont des paires de mise en correspondance.

1.3 Plan du mémoire

Ce mémoire est divisé en sept chapitres, chacun abordant un aspect particulier de l'algorithme permettant la construction de la mosaïque:

- Le deuxième chapitre présente l'état de la recherche sur les méthodes actuelles dans le domaine de construction de la mosaïque en présentant certains d'entre eux. Quelques-unes de ces méthodes traitent d'une problématique semblable à la nôtre.
- Le troisième chapitre présente l'extraction des points d'intérêt. Nous parlons de classification des détecteurs des points d'intérêt et nous étudions le détecteur de Harris, utilisé dans notre méthode, en détail. Nous proposons une technique du « seuil adaptatif » pour améliorer la distribution des points extraits dans l'image.
- Le quatrième chapitre illustre la mise en correspondance. Au début nous parlons de la mise en correspondance par corrélation qui est basée sur une mesure de la similarité ou de la dissemblance. Ensuite, nous décrivons la mise en correspondance finale, extraire toutes les paires de points qui sont « vraies » en utilisant deux étapes :
 - choix des paires de référence,
 - comparaison par triangulation pour extraire les vraies paires de correspondance.
- Le cinquième chapitre vise à déterminer le modèle de caméra de chaque image en utilisant la méthode de calibration affine (recherche directe) et aussi la formation de l'image finale.
- Les résultats ainsi que la conclusion sont discutés en dernier lieu.

Chapitre 2:

2 Méthodes actuelles

Il existe plusieurs types de reconstruction panoramique à partir d'images. Les mosaïques les plus simples sont créées à partir d'un ensemble d'images dont les déplacements mutuels sont des translations « image-plan ». C'est approximativement le cas avec les images satellites.

D'autres mosaïques sont créées en tournant la caméra autour d'un centre optique à l'aide d'un dispositif spécial et créant une image panoramique qui représente la projection de la scène sur un cylindre ou une sphère.

2.1 Différentes méthodes de reconstruction d'une mosaïque

De nombreuses méthodes ont été proposées pour la construction d'une mosaïque.

La méthode de Zoghlami [Zog97] est donnée par l'algorithme suivant :

- Pour chaque image, extraire des points d'intérêt par filtrage. Ces points correspondent souvent à des coins.
- Chercher les caractéristiques géométriques des coins (positions, angles d'ouverture et de fermeture du coin).
- Sélectionner les coins paramétriques présentant le plus d'intérêt.
- Pour chaque paire de coins dans chacune des deux images :
 - a- Calculer l'homographie envoyant les paramètres géométriques des coins de l'image A sur ceux de l'image B et ceci en se servant des relations de dualité entre lignes et points du plan projectif [Zog97].
 - b- Donner une valeur décrivant la qualité de cette homographie.
- Prendre l'homographie qui a la meilleure valeur.
- Plutôt que d'attendre que l'étape précédente lui donne la meilleure homographie, il essaye de raffiner systématiquement les homographies qu'il rencontre si la valeur qu'elles obtiennent (au voisinage des points qui servent à leur calcul) est inférieure à une limite donnée.

Rouso[Rou98] a résolu le problème dans le cas où la trajectoire inclinée et le mouvement de la caméra serait dans la direction de l'axe optique (vers l'avant).

2.2 Reconstruction de mosaïque

Dans cette section, on décrit les méthodes existantes des étapes de l'algorithme de la construction de mosaïque. À la section 2.2.1, on parle des méthodes actuelles pour détecter des points d'intérêt (on décrit le choix de points d'intérêt dans la première partie et l'historique des méthodes connues dans la deuxième partie). Ensuite, on décrit les méthodes connues de la mise en correspondance dans la section 2.2.2 et enfin, on décrit les méthodes connues de la calibration des caméras dans la section 2.2.3.

2.2.1 Détection des points d'intérêt

Parmi les différents types de caractéristiques bas-niveau, nous avons choisi d'utiliser les points d'intérêt. Un point d'intérêt correspond à un changement bidimensionnel de l'intensité de l'image. Les coins, les jonctions en T et les endroits où la texture varie fortement en sont des exemples. Ce choix repose sur le fait que le signal contient plus d'information en ces points qu'aux points correspondant à des changements unidimensionnels du signal (lignes de contours).



Figure 2.1: Différents types de points d'intérêt

a- Coin, b- jonction en T et c- points de fortes variations de texture.

L'utilité des points d'intérêt a été constatée par Brady [Bra87] qui a remarqué qu'ils imposent plus de contraintes sur le processus visuel que les contours. Selon lui, ces points fournissent des endroits de calcul fiable. De même, Dreschler et Nagel [Dre82] ont constaté que le flux optique peut être calculé uniquement aux endroits correspondants à des points d'intérêt.

On peut également citer le travail de Zhang et al. [Zha95] qui ont montré que l'utilisation des points d'intérêt pour le calcul de la géométrie épipolaire donne de bons résultats. Dans leur travail, les points détectés sont appariés par corrélation. Leur calcul est effectué sur une fenêtre locale, au moins en ce qui concerne les méthodes basées sur le signal.

Les points d'intérêt ont également un caractère général. Leur extraction fonctionne aussi bien pour des objets simples que pour des objets complexes.

Historique des méthodes connues

Les détecteurs des points d'intérêt peuvent être classés en trois catégories. La première contient les méthodes basées sur les contours. C'est-à-dire, à partir de chaînes de contours, les endroits avec une courbure maximale ou un point d'inflexion sont recherchés. La deuxième méthode extrait le point d'intérêt directement à partir du signal de niveaux de gris, et la dernière méthode est l'approximation des points recherchés avec un modèle théorique.

Méthodes basées sur les contours

Le principe des méthodes basées sur les contours est de rechercher les points de courbure maximale le long des chaînes de contours ou d'effectuer une approximation polygonale en vue d'en déduire des points particuliers (intersection, inflexion,...). De telles méthodes existent depuis longtemps, mais nous ne détaillerons par la suite que celles qui sont les plus récentes.

Asada et Brady [Asa86] ont extrait des points d'intérêt pour des objets 2D à partir de courbes planaires. Ils ont constaté que les courbes planaires ont des caractéristiques significatives qui sont les changements de courbure. Ces changements sont classés en plusieurs catégories : coin, terminaison, etc. Pour pouvoir les détecter d'une manière robuste, l'algorithme est intégré dans un cadre multi-échelle. Une approche similaire a été proposée par Mokhtarian et Mackworth [Mok86]. Au lieu d'utiliser les changements de courbure d'une courbe plane, ils utilisent des points d'inflexion de celle-ci.

Medioni et Yasumoto [Med87] approximent les contours avec des B-splines. Les points d'intérêt sont des maxima de courbure calculés à partir des coefficients de ces B-splines.

Horand et al. [hor90] recherchent des groupements dans une image de contours pour établir une représentation intermédiaire. Cette représentation repose sur la structuration de segments extraits de l'image. L'intersection de ces segments donne les points d'intérêt.

Plusieurs détecteurs ont été suggérés pour résoudre ces problèmes (Harris [Har88], Susan [Smi97], CSS [Mok98],...).

Méthodes basées sur le signal :

Les méthodes basées sur le signal ne dépendent pas des contours ni d'un modèle théorique du signal. La mesure qui indique s'il y a un point d'intérêt à un endroit donné est calculée directement à partir du signal.

Beaudet [Bea78] a été le premier à proposer un détecteur de points d'intérêt. Cet opérateur utilise les dérivées secondes du signal pour calculer la mesure « DET » :

$$DET = I_{xx}I_{yy} - I_{xy}^2$$

où $I(x, y)$ représente l'intensité de l'image,

I_{xx} est la dérivée seconde en x,

I_{yy} est la dérivée seconde en y,

I_{xy} est la dérivée mixte.

Cette mesure est invariante en rotation et liée à la courbure gaussienne du signal. Les points où cette mesure est maximale sont les points d'intérêt. Pour obtenir les points d'intérêt, la valeur absolue de cette mesure doit être seuillée. Il faut noter que cet opérateur détecte les points d'intérêt près des coins mais pas sur les coins, pour autant que la notion de coin existe dans le signal.

Moravec [Mor79, Mor81] a proposé un détecteur basé sur la fonction d'auto-corrélation du signal. Cette fonction mesure les différences entre une fenêtre du signal et ses quatre fenêtres voisines. En effet, le voisinage n'est considéré que de manière discrète et dans les directions parallèles aux lignes et aux colonnes de l'image. Lorsque le minimum de ces quatre différences est supérieur à un seuil, ceci indique la présence d'un point d'intérêt.

Dreschler et Nagel [Dre82] ont trouvé un défaut à l'approche de Beaudet[Bea78] car la courbure gaussienne peut devenir grande sur des contours marqués, c'est-à-dire sur des contours dont les deux niveaux de gris qui les définissent sont très différents. Ceci est dû au fait que la courbure gaussienne est le produit de deux courbures principales d'une surface, et sur un contour marqué, une des deux courbures devient très importante.

En utilisant un modèle théorique d'un coin, ils ont constaté qu'autour d'un coin la courbure gaussienne change de signe et qu'elle possède un maximum positif et un minimum négatif. Ils proposent donc de localiser un point d'intérêt sur la ligne joignant ce minimum et ce maximum, notamment à l'endroit où la pente du signal est maximale. À cet endroit la courbure s'annule et change de signe.

Harris et Stephens [Har88] ont amélioré l'approche de Moravec en calculant une matrice liée à la fonction d'auto-corrélation qui prend en compte les valeurs des dérivées premières du signal sur une fenêtre. Ceci est une amélioration par rapport à Moravec, car la discrétisation choisie n'est plus nécessaire. Ils obtiennent donc la matrice suivante :

$$e^{-\frac{x^2 + y^2}{2\sigma^2}} \otimes \begin{bmatrix} (I_x)^2 & I_x I_y \\ I_x I_y & (I_y)^2 \end{bmatrix}$$

où \otimes est l'opérateur de convolution et σ^2 est la variance.

Les valeurs propres de cette matrice sont les courbures principales de la fonction d'auto-corrélation. Si ces deux courbures sont grandes, ceci indique la présence d'un point d'intérêt. L'utilisation des courbures est plus précise que l'utilisation de la valeur minimale comme l'avait proposée Moravec. Toutefois, pour éviter d'extraire les valeurs propres, Harris et Stephens utilisent une mesure reposant sur le déterminant et la trace de la matrice. Cette mesure est supérieure à zéro dans le cas d'un coin.

Forstner [For87, For94] propose une approche basée sur la statistique locale d'une image. Ceci lui permet d'estimer les paramètres de son algorithme de manière automatique. La première étape de son algorithme est d'estimer la variance du bruit. Il utilise ensuite cette estimation pour restaurer le signal. Puis, les pixels sont classés dans les catégories : régions, contours et points d'intérêt. Pour ce faire, il utilise la fonction d'auto-corrélation de la même manière que Harris et Stephens. Enfin, il classe les points d'intérêt en deux catégories : soit une jonction, soit un point isolé.

Heitger et al. et Rosenthaler et al. [Hei92, Ros92] ont proposé une autre approche inspirée des mécanismes neuro-biologiques. Leur approche consiste à convoluer l'image avec des filtres directionnels pairs et impairs. Ces filtres sont des fonctions sinusoïdales sur une enveloppe gaussienne de moyenne nulle. Ils ressemblent à des filtres de Gabor. Les résultats des filtres

pairs et impairs d'une même direction permettent de calculer l'énergie locale de l'image dans cette direction. Cette énergie correspond aux caractéristiques 1D de l'image. Pour obtenir la caractéristique 2D de l'image, leur approche consiste à calculer pour chaque direction les dérivées premières et secondes de cette énergie. La mesure ainsi obtenue permet de détecter les caractéristiques 2D mais, aussi les fausses réponses sur les caractéristiques 1D. Une méthode reposant sur la nature systématique des erreurs permet ensuite d'éliminer les réponses sur les caractéristiques 1D. Enfin, les réponses restantes sont seuillées.

Méthodes basées sur un modèle théorique du signal

En ce qui concerne les méthodes basées sur un modèle théorique du signal, le but est d'obtenir une précision « sous-pixel » en approximant le signal par un modèle théorique. De telles méthodes ne sont utilisables que pour des types bien précis de points d'intérêt comme des coins par exemple. Une telle approche est inutilisable dans un contexte général de détection des points d'intérêt. Par contre les cercles, les lignes, etc. peuvent également être modélisés par une telle approche.

Roher [Roh92] modélise les jonctions de plusieurs lignes. Pour ce faire, il convolue un modèle binaire de jonction avec une gaussienne afin de modéliser le flou. Dans le cas d'un coin, les paramètres du modèle sont l'angle définissant l'orientation de l'axe de symétrie, l'angle définissant l'ouverture du coin, les niveaux de gris, la position du point et le flou. Ces paramètres sont ajustés pour que le signal théorique soit le plus proche possible du signal observé. Cette recherche repose sur une minimisation au sens des moindres carrés. Les coins obtenus par cette méthode sont très précis. Toutefois, la qualité de l'approximation repose sur une bonne estimation initiale de la position. Roher utilise les segments extraits pour déterminer les paramètres initiaux du modèle : le type du coin, les angles initiaux ainsi que la position du coin.

Deriche et Blaszkia [Der93a] ont proposé une amélioration de la méthode de Roher au niveau du temps de calcul en remplaçant la fonction gaussienne de lissage par une fonction exponentielle. Ils proposent en outre, une solution pour l'initialisation des paramètres. Une fois que la position du coin a été donnée de façon interactive par l'utilisateur, l'ouverture du coin est estimée de façon automatique à partir du gradient sur les bords de la fenêtre. Le point donné par l'utilisateur est ensuite amélioré avec une méthode de descente de la variance des

niveaux de gris. Deriche et Blaszkia ont montré la bonne précision en position des résultats en présence de bruit synthétique. Cette étude a aussi permis de voir que pour assurer la convergence de la minimisation, le support utilisé doit être assez grand.

Brand et Moher [Bra94] proposent une implémentation différente. Le modèle théorique est ajusté au signal en utilisant une transformation affine. L'importance de leurs travaux repose sur l'évaluation de la qualité de la localisation d'une telle approche sur des données réelles. Pour cette évaluation, ils mettent en place plusieurs tests : un test d'alignement, une reconstruction 3D et un calcul de la géométrie épipolaire. Ces tests permettent d'obtenir une précision de 0.1 pixel.

Deriche et Giraudon [Der93b] ont présenté une méthode différente pour améliorer la précision de la détection tout en utilisant un modèle théorique pour un coin. À partir de ce modèle, ils analysent d'une manière théorique le comportement de différents détecteurs. Ils constatent que la réponse de nombreux détecteurs varie suivant l'échelle. Toutefois, il existe une relation entre la position de la véritable caractéristique à détecter et les réponses dans l'espace d'échelle. Pour le détecteur de Beaudet [Bea78], les réponses se situent, pour un point donné, sur la bissectrice de l'angle définissant ce coin. Ceci, et le fait que le Laplacien du signal s'annule à l'endroit précis du coin indépendamment de l'échelle considérée, leur permet de proposer la méthode suivante : d'abord, un détecteur de Beaudet [Bea78] est appliqué à deux échelles différentes. Ensuite, les points correspondant à un même coin sont recherchés dans les deux images. La droite reliant ces deux points définit la bissectrice de ce coin. La position exacte du coin est alors déterminée sur cette droite à l'endroit le plus proche du point où le Laplacien s'annule.

2.2.2 Mise en correspondance

Pour deux images A et B d'une scène, la mise en correspondance consiste à déterminer, pour un élément de l'image A, l'élément qui lui correspond dans l'image B.

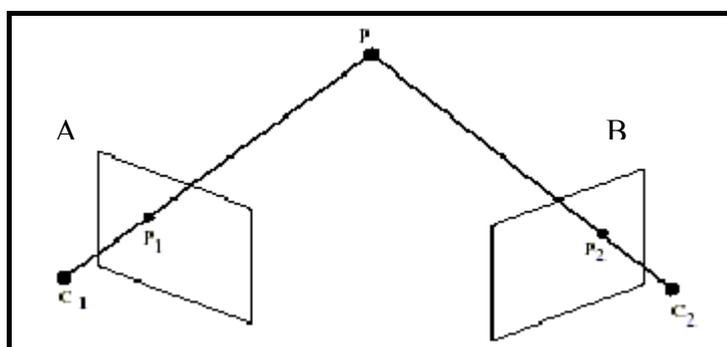


Figure 2.2: Deux points de mise en correspondance.

La figure 2.2 illustre la mise en correspondance (P désigne le point de la scène, p_1 et p_2 désignent les points dans les images A et B, et c_1 et c_2 désignent les caméras).

Méthodes existantes ou « évidentes »

De nombreuses méthodes ont été proposées dans le passé pour résoudre le problème de l'appariement. Dans ce qui suit, on parlera de quelques méthodes existantes. Au début, on décrit la méthode de relaxation, ensuite on parlera de la méthode de la recherche arborescente, et après on décrira la méthode de détection de la clique maximale. Pour finaliser, on parlera de la méthode de la programmation dynamique.

Relaxation

Une méthode populaire est la technique de relaxation [Hum83]. C'est un algorithme itératif et localement parallèle. D'abord, un ensemble initial des appariements possibles est construit en appariant chaque primitive de la première image avec chaque primitive de la deuxième image. Cet ensemble est organisé comme une collection de nœuds $\{a_i\}$, un nœud pour chaque primitive de la première image. On associe à chaque nœud a_i un vecteur x qui représente l'information géométrique de la primitive dans la première image, et l'ensemble d'étiquettes L_i qui représente les appariements possibles des primitives. Une étiquette spéciale, dite le caractère nul $*$, est initialement incluse dans L_i , qui signifie que la primitive en considération

n'a pas de primitive correspondante dans la deuxième image. À chaque étiquette e dans L_i , nous associons un nombre $p_i(e)$ qui est interprété comme une estimation de probabilité que e est effectivement la correspondance de la primitive en considération dans la première image. Donc, nous avons $p_i(e) \in [0,1]$ et $\sum_{e \in L_i} p_i(e) = 1$. Ces estimations de probabilité seront, au fur et à mesure, mises à jour en tenant compte de la cohérence dans le voisinage. Si beaucoup de primitives dans le voisinage sont relativement compatibles avec l'appariement e , alors $p_i(e)$ augmentera; sinon, $p_i(e)$ diminuera. Après un nombre suffisant d'itérations, un nœud est considéré comme appariable s'il a une étiquette avec une probabilité suffisamment élevée, sinon, il n'est pas considéré comme appariable. Ce qui est essentiel dans cette technique est la définition de la règle pour la mise à jour des probabilités.

Pour réduire les ambiguïtés des appariements initiaux, des contraintes unitaires peuvent être imposées [Bar80]: l'angle du coin, la grandeur du gradient, la longueur du segment de droite, le maximum de la disparité, etc . Les contraintes binaires utilisées déterminent la manière de l'évolution des probabilités. Barnard et Thompson [Bar80] utilisent des points d'intérêt où la variance est forte dans toutes les directions et montrent l'efficacité de la propriété de la continuité locale de la disparité. Cette propriété est une conséquence directe de la continuité des surfaces réelles dans l'espace. Leur méthode a donc des difficultés près de la discontinuité. Ogawa [Oga84] utilise aussi une technique de relaxation pour appairer deux ensembles de points. Les points représentent des étoiles, chacune ayant une magnitude qui peut être utilisée comme une contrainte unitaire. Au lieu d'utiliser des points individuellement, il considère des paires de points comme primitives. La mesure de compatibilité est calculée en tenant compte de la transformation géométrique à partir des deux paires en considération. Cheng et Huang [Che84] vont encore plus loin. Les segments de droite 2D sont utilisés. Au lieu d'utiliser des segments individuels, ils utilisent une représentation qu'ils appellent « une structure étoile ». Une étoile est un groupe de segments qui sont parallèles, co-linéaires ou adjacents. La technique de relaxation est utilisée, mais comme les primitives maintenant sont des étoiles qui contiennent beaucoup plus d'informations, l'espace de recherche est considérablement réduit.

Recherche arborescente

L'approche traditionnelle pour effectuer la mise en correspondance est la recherche arborescente en profondeur. Pour expliquer une telle procédure, nous appelons les primitives dans la première image des unités, et celles dans la deuxième image des étiquettes. La procédure assigne des étiquettes aux unités tant qu'elle peut trouver une étiquette pour chaque unité qui est compatible avec les étiquettes déjà fixées à des unités précédentes, compatible au sens des contraintes binaires définies. Aussitôt que la procédure n'arrive pas à trouver une étiquette pour une nouvelle unité, elle rebrousse chemin. Il peut donc y avoir beaucoup de rebroussements avec l'algorithme standard. Haralick et Elliott [Har80] montrent analytiquement et expérimentalement que des heuristiques comme suit donnent des performances supérieures à l'algorithme standard :

- Anticiper le futur et essayer d'abord les endroits où on sera très probablement amené à échouer.
- Mémoriser ce qu'on a déjà fait pour éviter de commettre la même erreur.

Détection de la clique maximale

Une technique bien connue pour l'isomorphisme du sous-graphe est la détection de la clique maximale, qui a été appliquée dans l'analyse de séquence d'images. D'abord, un graphe d'association est construit de la manière suivante. Pour chaque primitive p_1 de la première image et chaque primitive p_2 de la deuxième image, elles forment un nœud a_i si elles satisfont les relations unitaires. Deux nœuds a_i et a_j se connectent s'ils sont compatibles, c'est-à-dire s'ils satisfont les contraintes binaires. Le meilleur « matching » entre deux structures relationnelles est le plus grand ensemble de nœuds qui sont mutuellement connectés dans le graphe d'association c'est-à-dire la clique maximale. Des algorithmes pour la détection de la clique maximale existent dans la littérature [Rad84]. Ils sont en général coûteux en temps de calcul. Notons aussi que cette technique a été appliquée aussi pour la reconnaissance d'objets où la taille du graphe est bien limitée.

Programmation dynamique

La programmation dynamique est une méthode pour résoudre des problèmes d'optimisation non-linéaire quand les variables ne sont pas toutes inter-dépendantes. Un tel problème peut être décomposé en une séquence de sous-problèmes d'optimisation avec seulement une

variable (plus généralement, avec un sous-ensemble des variables). Considérons le problème suivant :

$$\text{Min}_{x_1, x_2, x_3} h(x_1, x_2, x_3)$$

où les x_i prennent des valeurs discrètes. Si on l'on sait que x_2 dépend seulement de x_1 , et que x_3 dépend seulement de x_2 , c'est-à-dire:

$$h(x_1, x_2, x_3) = h_1(x_1, x_2) + h_2(x_2, x_3)$$

alors, on peut minimiser h_1 sur x_1 et mettre dans un tableau la valeur minimale de $h_1(x_1, x_2)$ pour chaque x_2 , c'est-à-dire:

$$f_1(x_2) = \min_{x_1} h_1(x_1, x_2)$$

Continuons de la même manière et éliminons x_2 en calculant $f_2(x_3)$ comme suit:

$$f_2(x_3) = \min_{x_2} [f_1(x_2) + h_2(x_2, x_3)]$$

Le problème initial devient donc

$$\text{Min}_{x_1, x_2, x_3} h(x_1, x_2, x_3) = \text{Min}_{x_3} f_2(x_3)$$

La programmation dynamique est analogue à la recherche d'un chemin optimal dans un graphe. Les primitives de la première image sont ordonnées, d'une manière arbitraire ou non, et sont indiquées sur l'axe x . Les primitives de la deuxième image sont indiquées sur l'axe y . Supposons qu'on a N et M primitives dans les deux images respectivement. Un chemin de $i = 1$ à $i = N$ est une association des primitives de la deuxième image à celles de la première image. Le problème du « matching » est donc de trouver le chemin optimal. Le principe de la séparabilité dans la programmation dynamique est appliqué de la manière suivante : à chaque étape i , un nœud (j, k) est associé au chemin qui garantit que le coût cumulé est minimal jusqu'à l'étape i . Le coût cumulé du chemin terminé au nœud (j, k) est défini comme :

$$C(j, k) = \min_{1 \leq h \leq M} f[c(j-1, h), c(j-1, h; j, k)]$$

où c est le coût local pour lier le nœud $(j-1, h)$ au nœud (j, k) , et $f(C, c)$ est une combinaison appropriée quelconque de C et c . La complexité de cet algorithme est $O(NM^2)$. Si les primitives dans les deux images sont ordonnées de la même manière, ou si on a d'autres informations, l'espace de recherche peut être considérablement réduit.

2.2.3 Calibration des caméras

Calibrer une caméra revient à calculer ses paramètres (internes et externes) en utilisant uniquement une ou plusieurs images d'un objet de calibration. Ces paramètres sont donnés par une matrice M de 4×4 que l'on nomme matrice de transformation. Une telle matrice permet de représenter plusieurs modèles de caméra (perspective, perspective faible, orthographe, affine)

$$\begin{pmatrix} xh \\ yh \\ h \end{pmatrix} = M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

où (X, Y, Z) est un point 3D et (x, y) sa projection dans l'image, tous deux exprimés dans l'espace homogène.

Méthodes connues pour la calibration de caméras

Dans cette section, nous allons présenter quelques méthodes connues de calibration de caméras.

Modèle perspective

À l'aide d'une liste de points correspondants entre l'image $p_i = (x_i, y_i)$ et la scène $P_i = (X_i, Y_i, Z_i)$, on estime les paramètres internes et externes directement à partir des équations de projection de P vers p . Pour une caméra perspective, on a :

$$\begin{pmatrix} h_i x_i \\ h_i y_i \\ h_i \end{pmatrix} = JART \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$A = \begin{pmatrix} -f_x & 0 & o_x & 0 \\ 0 & -f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

$$T = \begin{pmatrix} 1 & 0 & 0 & T_x \\ 0 & 1 & 0 & T_y \\ 0 & 0 & 1 & T_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Tout revient à résoudre les équations :

$$x_i - o_x = -f_x \frac{r_{11}X_i + r_{12}Y_i + r_{13}Z_i + T_x}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z}$$

$$y_i - o_y = -f_y \frac{r_{21}X_i + r_{22}Y_i + r_{23}Z_i + T_y}{r_{31}X_i + r_{32}Y_i + r_{33}Z_i + T_z}$$

où les paramètres connus sont :

(x_i, y_i) : les points de l'image;

(X_i, Y_i, Z_i) : les points 3D de la scène;

(o_x, o_y) : le centre de l'image.

Les paramètres à estimer sont :

(f_x, f_y) : la distance focale;

$r_{11}, r_{12}, \dots, r_{33}$: les paramètres de la matrice de rotation de la caméra;

T_x, T_y, T_z : les paramètres de translation de la caméra.

C'est un système non-linéaire qu'on peut résoudre, cependant il est instable quand il y a du bruit dans l'image.

Modèle affine

À l'aide d'une liste de points correspondants entre l'image $p_i = (x_i, y_i)$ et la scène $P_i = (X_i, Y_i, Z_i)$, on peut estimer les paramètres d'un modèle plus général (modèle affine), puis on convertit ces paramètres vers ceux du modèle perspective.

Un modèle affine est donné par :

$$\begin{pmatrix} xh \\ yh \\ h \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}.$$

Alors tout revient à résoudre le système suivant :

$$x_i = \frac{m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

$$y_i = \frac{m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

On obtient les paramètres de la matrice du modèle affine. C'est un système linéaire facile à résoudre et qui est plus stable, mais il est plus général.

Géométrie épipolaire

Cette méthode consiste à calculer les paramètres d'une caméra en utilisant uniquement des points mis en correspondance entre deux images d'une même scène, prises sous des points de vue différents.

Soit $p_l = (x_l, y_l)$ et $p_r = (x_r, y_r)$ les points images correspondants. On a deux cas :

- Paramètres interne connus : alors il faut calculer les paramètres de la matrice essentielle E donnée par l'équation :

$$p_r^T E p_l = 0$$

- Paramètres internes inconnus : alors il faut calculer les paramètres de la matrice fondamentale F donnée par l'équation :

$$p_r'^T F p_l' = 0$$

Avec huit points le système d'équation linéaire est facile à résoudre.

Chapitre 3

3 Points d'intérêt

Dans ce chapitre, nous présentons la première étape de notre algorithme : l'extraction des points d'intérêt. Le choix des points d'intérêt est d'abord expliqué à la section 3.1, ensuite dans la section 3.2, nous décrivons l'algorithme de détection des points d'intérêt, en outre nous décrivons les différentes étapes de l'algorithme (choix du seuil, le filtrage de gradient par convolution). Enfin, la section 3.3 présente quelques résultats obtenus.

3.1 Notre choix de méthode (algorithme de Harris)

Parmi les méthodes étudiées, on a utilisé le détecteur de Harris [Har88]. La méthode de Harris est basée sur le calcul d'une matrice liée à la fonction d'auto-corrélation qui tient compte des valeurs des dérivées premières du signal sur une fenêtre. Cette matrice M est donnée par :

$$M = \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix}$$

Les valeurs propres de cette matrice sont les courbures principales de la fonction d'auto-corrélation. Si ces deux courbures sont grandes, ceci indique la présence d'un point d'intérêt. Toutefois, pour ne pas extraire les valeurs propres, Harris utilise une mesure qui repose sur le déterminant et sur la trace de la matrice. Cette mesure est supérieure à zéro dans le cas d'un coin.

Alors on cherche en chaque points de l'image les valeurs maximales de $R(x, y)$ donnée par :

$$R(x, y) = \det[M] - K(tr[M])^2$$

où k est une constante qui a comme valeur fixe 0.04 [Xav00] et $tr[M]$ est la trace de la matrice M qui est la somme des éléments de la diagonale de la matrice M .

I_x et I_y sont les dérivées des niveaux de gris de l'image en (x, y) , selon les axes x et y respectivement. L'image résultante est seuillée en fonction du nombre des points d'intérêt désirés. Voir la section 3.2.1 pour plus de détails.

3.2 L'algorithme de détection des points d'intérêt

L'algorithme de détection des points d'intérêt est constitué des étapes suivantes :

- Filtrage : la dérivée est très sensible au bruit. Une image doit donc être filtrée afin de ne pas générer trop de faux points d'intérêts. Cependant, le filtrage diminue la réponse de l'algorithme pour certains points d'intérêt réels. Il y a donc un compromis entre la diminution du bruit et la conservation des points d'intérêt.
- Calcul du gradient : nous avons utilisé l'opérateur de Sobel pour calculer le gradient par rapport à x et à y . Les composantes du gradient sont obtenues en convoluant l'image avec les masques :

$$\frac{1}{4} \begin{pmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{pmatrix} \text{ et } \frac{1}{4} \begin{pmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{pmatrix}$$

- Calculer $R(x, y) = \det[M] - K \operatorname{tr}^2[M]$
où K est une constante qui a comme valeur fixe 0.04 [Xav00].
- Seuil : pour détecter un point d'intérêt on fixe un seuil unique. Si $R(x, y)$, en un point, est supérieur à ce seuil, on le considère comme étant un point d'intérêt.

3.2.1 Le seuil

Les points d'intérêt sont les points de $R(x, y)$ qui ont les plus grandes valeurs positives. Après avoir appliqué le détecteur de Harris sur l'image, nous obtenons beaucoup de points d'intérêt. En général, nous avons besoin d'une centaine de points pour faire la mise en correspondance. Appliquer un seuil nous donne le nombre de points que nous voulons en conservant les points d'intérêt qui ont les plus grandes valeurs. Mais comment choisir le seuil?

Pour répondre à cette question, nous proposons trois seuils différents et nous choisissons un parmi eux qui va améliorer la distribution des points d'intérêt désirés obtenus dans l'image. Pour les deux premiers seuils, nous utilisons une valeur fixe pour l'appliquer à l'image. Lors du troisième seuil, la valeur varie dépendamment de chaque fenêtre $n \times n$ de l'image.

Seuil « Maximum »

Ce seuil, appelé S_{\max} , dépend de la valeur maximale de $R(x, y)$ telle que :

$$S_{\max} = \frac{\max[R(x, y)]}{a}$$

où a est une constante qui change selon le nombre des points désirés.

On applique ce seuil sur tous les éléments de la matrice R . L'élément qui a une valeur plus grande que ce seuil est considéré comme point d'intérêt, et celui qui est le plus petit est rejeté.

Autrement dit :

$$\text{Points d'intérêt} = \{(i, j) / R(i, j) \geq S_{\max}\}.$$

Le nombre de points d'intérêt dépend de la constante a . Si le nombre de points est petit, on réduit a pour augmenter le nombre de points d'intérêts. Si le nombre de points d'intérêt est très grand, on augmentera a pour réduire ce nombre, parce que si celui-ci est très grand :

- Le temps d'exécution du programme augmente.
- La difficulté de mettre les points en correspondance augmente.

Seuil « Moyen »

Les points d'intérêt obtenus en utilisant le S_{\max} dépendent de la présence de la texture dans l'image, comme on le verra dans les exemples. Pour cela, nous utilisons un autre seuil global, appelé S_{moy} , appliqué aux valeurs maxima obtenues de toutes les fenêtres.

Nous prenons pour chaque pixel (i, j) de l'image une fenêtre F de $n \times n$ pixels, et nous prenons le maximum $M(i, j)$ de ces valeurs comme point associé pour cette fenêtre

$$M(i, j) = \max_{\substack{i \leq l \leq n \\ j \leq k \leq n}} F(l, k)$$

Alors, pour chaque fenêtre, nous avons un point associé qui est le maximum $M(i, j)$. Nous calculons la moyenne des maxima obtenus et nous l'utilisons comme seuil.

$$S_{moy} = \frac{1}{N} \frac{\sum_{i,j} M(i, j)}{a}$$

où a est une constante qui change selon le nombre des points désirés obtenus, et N est le nombre des points maximum obtenus.

$$\text{Points d'intérêt} = \{(i, j) / R(i, j) \geq S_{moy}\}.$$

Seuil « Adaptatif »

Les deux premiers seuils sont globaux. Dans ce paragraphe, nous proposons un seuil local, appelé $S_{adaptatif}$, qui permet d'obtenir des points d'intérêt bien distribués dans l'image.

Nous prenons pour chaque pixel (i, j) de l'image une fenêtre F de $n \times n$ pixels et nous calculons le maximum $M(i, j)$ de cette fenêtre

$$M(i, j) = \max_{\substack{i-n \leq l \leq i+n \\ j-n \leq k \leq j+n}} F(l, k)$$

Nous disons que le pixel (i, j) est un point d'intérêt si et seulement si :

$$R(i, j) = M(i, j)$$

Autrement dit le pixel (i, j) est un point d'intérêt si sa valeur de Harris est la plus grande valeur obtenue dans la fenêtre.

$$\text{Points d'intérêt} = \{(i, j) / R(i, j) = M_{(i,j)}\}$$

En utilisant ce seuil, nous pouvons garantir que chaque fenêtre F de l'image contient au moins un point d'intérêt, et ce peu importe la texture présente dans la fenêtre.

3.2.2 Filtrage et gradient par convolution

Les parties 1 et 2 de l'algorithme des points d'intérêt sont calculées par la convolution de l'image avec un opérateur approprié. Cette section décrit comment nous réalisons ces opérations. Soit $f(i, j)$ l'image originale, $g(i, j)$ l'opérateur de convolution de dimension $(n \times m)$ et dont l'origine est (o_x, o_y) . Le résultat de la convolution est:

$$h(i, j) = f(i, j) \times g(i, j) = \sum_{k=1}^n \sum_{l=1}^m f(i+k-o_x, j+l-o_y) g(k, l)$$

3.3 Résultats

Dans cette section, nous présentons des résultats obtenus pour des points associés par le détecteur de Harris en utilisant les trois seuils proposés dans la section 3.2.1, soient les seuils « maximum », « moyen » et « adaptatif ».

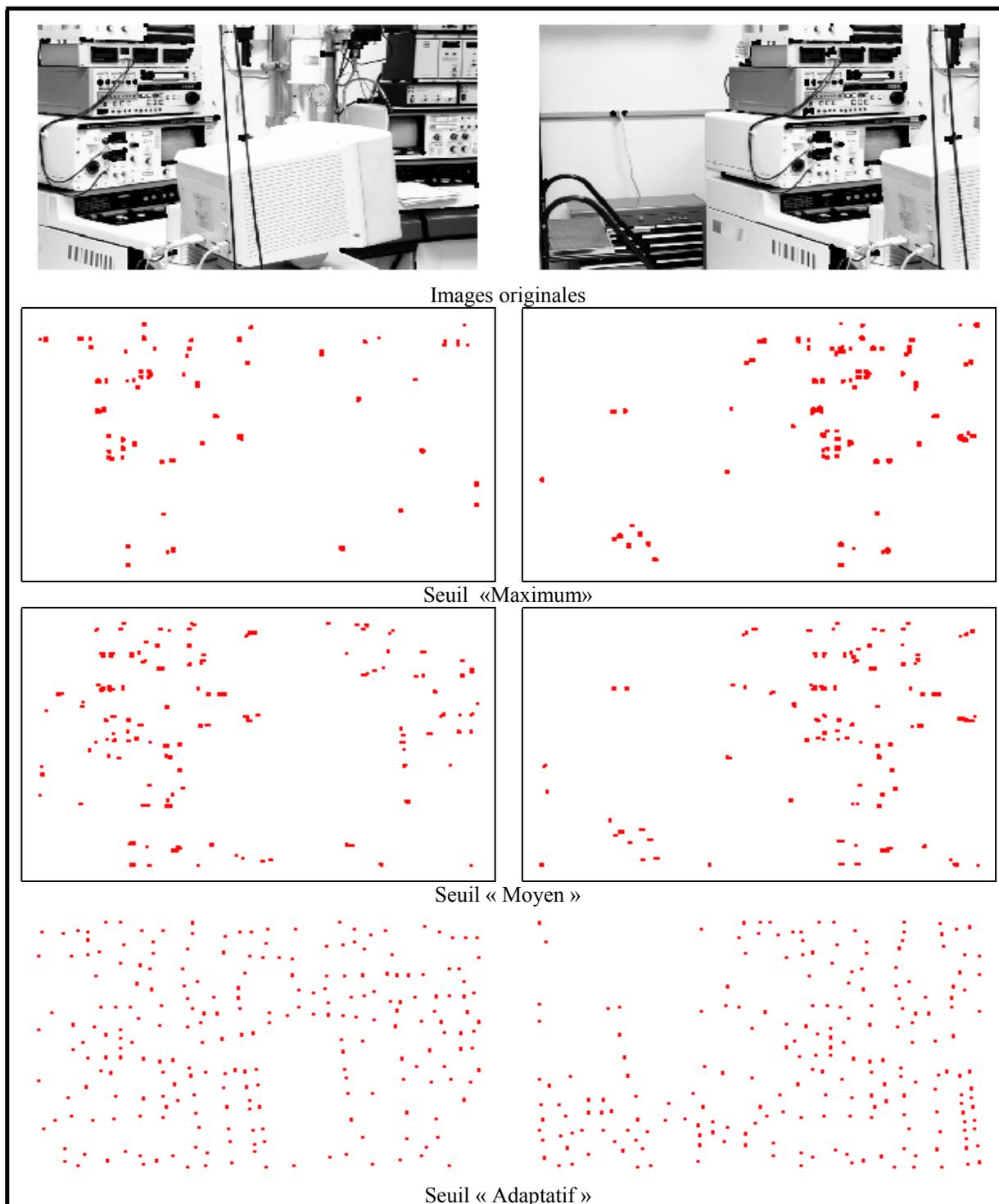


Figure 3.1: Exemple du détecteur de points d'intérêt en utilisant les trois seuils

Dans la figure 3.1 nous prenons les images originales bien texturées, surtout dans la zone de correspondance. Les points d'intérêt obtenus en utilisant le seuil « Maximum » ne couvrent pas bien l'image. La plupart de ces points appartiennent à une partie de la zone de correspondance. Les points d'intérêt obtenus en utilisant le seuil « Moyen » couvrent bien les endroits où il y a des textures dans l'image. Nous remarquons que la plupart de points détectés appartiennent à la zone de correspondance et ils la couvrent uniformément. Les points d'intérêt obtenus en utilisant le seuil « Adaptatif » couvrent très bien l'image. On voit qu'avec les trois seuils il y a des zones n'ayant pas des points d'intérêt (zones vides) qui diminuent en passant de « seuil maximum » au « seuil moyen » au « seuil adaptatif ». Ces zones sont causées par l'absence de textures dans les images originales. Nous remarquons qu'il y a beaucoup de points détectés à l'extérieur de la zone d'intersection de deux images, et qui ne seront pas nécessaires à la mise en correspondance.

seuil	Valeur de a	Taille de la fenêtre	Nb des points détectés	
			Image A	Image B
maximum	2.5		300	492
moyen		18 x 18	362	301
adaptatif		18 x 18	228	208

Tableau 3-1 : statistique sur le nombre des points d'intérêt du figure 3.1.

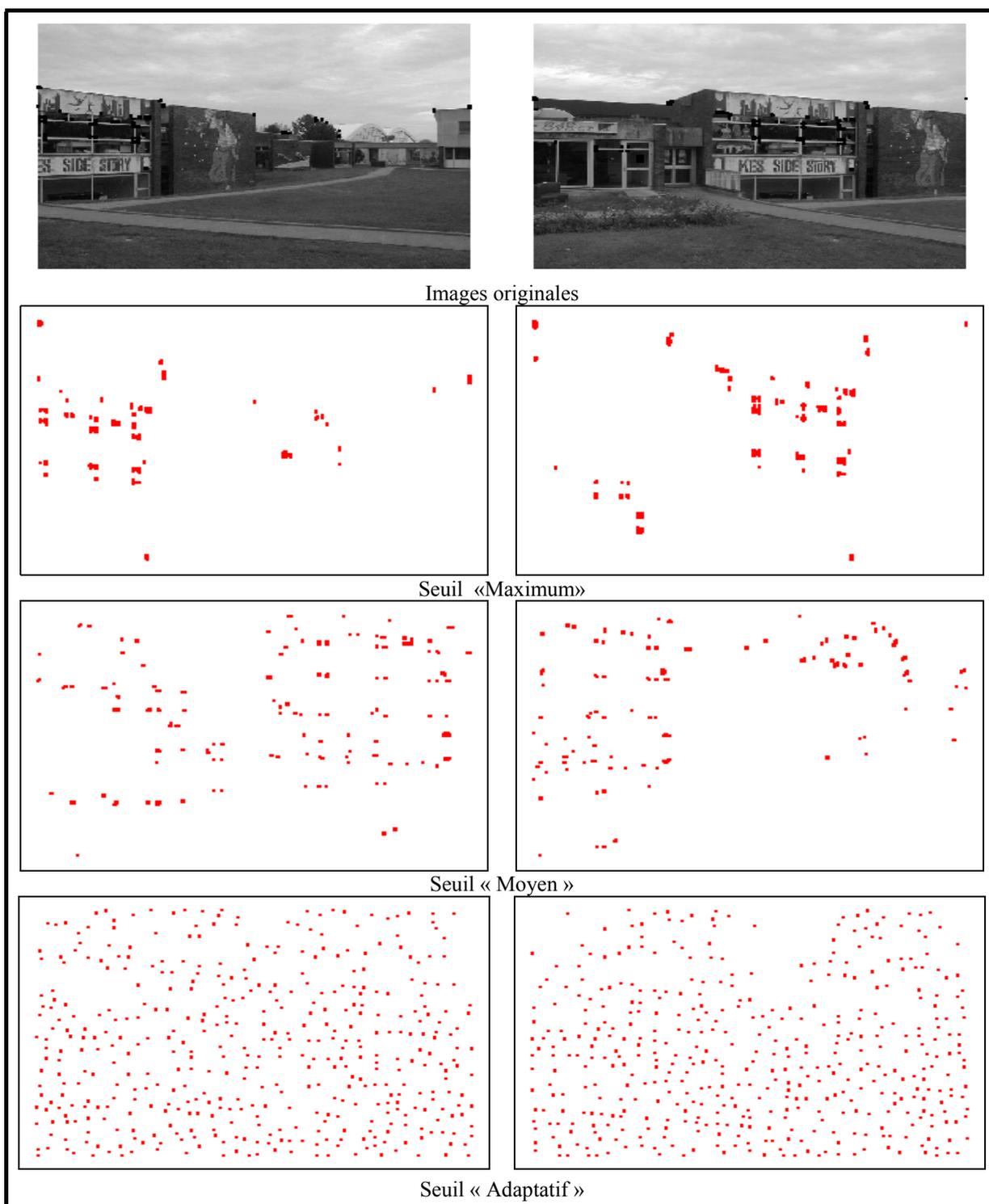


Figure 3.2: Exemple du détecteur de points d'intérêt en utilisant les trois seuils

Dans la figure 3.2 nous prenons les images originales moins texturées. Les points d'intérêt obtenus en utilisant le seuil « Maximum » ne couvrent pas bien l'image. La plupart de ces points appartiennent à une partie de l'image où il y a vraiment de la texture. Les points d'intérêt obtenus en utilisant le seuil « Moyen » couvrent bien les endroits où il y a des textures dans l'image. De même, il y a des points d'intérêt détectés dans le ciel nuageux ceci s'explique par l'effet que les maxima obtenus à cette zone sont plus grands que le seuil moyen. Nous remarquons que la plupart de points détectés appartiennent à la zone de correspondance. Les points d'intérêt obtenus en utilisant le seuil « Adaptatif » couvrent bien l'image. Nous remarquons qu'il y a beaucoup de points détectés à l'extérieur de la zone d'intersection de deux images et qui ne seront pas nécessaires à la mise en correspondance.

seuil	Valeur de a	Taille de la fenêtre	Nb des points détectés	
			Image A	Image B
maximum	2.5		179	235
moyen		18 x 18	298	241
adaptatif		18 x 18	434	438

Tableau 3-2 : statistique sur le nombre des points d'intérêt du figure 3.2.

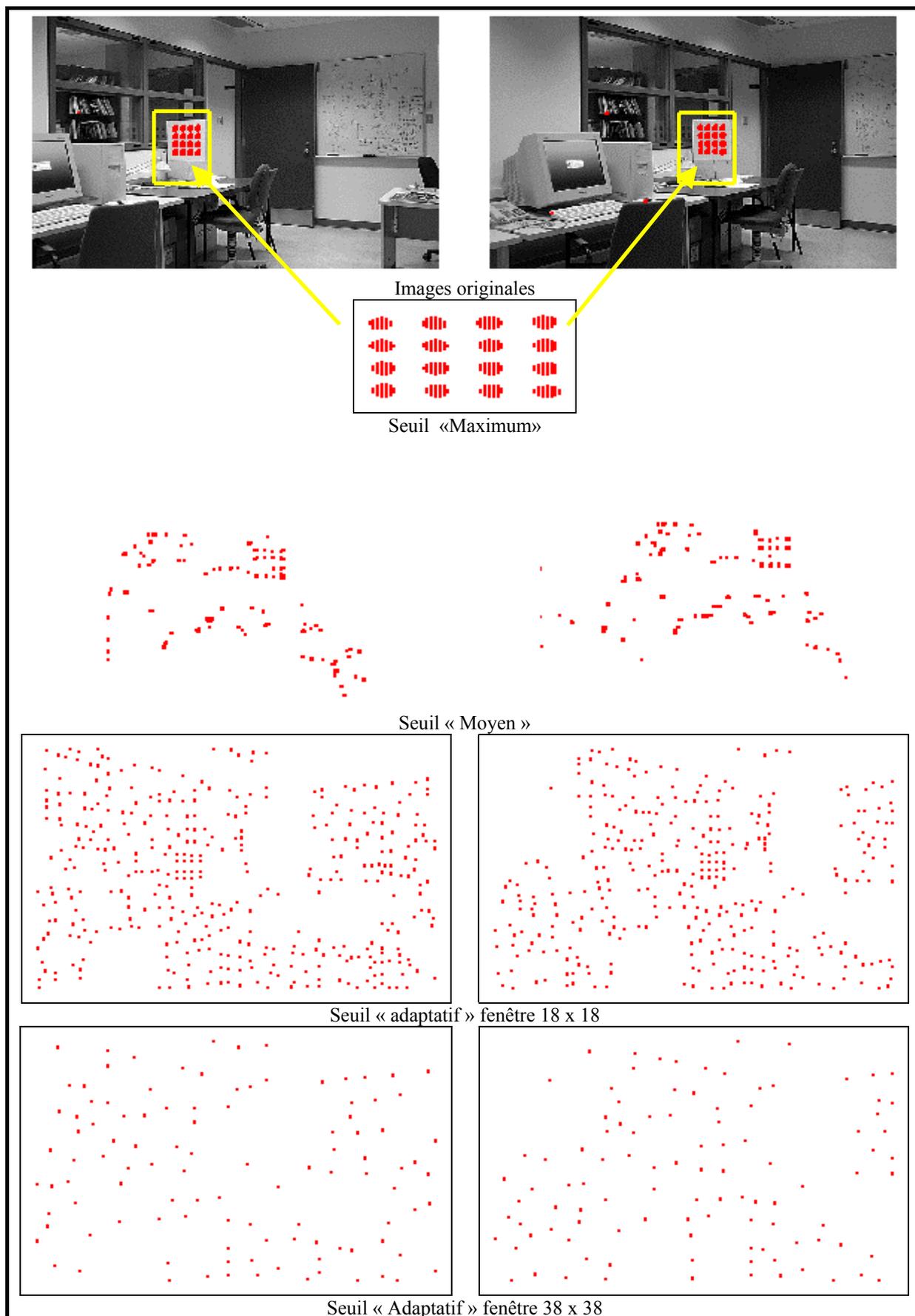


Figure 3.3: Exemple du détecteur de points d'intérêt en utilisant les trois seuils

Dans la figure 3.3 nous prenons les images originales bien texturées avec une partie très fortement texturée (illustrées par des rectangles dans les images originales) . Les points d'intérêt obtenus en utilisant le seuil « Maximum » ne couvrent pas toute l'image. La plupart de ces points appartiennent à la partie où la texture est très forte.

Les points d'intérêt obtenus en utilisant le seuil « Moyen » couvrent bien les endroits où il y a des textures dans l'image. Il réduit le nombre des points d'intérêt dans la partie très fortement texturée. Les points d'intérêt obtenus en utilisant le seuil « Adaptatif » couvrent très bien l'image. Nous remarquons qu'il y a beaucoup de points détectés à l'extérieur de la zone d'intersection de deux images et qui ne seront pas nécessaires à la mise en correspondance. Nous remarquons que le nombre des points est très petit dans la zone texturée, ce qui montre l'avantage du seuil « Adaptatif » sur les autres seuils.

seuil	Valeur de a	Taille de la fenêtre	Nb des points détectés	
			Image A	Image B
maximum	2.5		532	532
moyen		18 x 18	179	177
adaptatif		18 x 18	377	328
Adaptatif		38 x 38	107	103

Tableau 3-3 : statistique sur le nombre des points d'intérêt du figure 3.3.

3.4 Discussion de résultats obtenus

Nous remarquons que le seuil « Maximum » fonctionne bien s'il y a des textures partout dans l'image, ce qui est le cas du premier exemple à la figure 3.1. Il échoue là où il y a des textures sur une petite zone de l'image, ce qui est le cas de la figure 3.3. Par contre, nous trouvons que les points détectés sont bien distribués dans l'image en utilisant les seuils « Moyen » et « Adaptatif ». Ces deux seuils résolvent donc les problèmes où la texture est centrée sur une petite zone de l'image. Le seuil « Adaptatif » distribue uniformément les points d'intérêt obtenus, mais il donne plus de points à l'extérieur de la zone de correspondance.

Stabilité

Nous avons utilisé le détecteur de Harris et la précision observée est bonne. De plus, un « détecteur de Harris précis » est présenté dans [Bau96], et le détecteur atteint une excellente stabilité pour détecter les points d'intérêts (un point détecté dans l'image A à une bonne chance d'être détecté dans B). Sa stabilité provient principalement du fait que son implantation n'utilise que les dérivées de l'image d'ordre 1. Nous avons utilisé le détecteur de Harris simple car il est conforme aux exigences de nos expériences et parce que les erreurs du détecteur de Harris sont prises en compte par notre méthode.

Chapitre 4

4 Mise en correspondance

Dans ce chapitre, nous parlons de la mise en correspondance. Dans la section 4.1, nous décrivons les mesures de corrélation et les difficultés liées à l'utilisation de la corrélation. Dans la section 4.2, nous présentons notre méthode complète de mise en correspondance.

Définition de la mise en correspondance

On dispose de deux images d'un objet tridimensionnel, et la mise en correspondance consiste à retrouver la projection d'un point de cet objet dans chacune des deux images. Autrement dit, pour un point p_1 dans une des deux images, qui est la projection d'un point P de la scène, on cherche la projection p_2 de P dans l'autre image. Les points p_1 et p_2 sont correspondants l'un à l'autre et on parle d'appariement ou de mise en correspondance. La difficulté vient du fait qu'on cherche p_2 à partir de p_1 sans connaître P ; les deux images peuvent donc avoir été prises sous des points de vue différents, et ainsi, les conditions d'éclairage peuvent varier.

4.1 Mise en correspondance par corrélation

Une méthode commune de mise en correspondance consiste à mesurer par une corrélation la ressemblance entre deux sous-images centrées sur les points dont on veut savoir s'ils sont en correspondance. Pour une sous-image de la première image, on cherche la sous-image la plus ressemblante dans la deuxième image. Ainsi, cette fenêtre peut correspondre à de nombreuses fenêtres à droite, ce qui entraîne un problème d'ambiguïté.

De manière générale, la technique de recherche d'une correspondance par corrélation se déroule de la manière suivante (voir figure 4.1). Si on cherche le correspondant d'un point p_1 qui est un point de l'ensemble des points d'intérêt de l'image A, on définit tout d'abord la zone de recherche dans l'image B comme l'ensemble des points d'intérêt qu'elle contient. On utilise alors une fenêtre fixe de corrélation F_1 dans l'image A centrée sur le point p_1 , ainsi qu'une

fenêtre glissante F_2 dans l'image B qui va parcourir chaque point d'intérêt. Pour chaque position de F_2 , une valeur de corrélation « $score(x_1, y_1; x_2, y_2)$ » est calculée entre F_1 et F_2 (voir exemples à la figure 4.2).

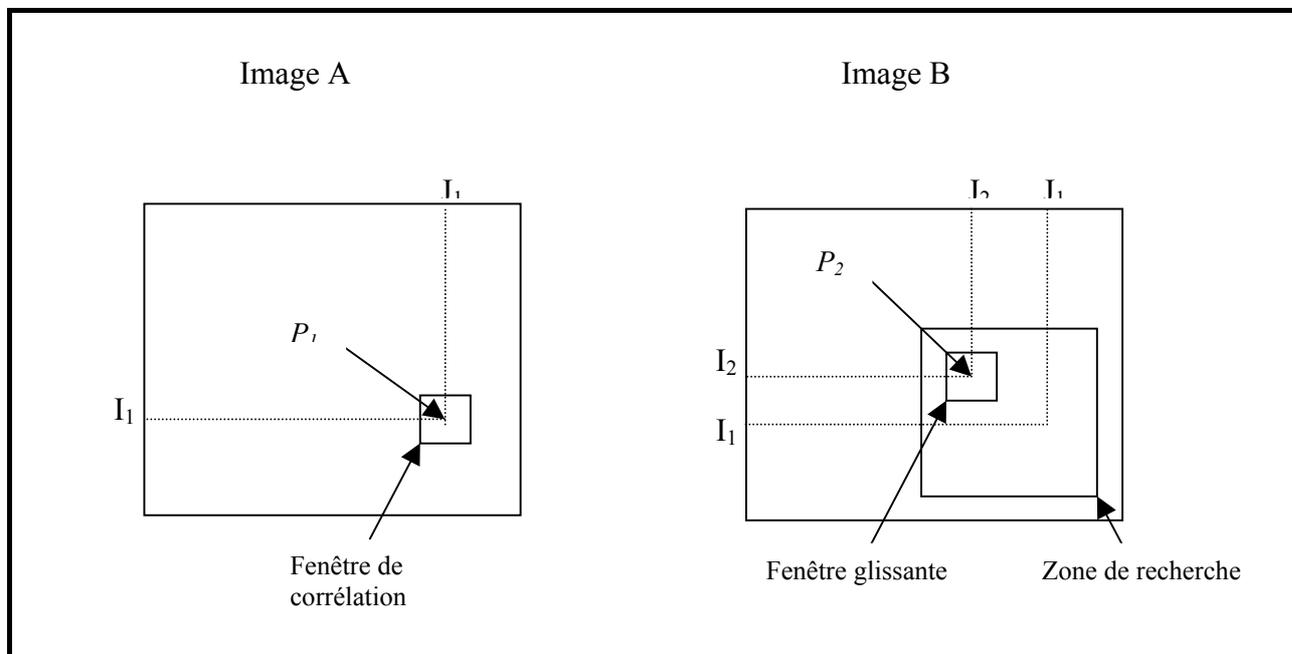


Figure 4.1: Recherche d'un lien par corrélation

Dans la figure 4.1 I_1 et J_1 sont identiques. Les valeurs des scores obtenues varient entre -1 et +1. La valeur -1 correspond à la correspondance d'images inverses; la valeur 0 signifie qu'il n'y a pas aucune correspondance; La valeur +1 désigne une mise en correspondance valide. L'ensemble des valeurs calculées sur la zone de recherche définit la surface de corrélation associée au point p_1 . Le point choisi sera alors celui pour lequel la valeur sera le plus grand.

4.1.1 Mesures de corrélation

Les critères de corrélation sont basés sur une mesure de similarité ou de dissemblance. Une mesure simple de corrélation basée sur la distance euclidienne est la somme des différences au carré :

$$score_1(x_1, y_1; x_2, y_2) = \sum_i \sum_j [I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j)]^2$$

pour une fenêtre de corrélation de taille $(2n + 1) \times (2m + 1)$

où $-n \leq i \leq n$ et $-m \leq j \leq m$.

La mesure $score_1$ se repose donc sur la distance euclidienne; d'autres distances peuvent aussi être utilisées. Le tableau suivant montre quelques mesures de corrélation:

Somme des différences au carré	$\sum_i \sum_j [I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j)]^2$
Somme des valeurs absolues des différences	$\sum_i \sum_j I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j) $
Somme des différences au carré centrées	$\sum_i \sum_j [I_1(x_1 + i, y_1 + j) - \hat{I}_1(x_1, y_1) - I_2(x_2 + i, y_2 + j) - \hat{I}_2(x_2, y_2)]^2$
Somme des valeurs absolues des différences centrées	$\sum_i \sum_j I_1(x_1 + i, y_1 + j) - \hat{I}_1(x_1, y_1) - I_2(x_2 + i, y_2 + j) - \hat{I}_2(x_2, y_2) $
Somme normalisée des différences au carré	$\frac{\sum_i \sum_j [I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j)]^2}{\sum_i \sum_j I_1(x_1 + i, y_1 + j) \sum_i \sum_j I_2(x_2 + i, y_2 + j)}$
Somme normalisée, centrées des différences au carré	$\frac{\sum_i \sum_j [I_1(x_1 + i, y_1 + j) - \hat{I}_1(x_1, y_1) - I_2(x_2 + i, y_2 + j) - \hat{I}_2(x_2, y_2)]^2}{\sqrt{\sum_i \sum_j (I_1(x_1 + i, y_1 + j) - \hat{I}_1(x_1, y_1))^2 \sum_i \sum_j (I_2(x_2 + i, y_2 + j) - \hat{I}_2(x_2, y_2))^2}}$

Table 4-1: Mesures de corrélation

Dans le tableau 4.1, $\hat{I}_1(x_1, y_1)$ et $\hat{I}_2(x_2, y_2)$ représentent les moyennes en intensités des fenêtres 1 et 2.

Ce tableau montre quelques mesures de corrélation connues dont les quatre premières sont considérées comme «classiques». Les autres mesures sont utilisées très fréquemment.

4.1.2 Difficultés liées à l'utilisation de la corrélation

On peut citer quelques difficultés liées à l'utilisation de la corrélation :

Images très bruitées

Lorsque les images à mettre en correspondance sont très bruitées, alors le voisinage de deux points correspondants peut devenir suffisamment différent pour que la valeur de corrélation s'éloigne de son optimum, et de ce fait un mauvais appariement peut être préféré à l'appariement correct.

Différences de luminosité

Lorsque deux caméras sont utilisées, il peut arriver que leurs réglages soient différents. Ceci a pour effet de produire des images avec des luminosités significativement différentes. Dans ce cas, l'invariance de la mesure de corrélation choisie aux changements globaux de luminosité peut s'avérer déterminante.

Occlusions

Lorsqu'une partie d'un objet est visible dans une image et absente dans l'autre, l'appariement des points au voisinage de la frontière d'occlusion pose des difficultés. En effet, l'hypothèse de similarité du voisinage est ici invalidée. Il serait nécessaire de savoir quels sont les points du voisinage qui proviennent effectivement de l'occlusion.

Zones uniformes

Un autre problème très important provient de la présence éventuelle de zones uniformes ou quasi-uniformes. En effet, dans ces régions, les voisinages ne se distinguent que par un éventuel bruit. Inévitablement, des erreurs d'appariement font leur apparition.

C'est pourquoi les résultats d'appariements seront meilleurs si les images sont suffisamment texturées, dans le sens où elles sont fournies en détails. En revanche, une texture répétitive ne convient pas car des points différents présentent alors des voisinages identiques. Le choix du correspondant d'un point ne dépend alors que du bruit.

4.2 Mise en correspondance finale

L'usage de la corrélation est efficace, mais entraîne de nombreuses ambiguïtés. Par exemple, pour une fenêtre de corrélation à gauche, on aura plusieurs fenêtres semblables à droite.

Dans la mise en correspondance par corrélation, l'ensemble des valeurs calculées sur la zone de recherche définit la surface de corrélation associée au point p_l . Le point choisi sera alors celui pour lequel la valeur sera la plus grande. Ce processus sera appliqué à tous les points détectés de l'image A. La plupart des valeurs obtenues sont plus grandes que 0.9, et ils ne sont pas tous de vraies paires de correspondance (voir figure 4.2). Pour cela, l'utilisation d'un seuil nous donne beaucoup de mauvaises correspondances qui proviennent de la ressemblance entre les fenêtres des points cherchées.

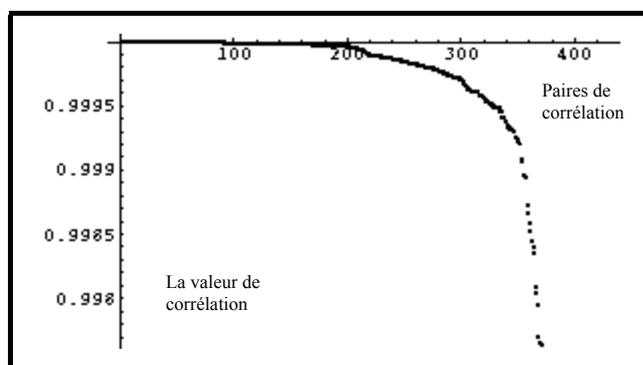


Figure 4.2: La valeur des points détectés.

La figure 4.2 nous montre que la majorité des points détectés ont une valeur plus grande que 0.999. La plupart des paires sont de mauvaises paires de correspondance. Le résultat obtenu est pour une fenêtre de taille 3×3 . Si on augmente la taille, on réduit le nombre des mauvaises paires mais la localisation des points correspondants sera moins précise.

4.2.1 La méthode des triangles

Dans le but de réduire le problème des ambiguïtés de mise en correspondance, nous allons utiliser les relations « géométriques » entre les paires de points pour améliorer la sélection.

La méthode des triangles est constituée de deux étapes :

- choix des paires de référence,
- comparaison par triangulation pour extraire les vraies paires de correspondance.

4.2.2 Extraction des paires de référence

Nous utilisons le critère de corrélation avec une petite fenêtre 3 x 3 pour donner les liens entre les points de l'image A et ceux de l'image B. Ainsi, nous obtenons les paires (a_i, b_j) qui sont en grande partie erronées. Comme nous ne connaissons pas l'angle de rotation de la caméra entre la prise de la première image et la deuxième image, ni le sens de la rotation de la caméra, nous ne pouvons pas prédire la position dans la seconde image d'un pixel de la première image.

Le but est de trouver les deux meilleures paires pour les utiliser comme paires de référence et de les utiliser pour extraire ensuite les vraies paires de la mise en correspondance.

Pour chaque point d'intérêt a_i de l'image A :

- Nous calculons les distances d_{ij}^A entre a_i et tous les autres points a_j .
- Nous calculons le signe de l'angle entre le vecteur $a_i a_j$ et l'axe des abscisses (x).

Pour chaque point b_i de l'image B :

- Nous calculons les distances d_{ij}^B entre b_i et tous les autres points b_j .
- Nous calculons le signe de l'angle entre le vecteur $b_i b_j$ et l'axe des abscisses (x).

Nous comparons les résultats obtenus pour l'image A avec les résultats obtenus pour l'image B. Les deux paires qui donnent le meilleur résultat sont les paires de référence.

Mathématiquement

Parmi les paires trouvées en utilisant la fenêtre de corrélation, on veut trouver deux paires de points qui sont « vraies ».

Soit F, l'ensemble des points d'intérêt détectés dans la première image :

$$F = \{ a_i / a \in \text{image A} \}$$

et G l'ensemble des points d'intérêt détectés dans la deuxième image :

$$G = \{ b_i / b \in \text{image B} \}$$

Soient :

L_{ij}^A : ligne qui relie les points a_i et a_j

L_{ij}^B : ligne qui relie les points b_i et b_j .

Les distances euclidiennes d_{ij}^A et d_{ij}^B utilisées pour calculer les distances entre les points a_i et a_j de l'image A et b_i et b_j de l'image B successivement sont données par :

$$d_{ij}^A = \| L_{ij}^A \|$$

$$d_{ij}^B = \| L_{ij}^B \|$$

Soit S_{ij}^A le signe de l'angle du vecteur $a_i a_j$ et l'axe des abscisses :

$$S_{ij}^A = \text{Signe}((a_j - a_i) \cdot (1, 0))$$

où \cdot est produit scalaire.

Soit S_{ij}^B le signe de l'angle du vecteur $b_i b_j$ et l'axe des abscisses :

$$S_{ij}^B = \text{Signe}((b_j - b_i) \cdot (1, 0))$$

Cette mesure n'est pas fiable dans les cas où les angles ont 90° et 180° comme valeurs, mais ces cas se présentent rarement en pratique.

Nous définissons le critère de signe et le critère de distance

$$S_{ij}^A = S_{ij}^B$$

$$|d_{ij}^A - d_{ij}^B| < D_{\min}$$

D_{\min} est une valeur très petite qui varie entre 2 pixels et 8 pixels.

Soit E_i l'ensemble des points qui vérifient les deux critères .

$$E_i = \{j / |d_{ij}^A - d_{ij}^B| < D_{\min} \text{ et } S_{ij}^A = S_{ij}^B, \forall j \neq i\}$$

Soit V_i le nombre « bons points » défini comme $V_i = \#E_i$. Nous choisissons un ensemble I_{ref} de deux points de référence qui satisfont « géométriquement » au plus grand nombre de points, ce sont les deux points qui donnent le meilleur résultat en appliquant les deux critères (critère de distance et critère de signe) aux autres points et qui est donné par la relation :

$$I_{ref} = \{(a_i, b_i), (a_j, b_j) / V_i \geq V_j \geq V_k, k = 1 \dots N, k \neq i, k \neq j\}$$

où N est le nombre des paires (a_i, b_i) .

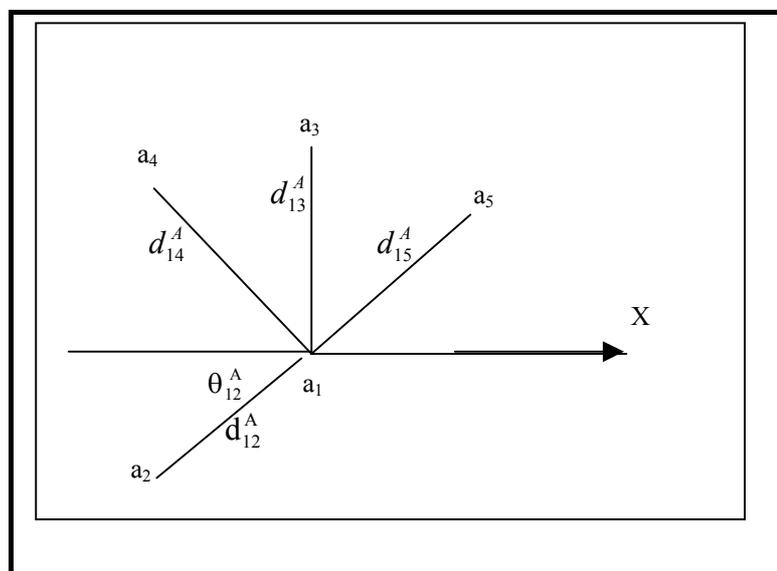
Ces deux points sont considérés « fiables » et serviront de référence pour déterminer le reste des mises en correspondance.

4.2.3 Explication avec un exemple

Voici un exemple qui illustre le choix de l'ensemble de deux points de référence I_{ref} .

On suppose que l'on a l'ensemble de correspondance suivant après corrélation

$$\{\{a_1, b_1\}, \{a_2, b_2\}, \{a_3, b_3\}, \{a_4, b_4\}, \{a_5, b_5\}\}$$



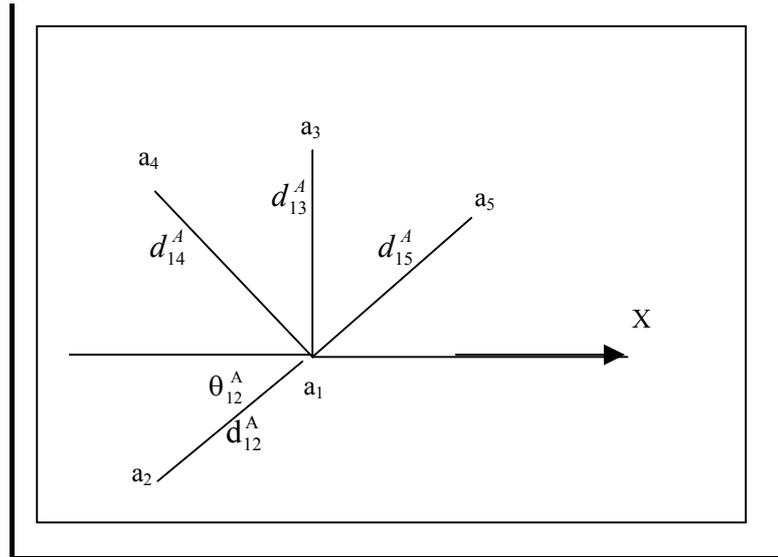


Figure 4.3: Première étape de la méthode du triangle

Les ensembles A et B sont donnés par :

$$A = \{a_1, a_2, a_3, a_4, a_5\}$$

$$B = \{b_1, b_2, b_3, b_4, b_5\}$$

Pour chaque point a_i de A, calculons d_{ij}^A et S_{ij}^A par rapport aux points a_j , tel que $j = \{1 \dots 5, j \neq i\}$

$$d_{1j}^A = \|L_{1j}^A\| = \|a_1 - a_j\|$$

$$S_{1j}^A = \text{Signe}((a_j - a_1) \cdot (1, 0))$$

Pour chaque point b_i de B, calculons d_{ij}^B et S_{ij}^B par rapport aux points b_j , tel que $j = \{1, \dots, 5, j \neq i\}$.

$$d_{ij}^B = \|L_{ij}^B\| = \|b_1 - b_j\|$$

$$S_{1j}^B = \text{Signe}((b_j - b_1) \cdot (1, 0))$$

Nous cherchons E_1 , l'ensemble de points qui vérifient les deux critères pour la paire (a_1, b_1)

$$E_1 = \{j / |d_{1j}^A - d_{1j}^B| < D_{\min} \text{ et } S_{1j}^A = S_{1j}^B \forall j \neq i\}$$

Comme illustré à la figure 4.3 la paire (a_5, b_5) ne vérifie pas le premier critère et sera rejetée.

De même la paire (a_2, b_2) ne vérifie pas le deuxième critère et sera rejetée.

Alors pour la paire (a_1, b_1) :

$$E_1 = \{(a_3, b_3), (a_4, b_4)\}$$

$$\text{et } V_1 = 2$$

Si on refait le même calcul pour chaque paire, on obtient :

- Pour la paire (a_2, b_2) :

$$E_2 = \{\}$$

$$\text{et } V_2 = 0$$

- Pour la paire (a_3, b_3)

$$E_3 = \{(a_1, b_1), (a_4, b_4)\}$$

$$\text{et } V_3 = 2$$

- Pour la paire (a_4, b_4)

$$E_4 = \{(a_1, b_1), (a_3, b_3)\}$$

$$\text{et } V_4 = 2$$

- Pour la paire (a_5, b_5)

$$E_5 = \{\}$$

$$\text{et } V_5 = 0$$

Et finalement, on cherche $I_{ref} = \{(a_i, b_i), (a_j, b_j) / V_i \geq V_j \geq V_k, k = 1 \dots N, k \neq i, k \neq j\}$ et on obtient les points

$$I_{ref} = \{(a_1, b_1), (a_3, b_3)\}$$

Donc, les points de référence pour la figure 4.3 sont les points 1 et 3. Nous avons obtenu trois paires qui donnent les mêmes résultats; nous avons donc choisi arbitrairement une paire parmi les trois.

4.2.4 Exemples de la recherche des points de référence

Dans cette section, nous présenterons des exemples pour montrer le choix des deux paires de points de référence en appliquant les deux critères présentés à la section 4.2.2, et nous aurons des exemples avec un seul critère (i.e. sans utiliser la comparaison de signe des angles). Il y a des exemples déjà illustrés sur les points d'intérêt détectés par Harris aux figures 3.1, 3.2 et 3.3.

Exemple1-a

Dans cet exemple, nous avons utilisé les deux critères pour extraire les deux paires de référence. Nous voyons que la position des deux meilleures paires (i.e. le plus grand V_i) de référence obtenue dans les deux images est la même.

Dans cet exemple nous avons 246 points détectés dans l'image A et 130 points détectés dans l'image B. Nous avons utilisé la méthode du « seuil moyen » pour cet exemple.

Le nombre de vraies paires obtenues est 44, c'est-à-dire qu'elles satisfont le critère géométrique de choix des points de référence (distance et signe de l'angle).

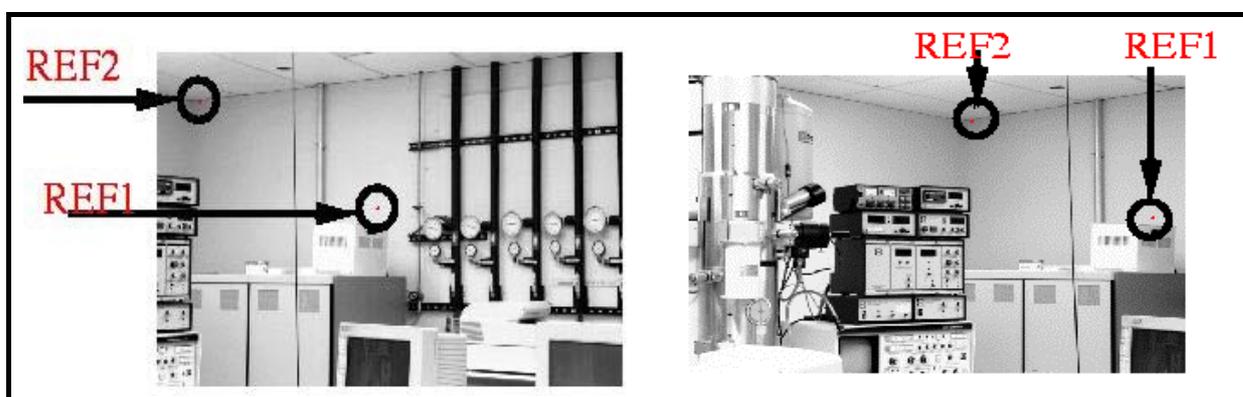


Figure 4.4: Exemple1-a en utilisant les deux critères

Exemple1-b

Ici, on utilise les mêmes images que l'exemple1-a, mais en annulant le critère de comparaison du signe de l'angle. Nous voyons que la position des deux premières paires de référence obtenues dans les deux images est erronée.

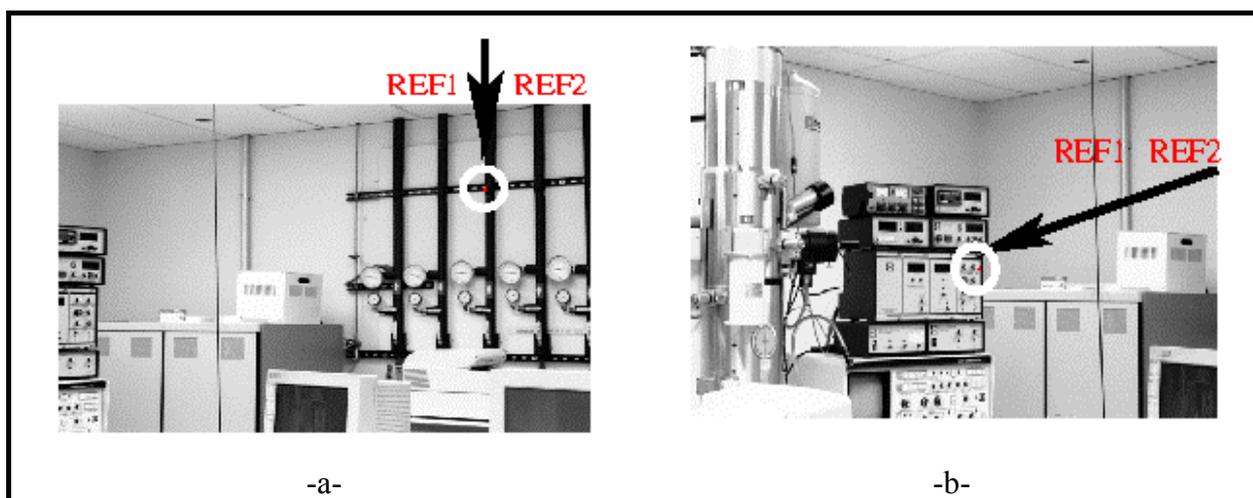


Figure 4.5: Exemple1-b deux points de référence avec un seul critère

Exemple 2

Dans cet exemple, nous avons utilisé les deux critères pour extraire les deux paires de référence. Nous voyons que la position des deux premières paires de référence obtenues dans les deux images est correcte.

Dans cet exemple nous avons 167 points détectés dans l'image A et 270 points détectés dans l'image B. Le nombre de vraies paires qui satisfont au critère géométrique est 28.

La figure 4.6-c montre la distribution des valeurs obtenues pour chacun des points considérés lors de la sélection des points de référence.

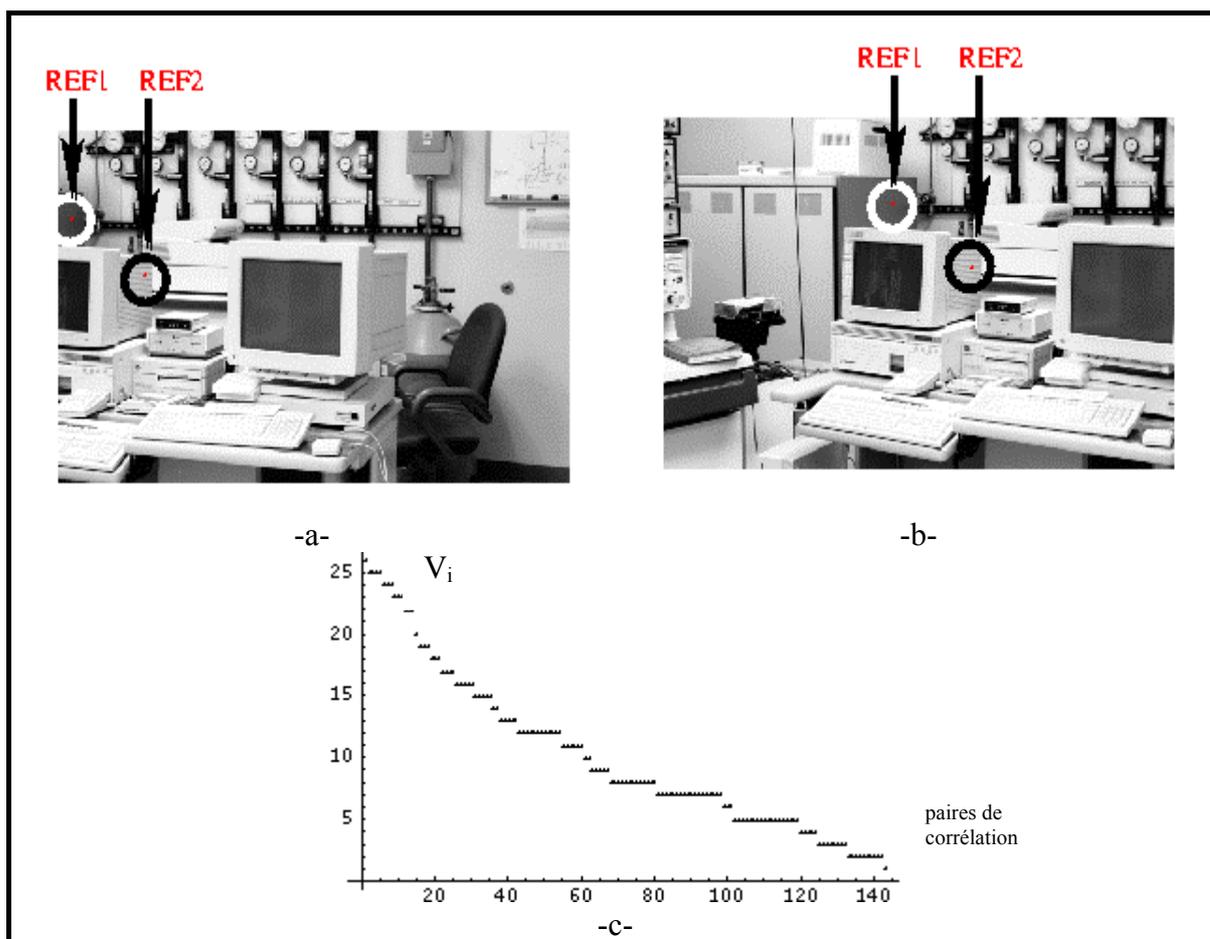


Figure 4.6: Exemple 2 en utilisant les deux critères

Exemple 3

Dans cet exemple, nous avons utilisé les deux critères pour extraire les deux paires de référence. Nous voyons que la position des deux premières paires de référence obtenues dans les deux images est bonne.

Dans cet exemple nous avons 115 points détectés dans l'image A et 190 points détectés dans l'image B.

La figure 4.7-c montre la distribution des valeurs obtenues pour chacun des points considérés lors de la sélection des points de référence.

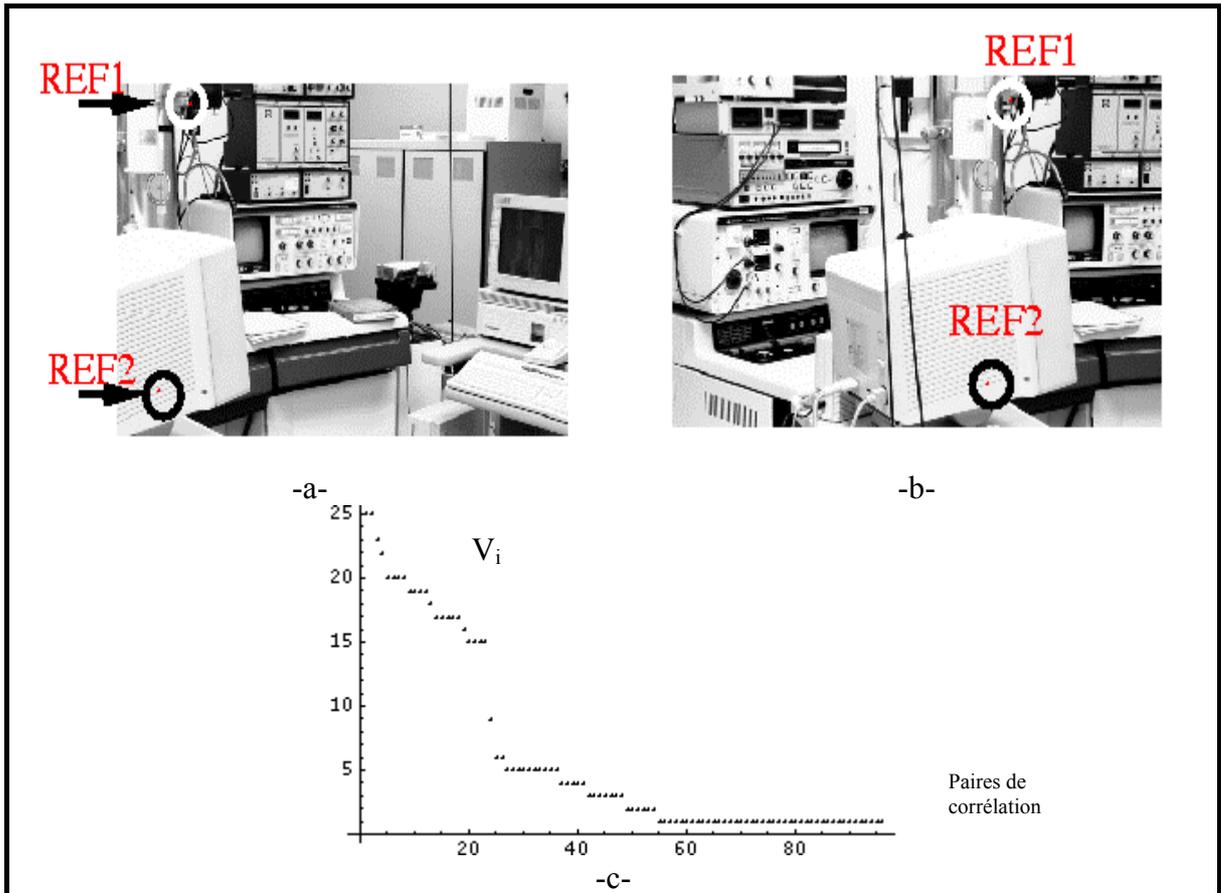


Figure 4.7: Exemple 3 deux points de référence

Exemple 4

Dans cet exemple, nous avons utilisé les deux critères pour extraire les deux paires de référence. Nous voyons que la position des deux premières paires de référence obtenues dans les deux images est bonne.

Dans cet exemple nous avons 207 points détectés dans l'image A et 230 points détectés dans l'image B. Le nombre de vraies paires qui satisfont le critère géométrique est 38.

La figure 4.8-c montre la distribution des valeurs obtenues pour chacun des points considérés lors de la sélection des points de référence.

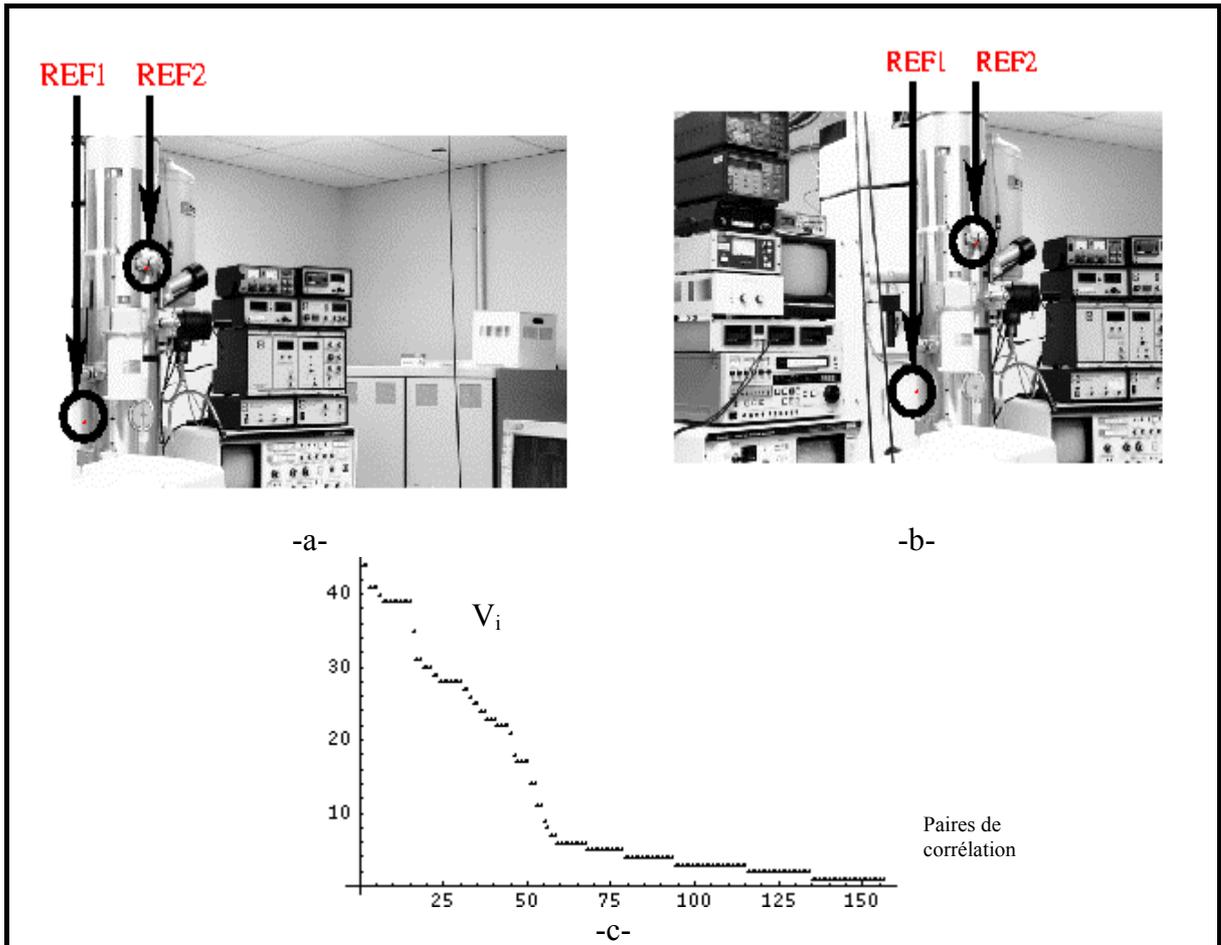


Figure 4.8: Exemple 4 deux points de référence en utilisant les deux critères

Exemple 5-a

Dans cet exemple, nous avons utilisé les deux critères pour extraire les deux paires de référence. Nous voyons que la position des deux premières paires de référence obtenues dans les deux images est correcte.

Dans cet exemple nous avons 197 points détectés dans l'image A et 170 points détectés dans l'image B. Le nombre de vraies paires qui satisfont le critère géométrique est 31.

La figure 4.9-c montre la distribution des valeurs obtenues pour chacun des points considérés lors de la sélection des points de référence.

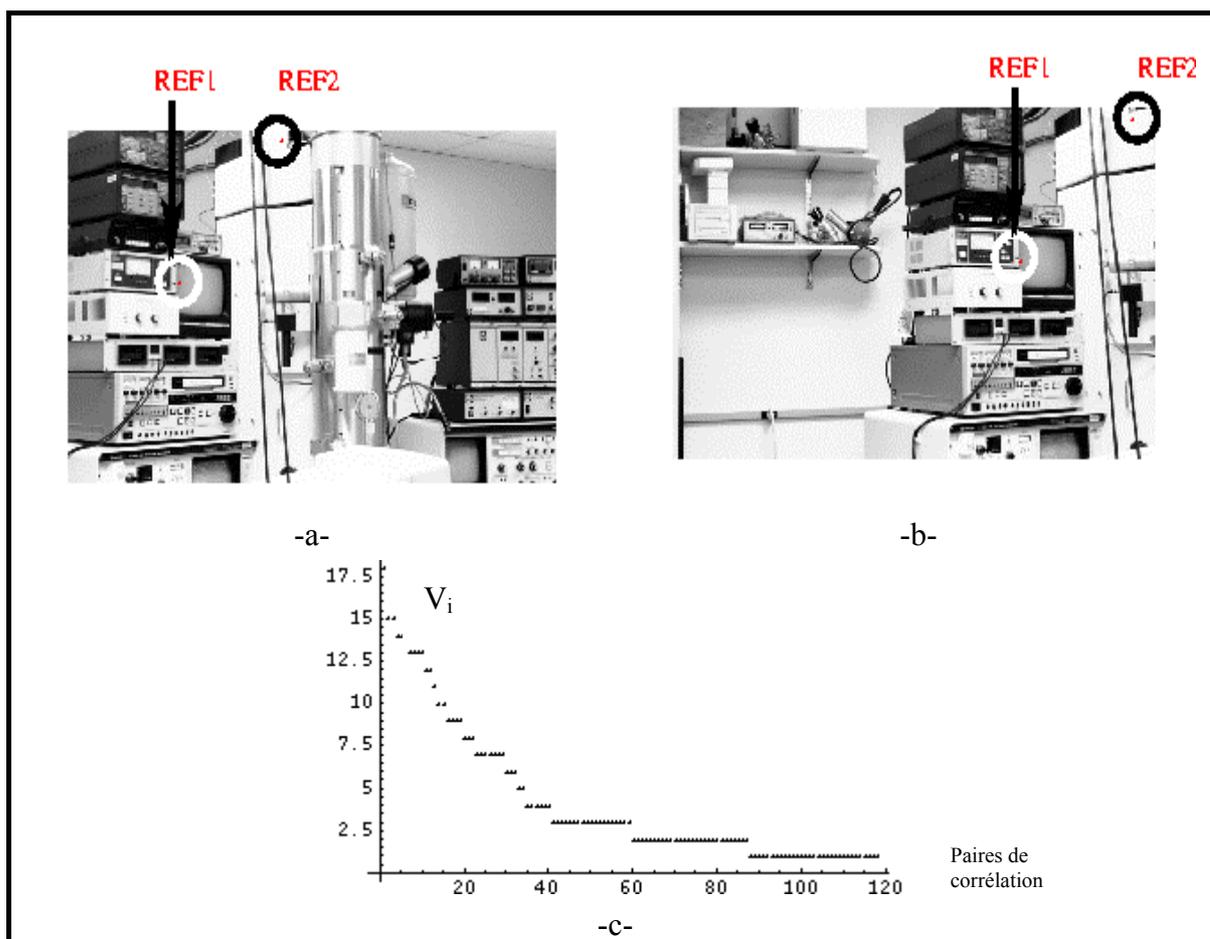


Figure 4.9: Exemple 5 deux points de référence en utilisant les deux critères

Exemple 5-b

Ici, on utilise les mêmes images que l'exemple 5-a, mais sans utiliser le critère de comparaison du signe de l'angle. Nous voyons que la position des deux premières paires de référence obtenues dans les deux images est erronée.



Figure 4.10: Exemple 5 deux points de référence en annulant la comparaison des angles

4.2.5 Comparaison par triangulations

Après avoir travaillé sur l'ensemble de tous les points obtenus par le détecteur de Harris, on doit considérer seulement ceux qui vérifient les deux critères pour l'extraction des points de référence (section 4.2.2).

On peut noter à ce niveau que les points obtenus ne sont pas tous des vraies paires de mise en correspondance. Certains points vérifient les deux critères théoriquement, mais ne correspondent pas, en pratique, à des vraies paires.

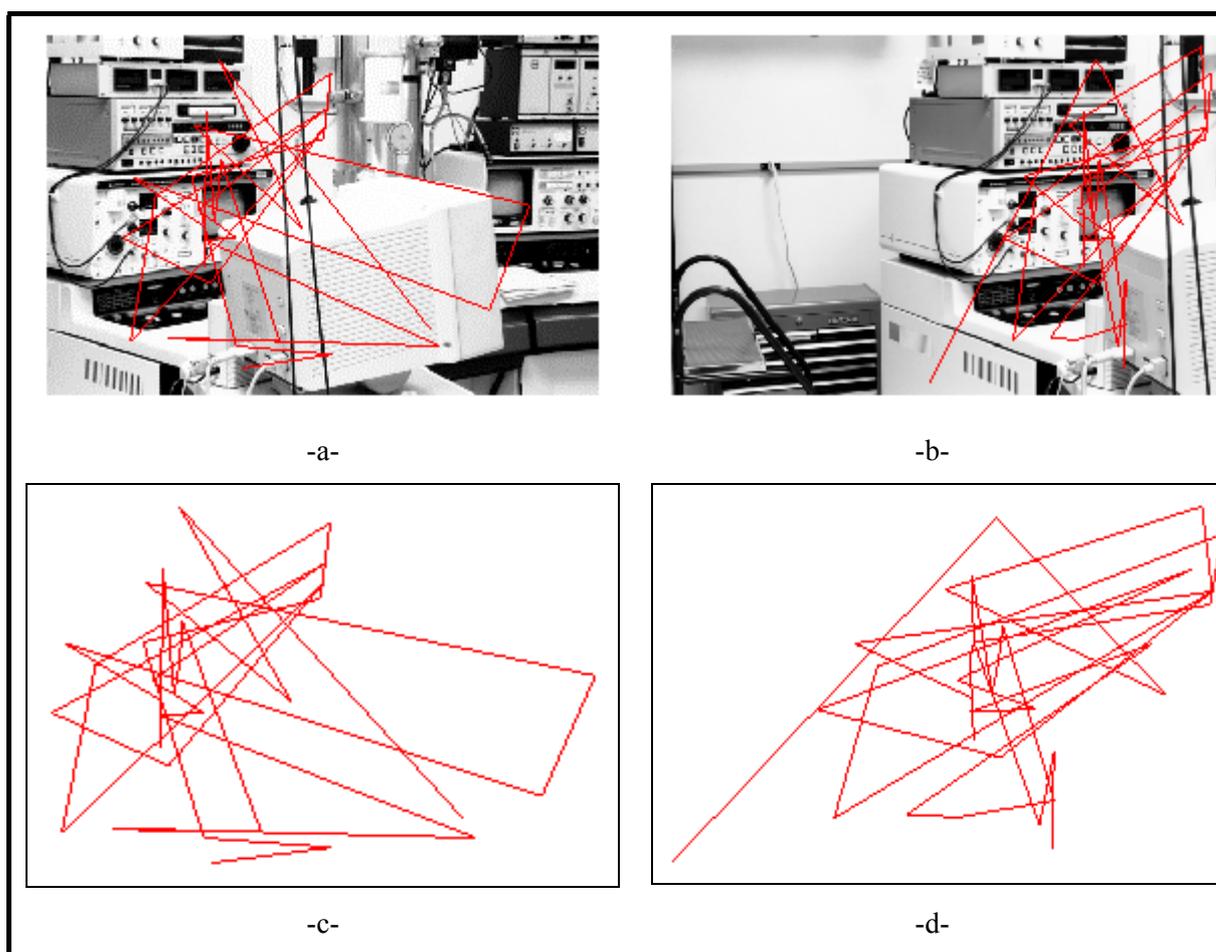


Figure 4.11 : illustre des paires de correspondance erronées.

Par conséquent, il est essentiel d'avoir recours à une deuxième étape, qui permet de trier, filtrer et extraire les vraies paires (qui vérifient la théorie et la pratique).

Cette technique vise à comparer deux triangles, le premier sur l'image A, et le deuxième sur l'image B. Ce triangle sera formé par deux points de référence, ainsi que par la paire de points à vérifier. Si les deux triangles sont idéalement égaux et congrus, avec leurs angles

homologues de mêmes signes, on peut conclure que c'est une vraie paire de points, et on l'accepte. Si ce n'est pas le cas, on rejette cette paire de points.

Pour la comparaison des deux triangles, on utilise les deux critères suivants :

- Premier critère : utilisation d'une fonction coût qui est basée sur la longueur de chaque côté homologue des deux triangles :

- Le triangle de l'image A est formé par les points suivants : REF_1^A , REF_2^A et a_{ij} .
- Le triangle de l'image B est formé par les points suivants : REF_1^B , REF_2^B et b_{ij} .

Cette fonction est donnée par

$$\text{Coût} = d_1 + d_2 + d_3$$

où d_1 , d_2 et d_3 sont les longueurs de côtés du triangle.

- Deuxième critère : les angles homologues des deux triangles celui de l'image A, et celui de l'image B doivent être du même signe.

Mathématiquement, on peut donc traduire qu'une paire de points (a_{ij}, b_{ij}) sera vraie si elle vérifie les deux critères précédant c'est-à-dire :

- $|\text{Coût}_1 - \text{Coût}_2| < \text{seuil}$.
- Les angles homologues de deux triangles sont du même signe.

À cause du processus de sélection de points de référence qui dépend de valeur V , la comparaison par triangulation n'est utilisée que pour détecter les points invalides de l'ensemble de V points considérés valides lors de la sélection de points de référence.

4.2.6 Explication par un exemple

Dans notre exemple illustré à la figure 4.12, on a :

(REF_1^A, REF_1^B) et (REF_2^A, REF_2^B) sont deux paires de points de référence, il faut maintenant vérifier si les autres paires (a_{11}, b_{11}) et (a_{12}, b_{12}) sont des correspondances valides.

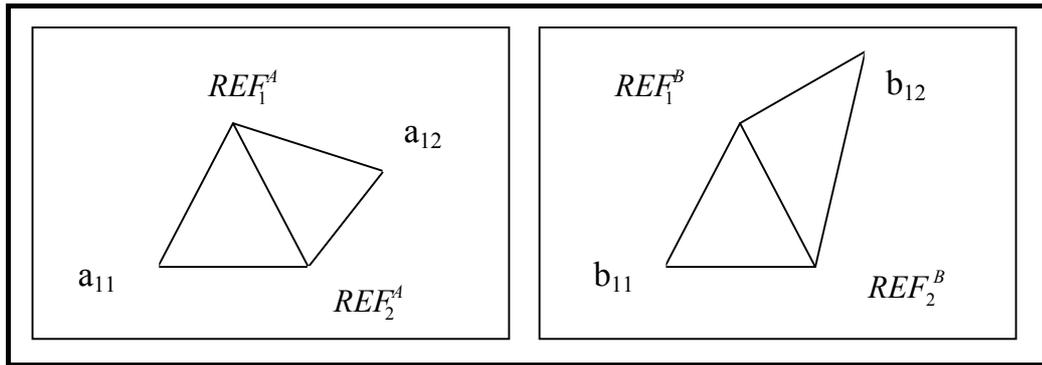


Figure 4.12: Illustration de la méthode de la triangulation par un exemple

Nous vérifions si les deux triangles $(REF_1^A, REF_2^A, a_{11})$ et $(REF_1^B, REF_2^B, b_{11})$ sont égaux pour montrer que (a_{11}, b_{11}) est une vraie paire.

Calculons les fonctions de coût correspondant à ces deux triangles :

$$Coût_1 = d_1^A + d_2^A + d_3^A$$

$$Coût_2 = d_1^B + d_2^B + d_3^B$$

ou nous pouvons comparer si les cotés de deux triangles sont égaux :

$$d_1^A = d_1^B$$

$$d_2^A = d_2^B$$

$$d_3^A = d_3^B$$

avec

$$d_1^A = \|REF_1^A - REF_2^A\|$$

$$d_2^A = \|REF_1^A - a_{11}\|$$

$$d_3^A = \|a_{11} - REF_2^A\|$$

$$d_1^B = \|REF_1^B - REF_2^B\|$$

$$d_2^B = \|REF_1^B - b_{11}\|$$

$$d_3^B = \|b_{11} - REF_2^B\|$$

Si la différence entre $Coût_1$ et $Coût_2$ est plus petite qu'un seuil donné, alors la paire

(a_{11}, b_{11}) est considérée valide. Si non, on la rejette.

Dans cet exemple la paire (a_{12}, b_{12}) est une mauvaise paire parce que les deux triangles $(REF_1^A, REF_2^A, a_{12})$ et $(REF_1^B, REF_2^B, b_{12})$ ne sont pas égaux.

4.2.7 Exemples de la méthode du triangle

Dans cette section, nous présentons des exemples pour montrer les paires de mise en correspondance extraites en utilisant la méthode de triangulation (section 4.2.5).

Exemple 1

Dans la figure 4.13 nous voyons la ressemblance du dessin obtenu dans les deux figures c et d en reliant les points de mise en correspondance extraits avec la méthode de triangulation. Les points d'intérêt sont extraits avec un seuil fixe.

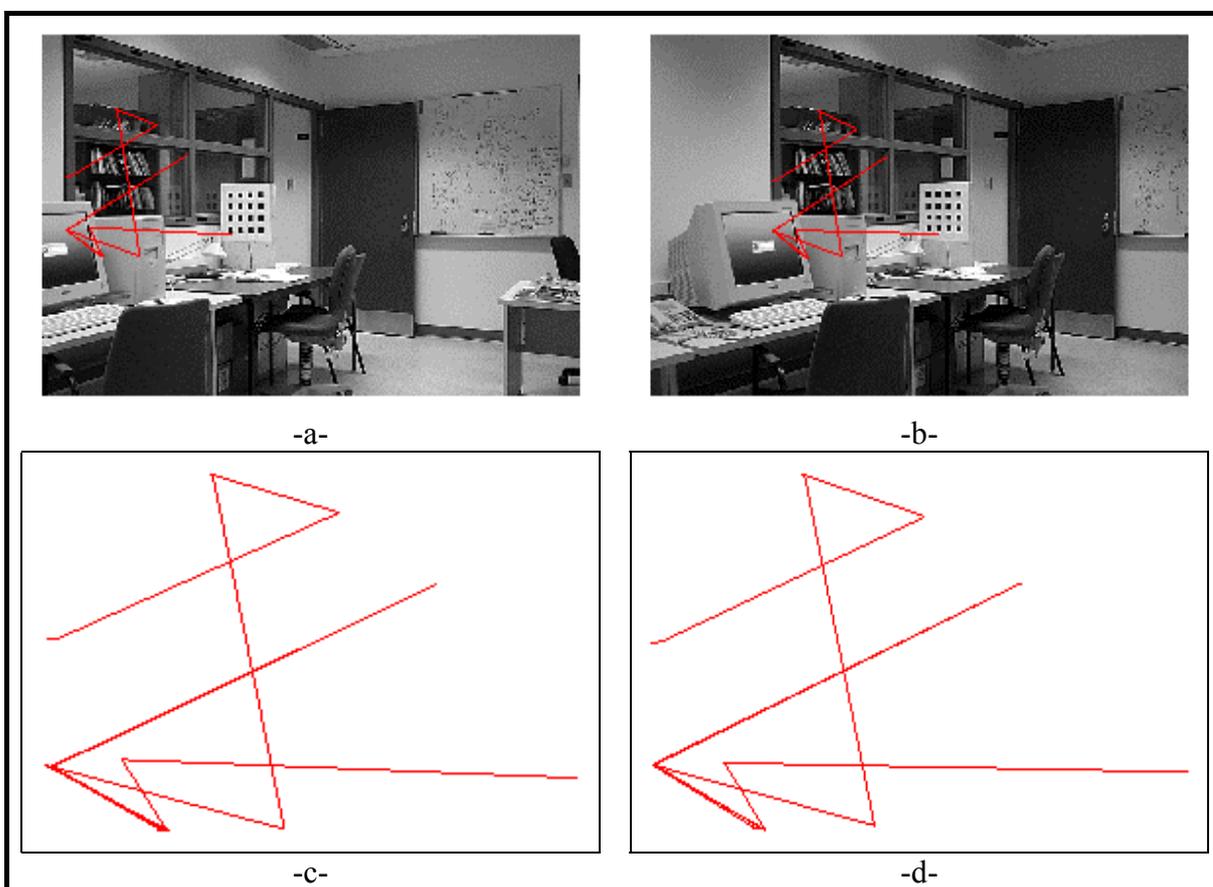


Figure 4.13: Exemple 1 mise en correspondance par triangulation avec un seuil fixe

Exemple 2

Dans la figure 4.14, nous voyons la ressemblance du dessin obtenu dans les deux figures c et d en reliant les points de mise en correspondance extraits en utilisant la méthode de triangulation. Les points d'intérêt sont extraits avec un seuil fixe. Puisque nous avons détecté un grand nombre de points d'intérêt, nous allons obtenir plus de paires de mise en correspondance valides.

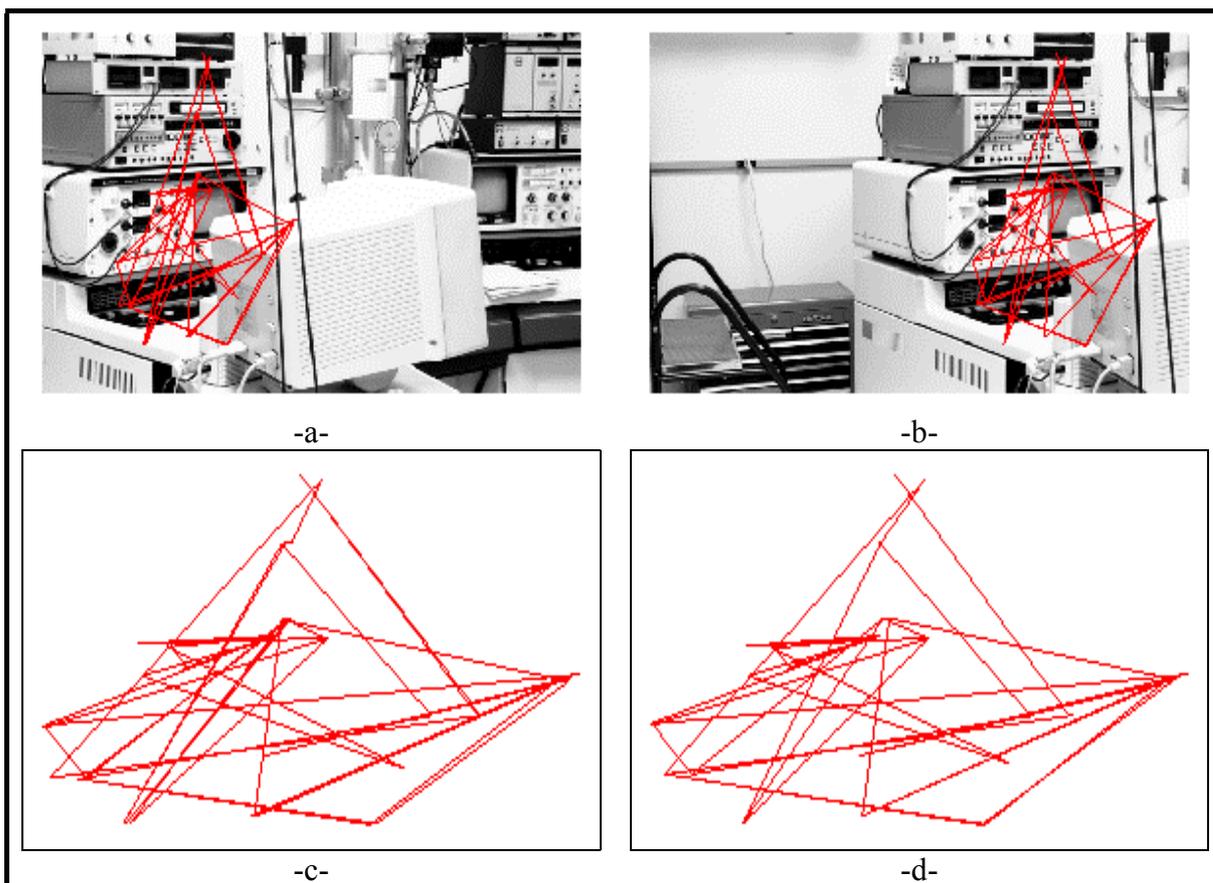


Figure 4.14: Exemple 2 mise en correspondance par triangulation avec un seuil fixe

Exemple 3 :

Dans la figure 4.15, nous voyons la ressemblance du dessin obtenu dans les deux figures c et d en reliant les points de mise en correspondance extraits en utilisant la méthode de triangulation. Les points d'intérêt sont extraits avec un seuil fixe.

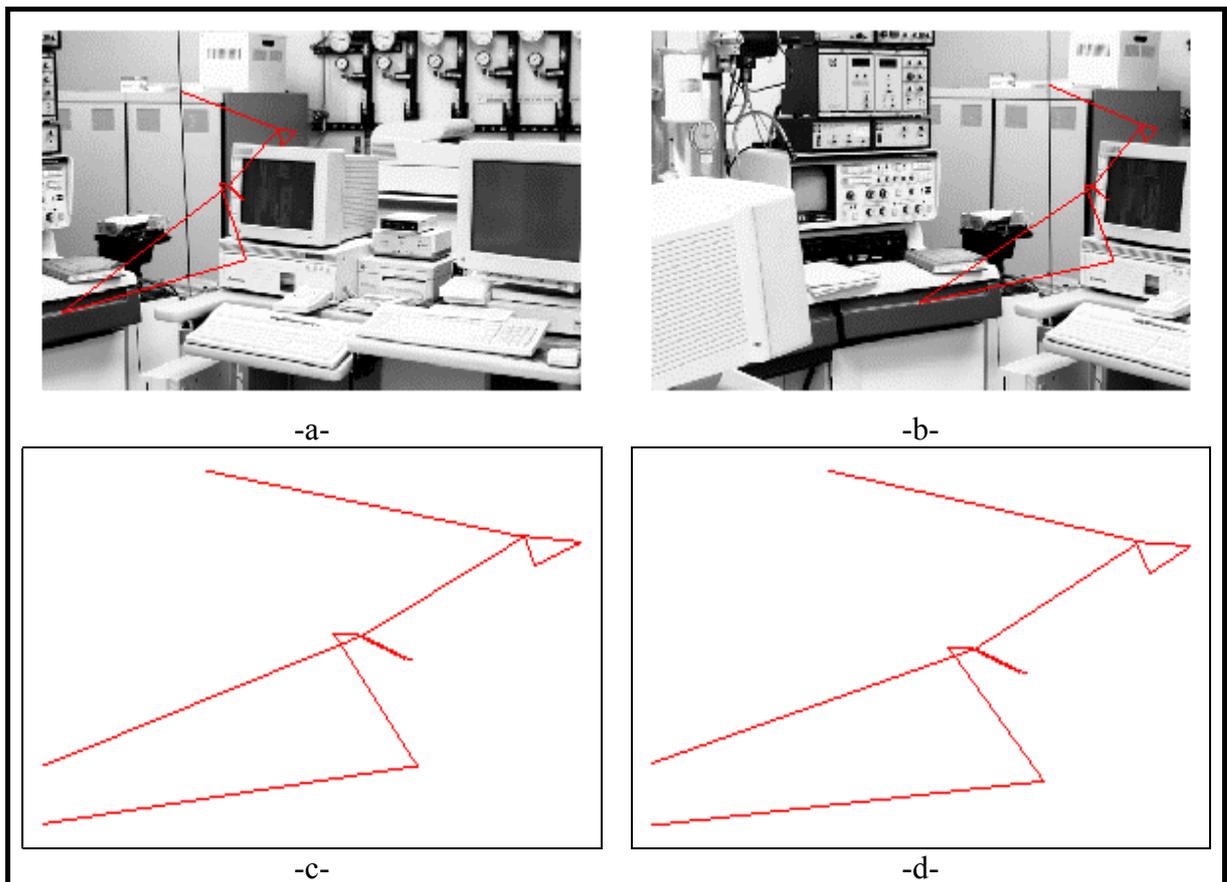


Figure 4.15: Exemple 3 mise en correspondance par triangulation avec un seuil fixe

Exemple 4 :

Dans la figure 4.16, nous voyons la ressemblance du dessin obtenu dans les deux figures c et d en reliant les points de mise en correspondance extraits en utilisant la méthode de triangulation. Les points d'intérêt sont extraits avec un seuil fixe. Même avec le grand nombre de fenêtres de ressemblance qu'on peut trouver dans l'image droite pour plusieurs fenêtres de l'image gauche. Malgré le caractère répétitif des images entraînant une grande ambiguïté de mise en correspondance, notre méthode ne conserve que des paires valides.

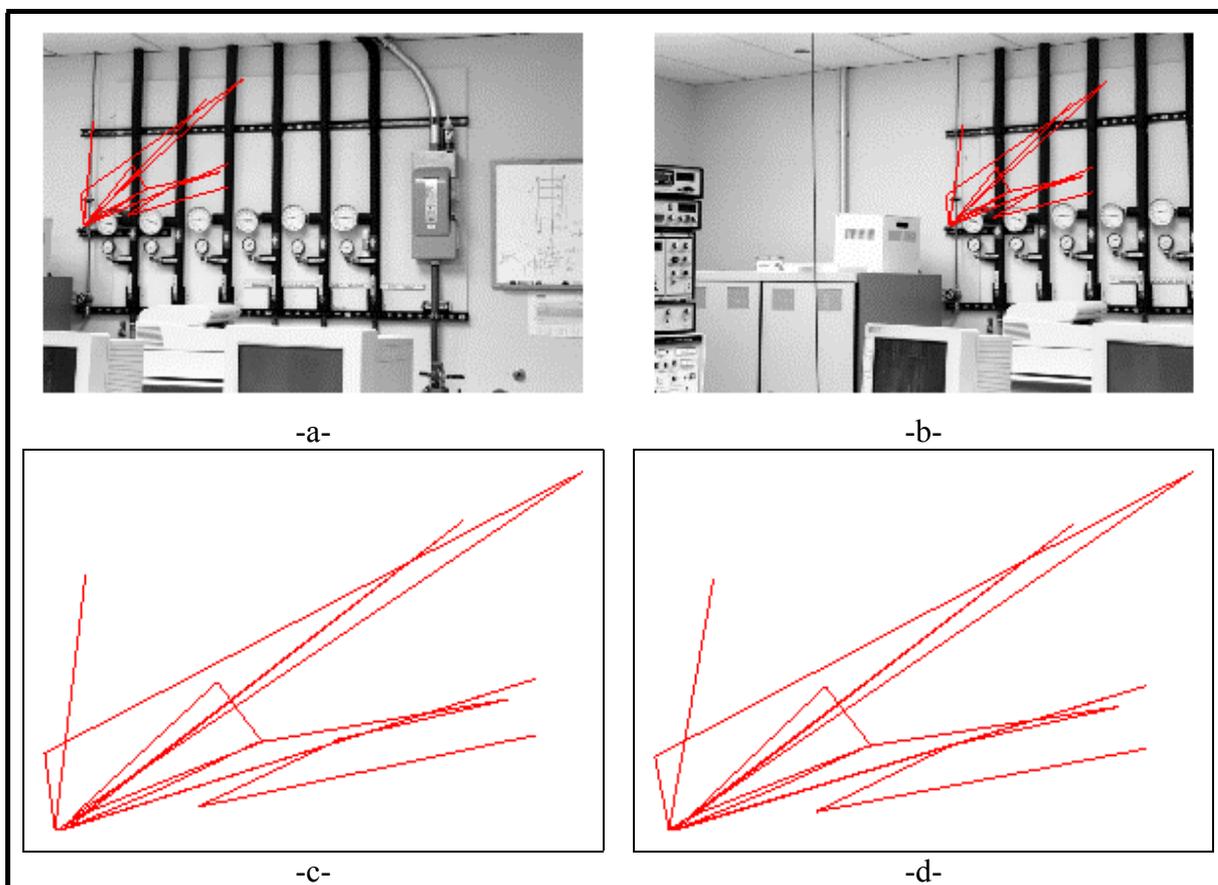


Figure 4.16: Exemple 4 mise en correspondance par triangulation avec un seuil fixe

Exemple avec un seuil variable

Dans la figure 4.17, nous voyons la ressemblance du dessin obtenu dans les deux images (figures 4.16-e et 4.16-f) en reliant les points de mise en correspondance extraits en utilisant la méthode de triangulation. Les points d'intérêt sont extraits avec un seuil adaptatif, ce qui uniformise leur distribution dans les images.

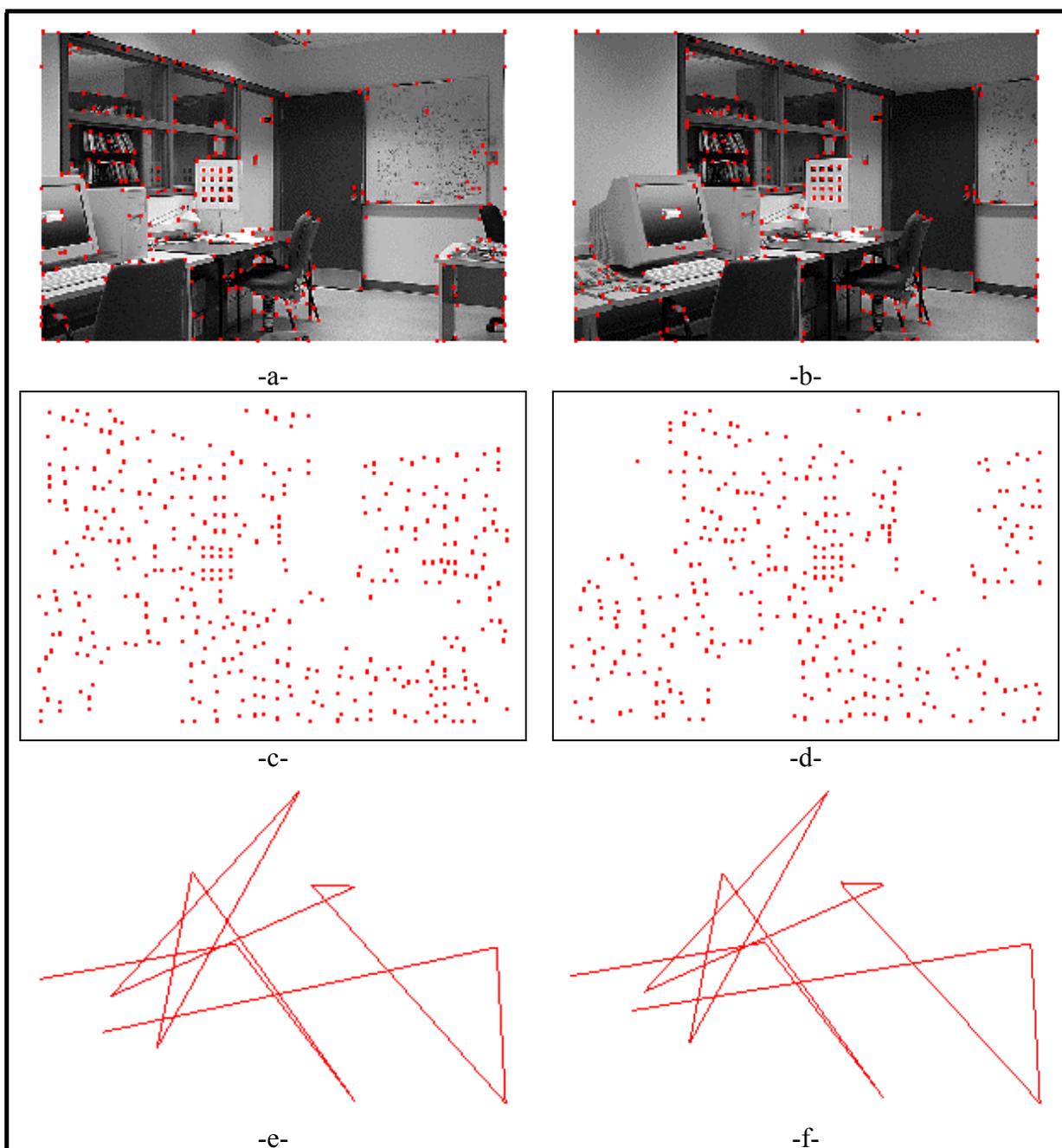


Figure 4.17: Exemple mise en correspondance avec un seuil adaptatif

	Points d'intérêt	V	Paires de mise en correspondance
Figure 4.13	179	32	24
Figure 4.14	361	55	45
Figure 4.15	250	31	15
Figure 4.16	450	30	23
Figure 4.17	328	20	13

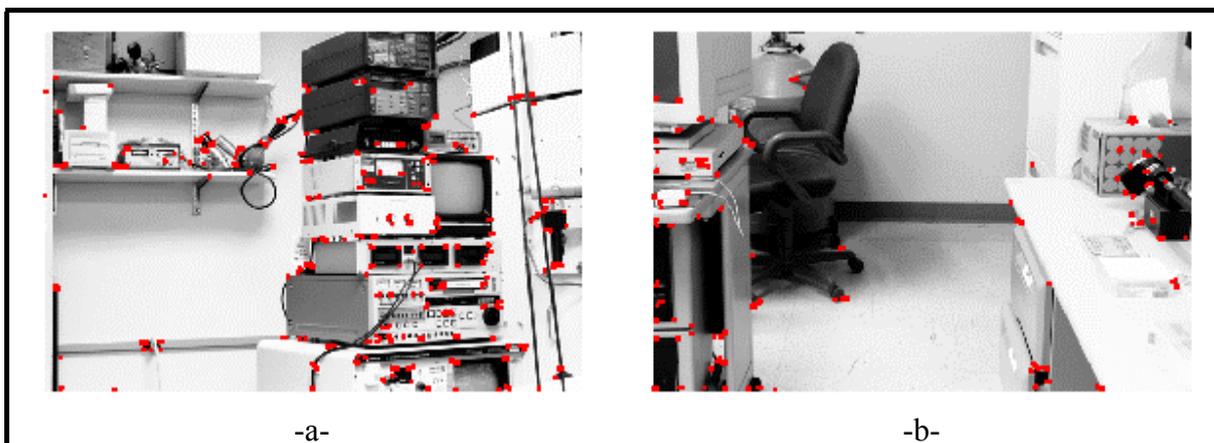
Tableau 4-1 :statistique sur le nombre des points de référence

4.2.8 Échec de la méthode

La méthode est basée sur l'extraction de deux paires de référence. Si l'un d'entre eux n'est pas une paire de correspondance valide, la méthode échouera.

De même, si la corrélation ne donne aucune vraie correspondance, c'est un échec de la méthode. Par exemple, ces situations d'échec de la méthode se présentent lorsque des images de scènes différentes sont mises en correspondance.

Exemple avec deux scènes différentes



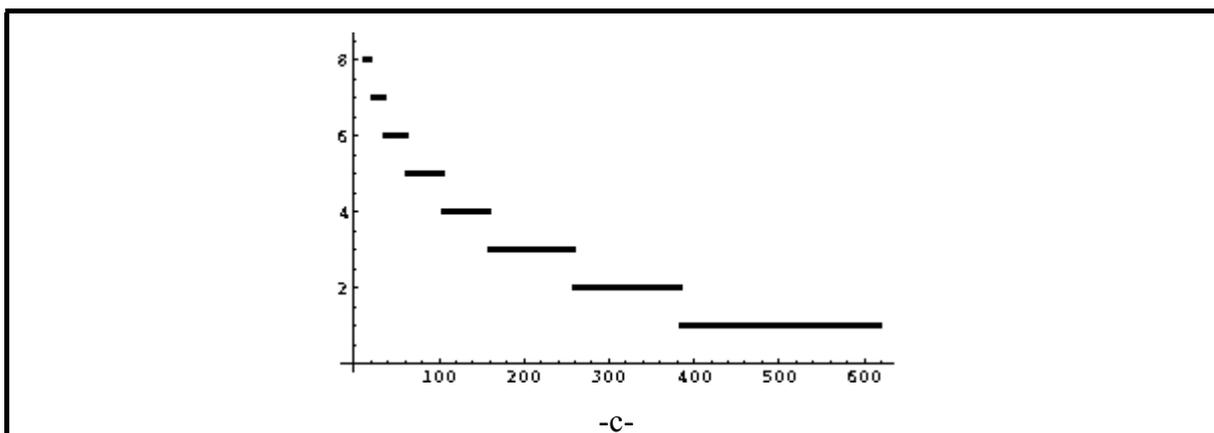


Figure 4.18: Valeurs des scores

Dans cet exemple nous avons 637 points détectés dont la première image et 367 points détectés dans la deuxième image.

On trouve qu'il y a un maximum de huit paires de mise en correspondance qui satisfont aux critères géométriques de sélection des points de référence, et c'est un nombre très petit par rapport aux points détectés (voir le tableau 4-1).

On peut donc facilement détecter les cas où il n'y a aucune correspondance entre les images.

4.2.9 Avantage de la méthode

La corrélation « classique » avec un seuil donne souvent trop de points d'intérêt similaires, et qui ne peuvent pas être mis en correspondance sans ambiguïté. Notre méthode corrige cela en tenant compte de la géométrie des structures triangulaires correspondantes. Cette information supplémentaire aide à réduire l'ambiguïté.

De plus, notre seuil adaptatif uniformise la distribution des points d'intérêt, ce qui rend la méthode plus robuste aux variations de texture présentées dans les images.

Les points qui sont ajoutés dans l'image sont des points moins forts pour le détecteur de Harris mais ils sont néanmoins valides et ils ont des correspondances à l'autre image.

Chapitre 5 :

5 Calibration et construction de mosaïque

Dans ce chapitre, nous présentons les deux dernières étapes de notre algorithme : la calibration des caméras et la construction de la mosaïque. La calibration et l'homographie sont d'abord expliquées à la section 5.1, ensuite dans la section 5.2, nous décrivons la construction de la mosaïque.

5.1 Calibration et homographie

Dans cette section, nous allons décrire le modèle géométrique associé au processus de saisie d'images à l'aide d'une caméra. Ce modèle est caractérisé par un certain nombre de paramètres que nous estimons par étalonnage (ou calibration, ou calibrage). Calibrer une caméra revient à établir les paramètres de la projection des coordonnées du monde (3D) vers celles de l'image (2D). Ces paramètres sont nécessaires dans la plupart des applications de reconstruction de modèles où la relation entre les coordonnées est utile.

5.1.1 Modèle géométrique d'une caméra

Le modèle sténopé se caractérise par un plan de projection, le plan image, et un centre de projection, le centre optique F . La figure 5.1 montre qu'un point P de la scène se projette sur le plan image en un point p , qui est l'intersection de la droite (FP) avec le plan image.

Avec ce modèle, le processus d'obtention d'une image peut être décrit de manière synthétique par la matrice de transformation. Il s'agit en fait de la matrice qui décrit la transformation qui permet de passer des coordonnées de P exprimées dans un repère absolu tridimensionnel, à sa projection p sur le plan image décrite par ses coordonnées exprimées en pixels dans un repère bidimensionnel lié à l'image plane.

Si les coordonnées de P dans le repère absolu sont (X, Y, Z) alors les coordonnées (x, y) de sa projection p dans le repère image sont obtenues par une transformation linéaire M en coordonnées homogènes qui s'écrit sous forme matricielle :

$$\begin{pmatrix} u \\ v \\ h \end{pmatrix} = \begin{pmatrix} hx \\ hy \\ h \end{pmatrix} = M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Les coordonnées cartésiennes de p dans le repère image sont simplement données par :

$$\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \frac{hx}{h} \\ \frac{hy}{h} \end{pmatrix} = \begin{pmatrix} \frac{u}{h} \\ \frac{v}{h} \end{pmatrix}$$

où M est la matrice de transformation.

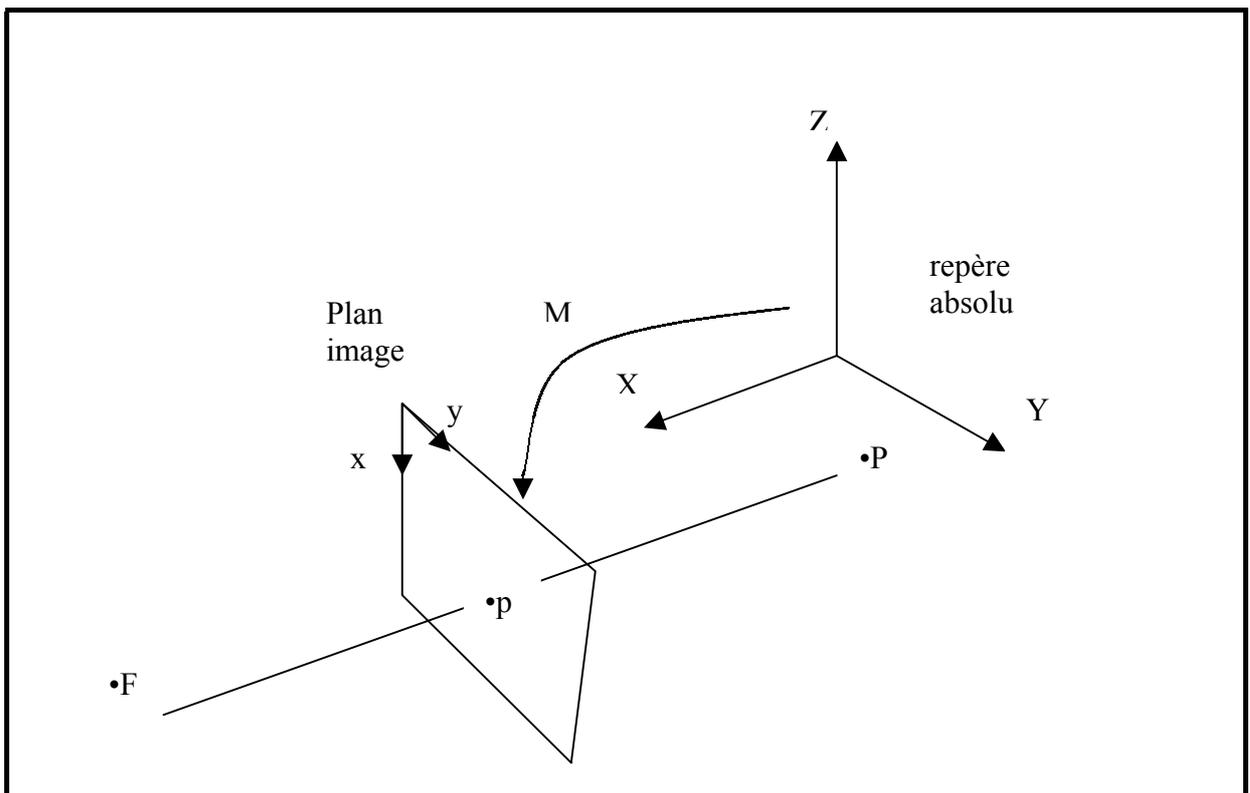


Figure 5.1: Modèle géométrique d'une caméra

5.1.2 Paramètres d'une caméra

Les paramètres du modèle de caméra décrivent la transformation entre les systèmes de coordonnées (Image, Caméra, Scène).



On sépare les paramètres en deux grandes catégories :

- Paramètres externes décrivant la transformation entre la scène et la caméra.
- Paramètres internes décrivant la transformation entre la caméra et l'image.

Matrice des paramètres externes

Cette matrice exprime simplement le passage du repère mire au repère caméra. C'est une transformation rigide (rotation + translation) qui peut s'écrire sous la forme :

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

La transformation rigide peut s'écrire sous la forme d'une transformation homogène :

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{pmatrix} r & t \\ 0 & 1 \end{pmatrix}$$

Cette matrice représente une transformation rigide (3 rotations et 3 translations).

Matrice des paramètres internes

Cette matrice notée A représente la transformation qui permet de passer des coordonnées images normalisées (dans le système de coordonnées de la caméra) vers des coordonnées images (x, y) en pixel. Cette transformation peut s'écrire sous la forme

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = A \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix}$$

La matrice A est donnée par

$$A = \begin{pmatrix} -f_x & 0 & o_x \\ 0 & -f_y & o_y \\ 0 & 0 & 1 \end{pmatrix}$$

où (o_x, o_y) sont les coordonnées de l'origine dans le repère image (mesurées en pixels) du centre de projection, f_x et f_y sont les coordonnées du point focal dans le système 2D de l'image.

Différents modèles

Les différences de modèles de caméra proviennent de la matrice de projection utilisée dans les paramètres internes. La scène se projette sur un plan image à partir d'un modèle caméra. Cette projection s'obtient par la relation:

$$\begin{pmatrix} u \\ v \\ h \end{pmatrix} = \begin{pmatrix} hx \\ hy \\ h \end{pmatrix} = JART \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

où

A : paramètres internes du plan image donnée par

$$A = \begin{pmatrix} -f_x & 0 & o_x & 0 \\ 0 & -f_y & o_y & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

R : matrice de rotation donnée par

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

T : matrice de translation donnée par

$$T = \begin{pmatrix} 1 & 0 & 0 & t_x \\ 0 & 1 & 0 & t_y \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

J : matrice de projection différente d'un modèle à l'autre

a- Modèle perspectif

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

b- Modèle perspectif faible (\bar{Z} est la profondeur moyenne dans la scène)

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & \bar{Z} \end{pmatrix}$$

c- Modèle orthographique

$$J = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

5.1.3 Calibration d'un modèle affine

Le modèle affine diffère des autres modèles présentés, car les paramètres internes et externes sont donnés par une seule matrice qu'on appelle matrice de transformation M :

$$\begin{pmatrix} u \\ v \\ h \end{pmatrix} = \begin{pmatrix} hx \\ hy \\ h \end{pmatrix} = \underbrace{\begin{pmatrix} m_1 & m_2 & m_3 & m_4 \\ m_5 & m_6 & m_7 & m_8 \\ m_9 & m_{10} & m_{11} & m_{12} \end{pmatrix}}_M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Calibrer une caméra va consister à estimer la matrice M en utilisant un minimum de 8 points non coplanaires mis en correspondance.

Soit une liste de points correspondants entre l'image $p_i = (x_i, y_i)$ et la scène

$$P_i = (X_i, Y_i, Z_i).$$

On doit résoudre le modèle suivant :

$$\begin{pmatrix} u_i \\ v_i \\ h_i \end{pmatrix} = \begin{pmatrix} h_i x_i \\ h_i y_i \\ h_i \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix}$$

Chaque paire donne deux équations :

$$x_i = \frac{m_{11}X_i + m_{12}Y_i + m_{13}Z_i + m_{14}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

$$y_i = \frac{m_{21}X_i + m_{22}Y_i + m_{23}Z_i + m_{24}}{m_{31}X_i + m_{32}Y_i + m_{33}Z_i + m_{34}}$$

On définit le système linéaire homogène :

$$Am = 0$$

avec

$$A = \begin{bmatrix} u_1 \\ v_1 \\ u_2 \\ v_2 \\ \cdot \\ \cdot \\ u_n \\ v_n \end{bmatrix} \quad m = \begin{bmatrix} m_{11} \\ m_{12} \\ \cdot \\ \cdot \\ \cdot \\ m_{34} \end{bmatrix}$$

avec $h=1$

$$u_i = [X_i \quad Y_i \quad Z_i \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad -x_i X_i \quad -x_i Y_i \quad -x_i Z_i \quad -x_i \quad]$$

$$v_i = [0 \quad 0 \quad 0 \quad 0 \quad X_i \quad Y_i \quad Z_i \quad 1 \quad -y_i X_i \quad -y_i Y_i \quad -y_i Z_i \quad -y_i \quad]$$

Donc, on résout par la décomposition en valeurs singulières (SVD) de A telle que $A = UDV^T$, et en choisissant la colonne de V^T correspondant à la plus petite valeur singulière.

5.2 Construction de la mosaïque

La mosaïque est le résultat qu'on cherche à obtenir en projetant plusieurs images à partir d'un repère cylindrique ou sphérique.

La construction de la mosaïque se fait selon la manière suivante. On sait que chaque point sur la mosaïque a une correspondance unique dans la scène, et que chaque point de la scène peut avoir une correspondance dans plusieurs images.

Nous considérons comme hypothèse qu'il n'y a aucune translation pour la caméra cela signifie que la matrice de translation est la matrice identité. Donc, notre modèle est:

$$\begin{pmatrix} h_i x_i \\ h_i y_i \\ h_i \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & 0 \\ m_{21} & m_{22} & m_{23} & 0 \\ m_{31} & m_{32} & m_{33} & 0 \end{pmatrix} \begin{pmatrix} X_i \\ Y_i \\ Z_i \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} \\ m_{21} & m_{22} & m_{23} \\ m_{31} & m_{32} & m_{33} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

On ne fait pas d'hypothèses sur les paramètres internes des caméras. Ils peuvent varier et ils sont implicitement pris en compte par le modèle affine que nous utilisons.

En choisissant un repère cylindrique, on peut représenter un point de la mosaïque qui est donné par les coordonnées (θ, t) , où θ est l'angle autour de l'axe du cylindre et t la hauteur sur le cylindre. Dans la mosaïque, θ correspond à l'axe x de l'image panoramique et h correspond à l'axe y de l'image, r est le rayon du cylindre. Un point de la scène est donné par $P(X, Y, Z, 1)$ dont la relation entre ces deux points est la suivante :

$$\begin{aligned} X &= r \cos\theta \\ Y &= t \\ Z &= r \sin\theta \end{aligned}$$

Il faut tout d'abord, pour un point de la mosaïque (θ, t) trouver le point correspondant $P(X, Y, Z, 1)$ dans la scène.

Il s'agit ensuite de projeter le point obtenu sur toutes les caméras, en utilisant la matrice de projection de chaque caméra M_i pour connaître les images qui contiennent ce point

$$\begin{pmatrix} u \\ v \\ h \end{pmatrix} = M_i \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = M_i \begin{pmatrix} r \cos\theta \\ t \\ r \sin\theta \end{pmatrix}$$

On peut calculer finalement l'intensité du point (θ, t) par la moyenne des intensités de la projection du point $P(X, Y, Z, 1)$ dans chaque image.

Chapitre 6 :

6 Résultats et discussion

Nous avons validé notre méthode de construction de mosaïque avec une série d'expérimentations sur des images réelles. Les cas étudiés sont :

- Seuil moyen
- Seuil adaptatif

Les images mosaïques obtenues peuvent avoir l'air flou car ce sont des résultats obtenus à basse résolution dans le logiciel Mathématicien.

Exemples avec des seuils moyens

La figure 6.1 montre un exemple qui contient beaucoup d'ambiguïtés de mise en correspondance (voir figures 6.1-a et 6.1-b) à cause de la nature répétitive de la scène. Les points correspondants (reliés par des lignes) sont illustrés aux figures 6.1-c et 6.1-d, et la figure 6.1-e montre l'image mosaïque obtenue.

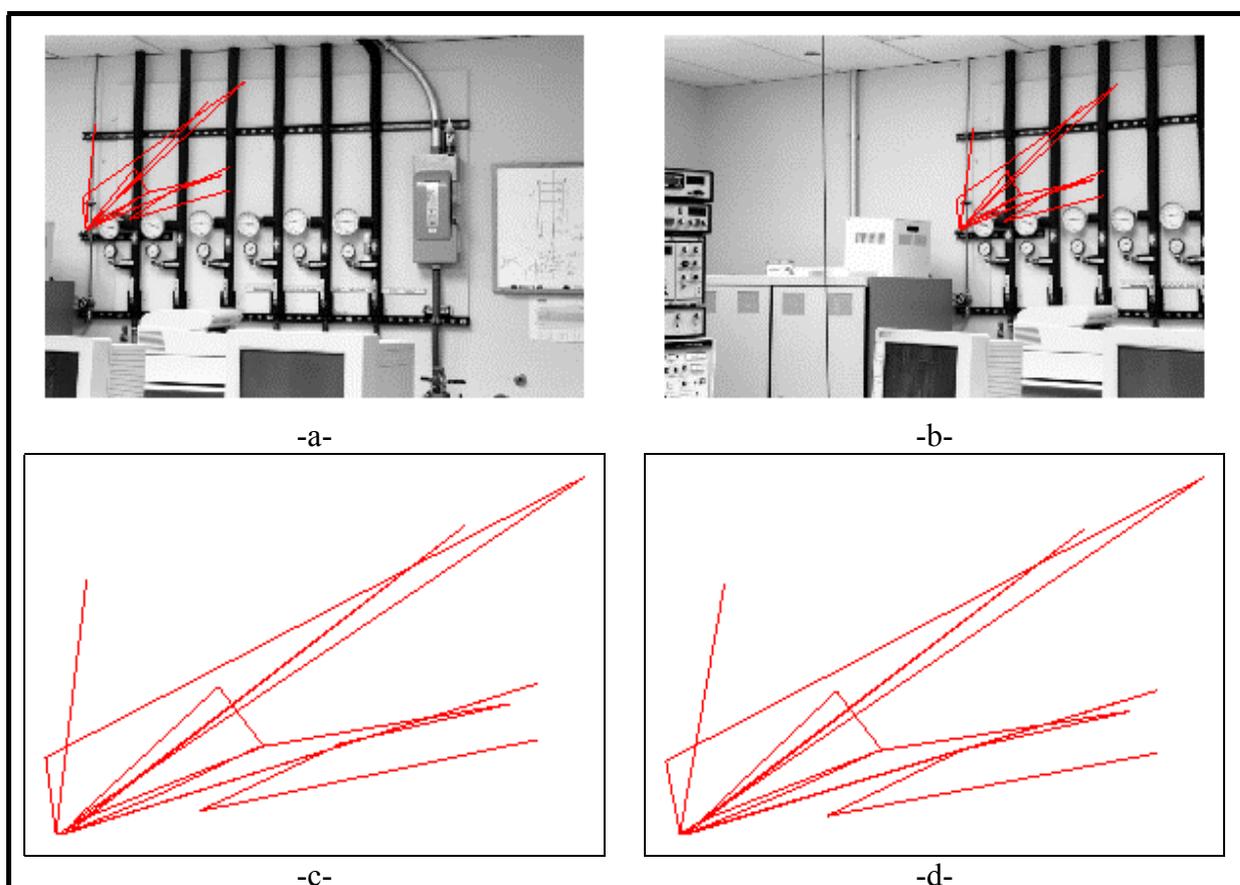
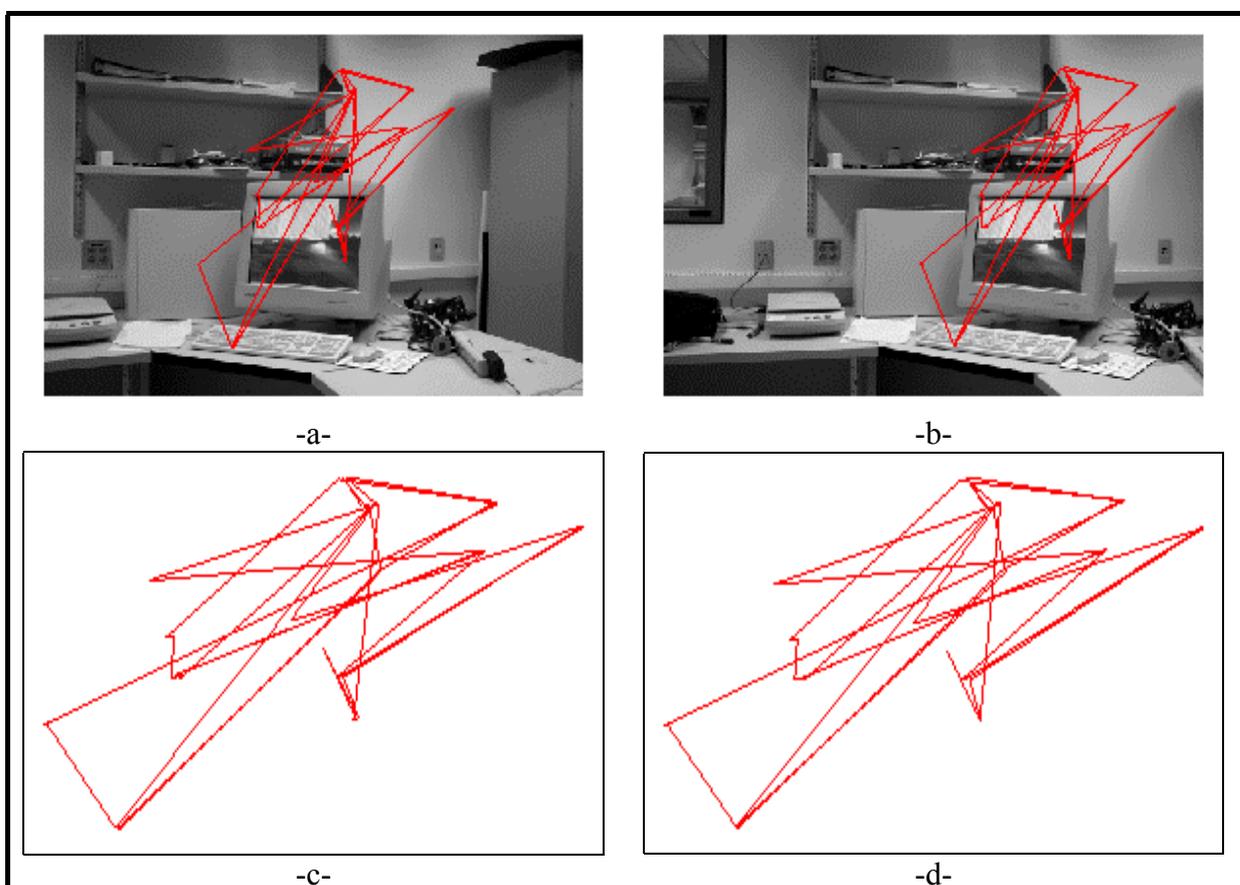




Figure 6.1: Exemple de construction de la mosaïque avec un seuil moyen

La figure 6.2 montre un exemple où il y a des textures partout dans les images (figures 6.2-a et 6.2-b). On remarque qu'il y a plus de points de mise en correspondance que dans l'exemple précédent. Les points correspondants (reliés par des lignes) sont illustrés aux figures 6.2-c et 6.2-d. Les figures 6.2-e et 6.2-f illustrent la calibration des caméras et montre l'efficacité de la mise en correspondance puisque nous avons une ligne qui relie chaque point du premier plan (image A) à sa projection dans la scène en passant par sa correspondance se trouvant dans le deuxième plan (image B). La figure 6.2-g montre la mosaïque obtenue.



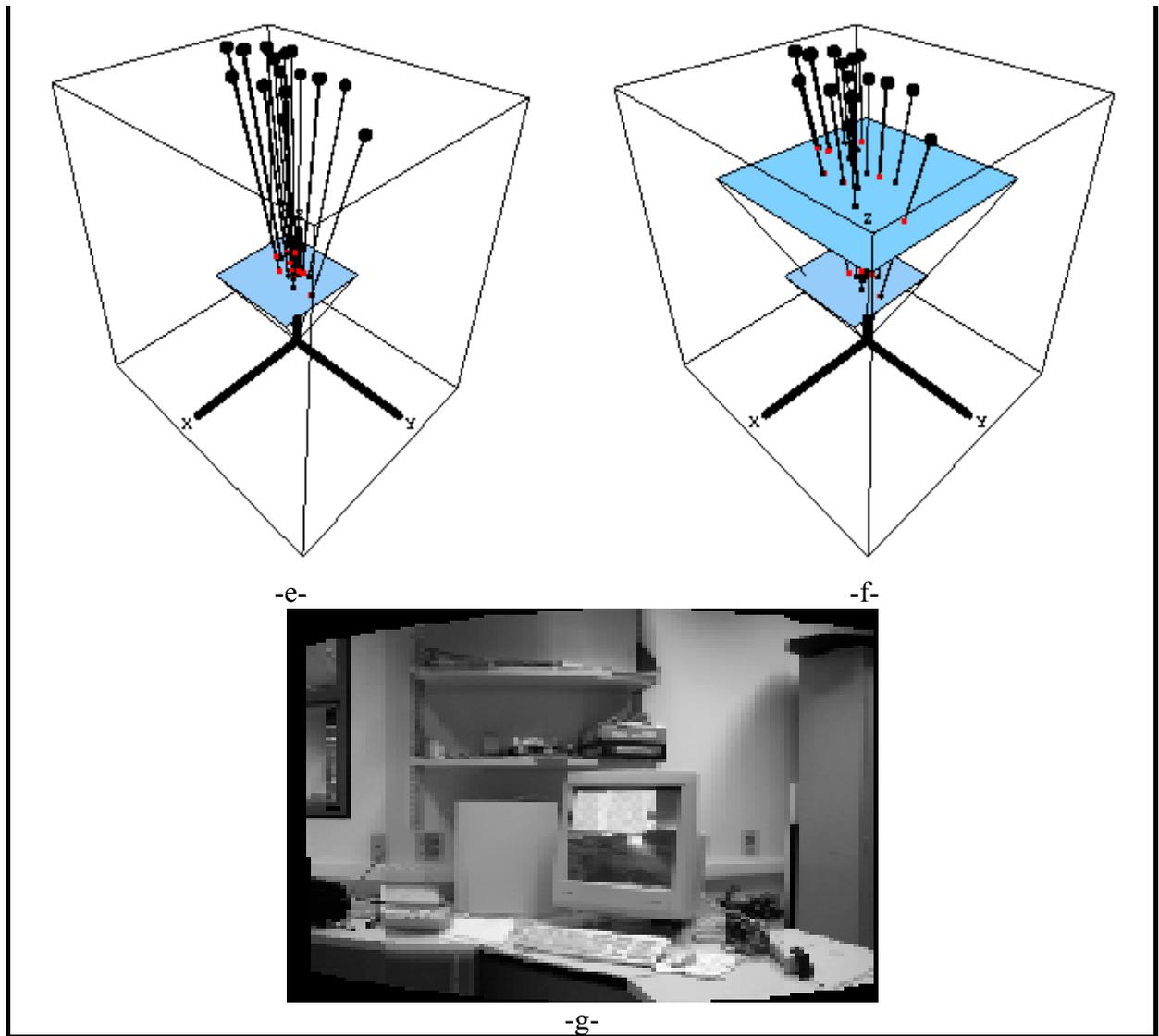
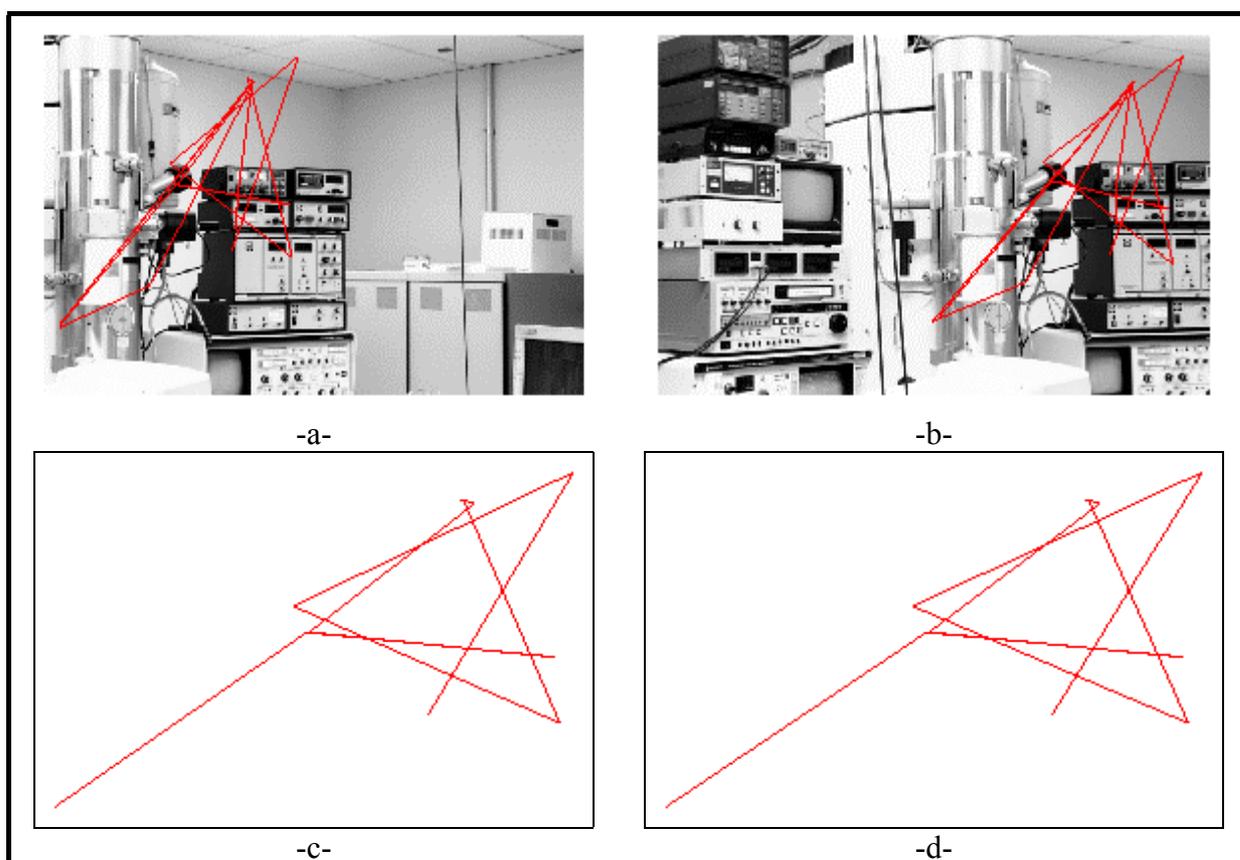


Figure 6.2: Exemple de construction de la mosaïque avec un seuil moyen

La figure 6.3 montre un exemple dont une des images est texturée partout figure 6.3-b et l'autre est à moitié texturée figure 6.3-a. Les points correspondants (reliés par des lignes) sont illustrés aux figures 6.3-c et 6.3-d qui montrent qu'il y a moins de triangles obtenus parce que nous avons moins de vraies paires de mise en correspondance. Les figures 6.3-e, 6.3-f et 6.3-g illustrent la calibration des caméras et montrent l'efficacité de la mise en correspondance puisque nous avons une ligne qui relie chaque point du premier plan (image A) à sa projection dans la scène en passant par sa correspondance se trouvant dans le deuxième plan (image B). La figure 6.3-h montre l'image mosaïque obtenue.



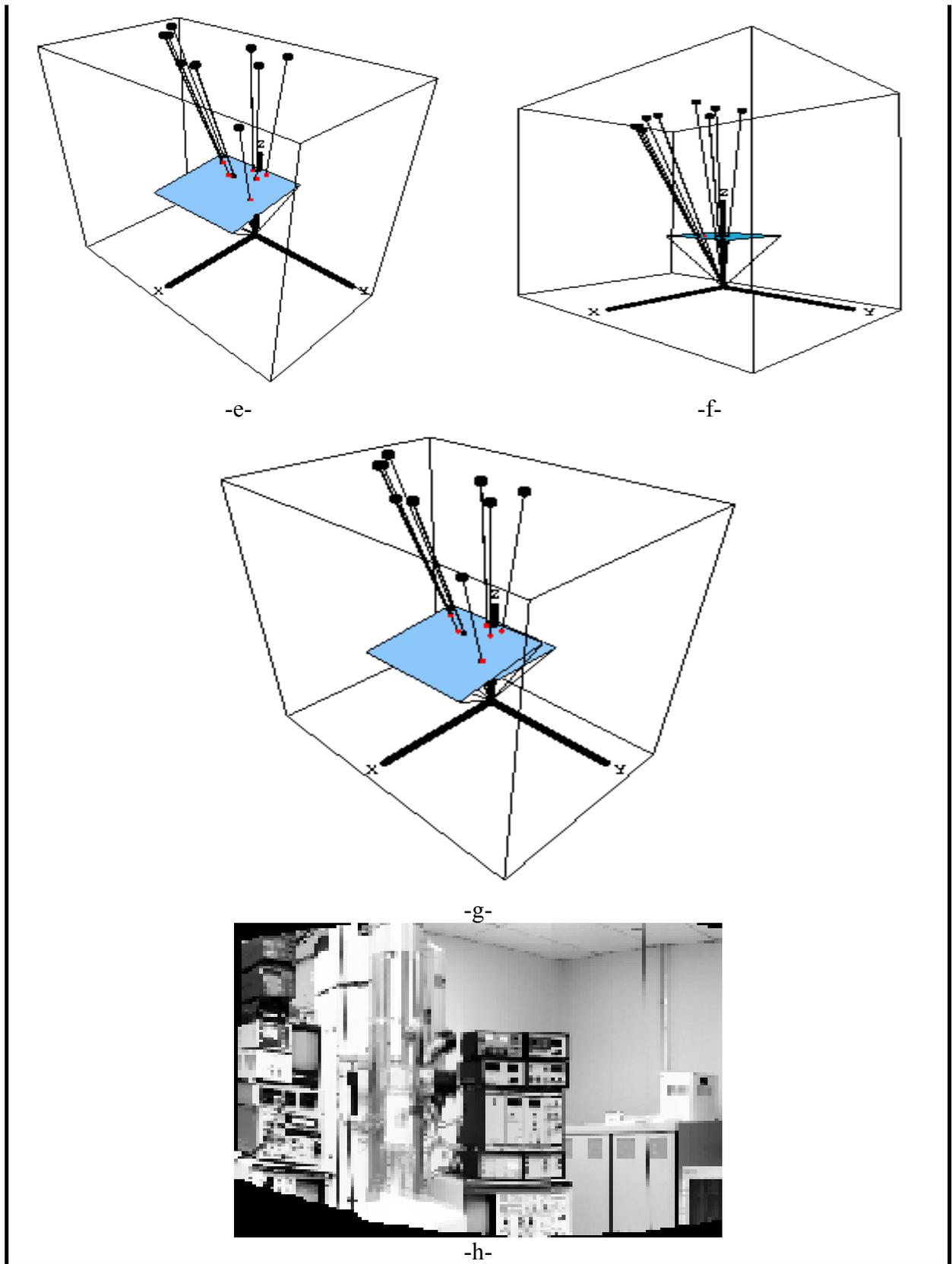
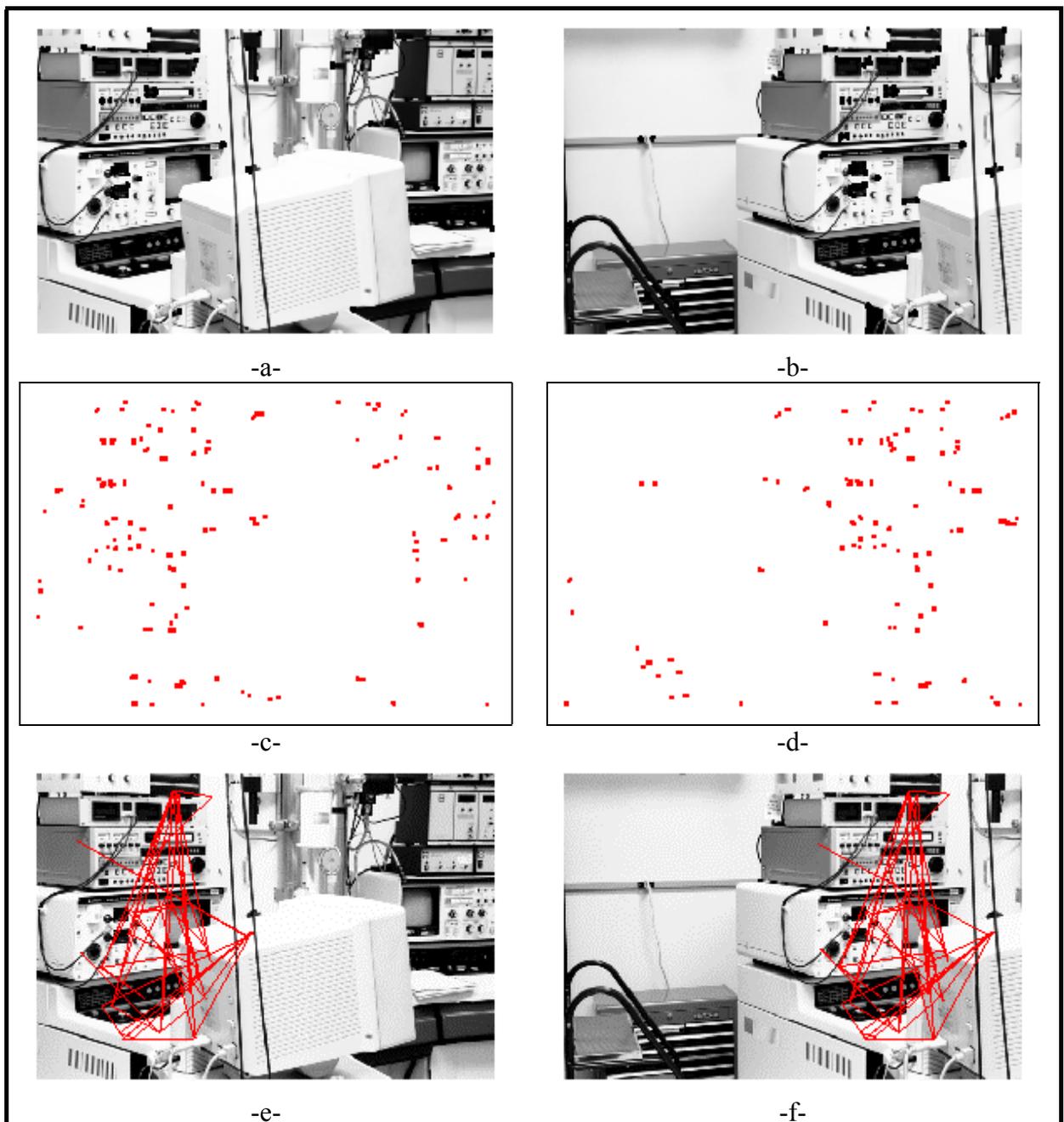
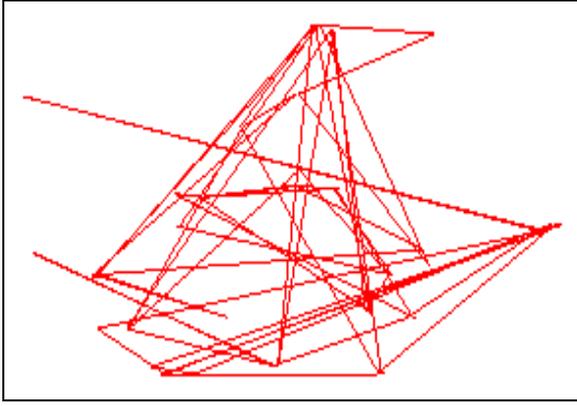


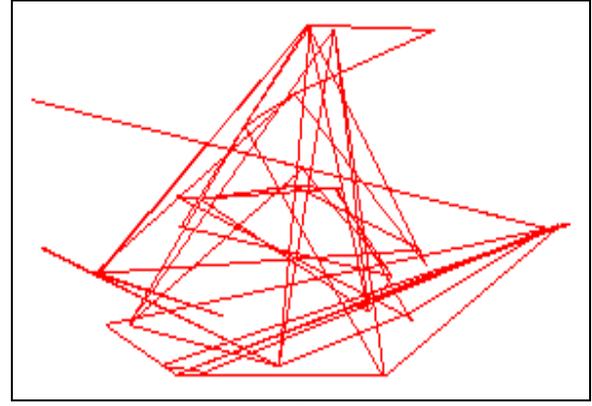
Figure 6.3: Exemple de construction de la mosaïque avec un seuil moyen

La figure 6.4 montre un exemple qui contient des textures très fortes dans la zone d'intersection de deux images (figures 6.4-a et 6.4-b). Les points correspondants (reliés par des lignes) sont illustrés aux figures 6.4-g et 6.4-h et montrent qu'il y a plus de triangles obtenus lorsque la texture est forte. Les figures 6.4-i, 6.4-j et 6.4-k montrent l'efficacité de la mise en correspondance puisque nous avons une ligne qui relie chaque point du premier plan (image A) à sa projection dans la scène en passant par sa correspondance se trouvant dans le deuxième plan (image B). La figure 6.4-l montre l'image mosaïque obtenue.

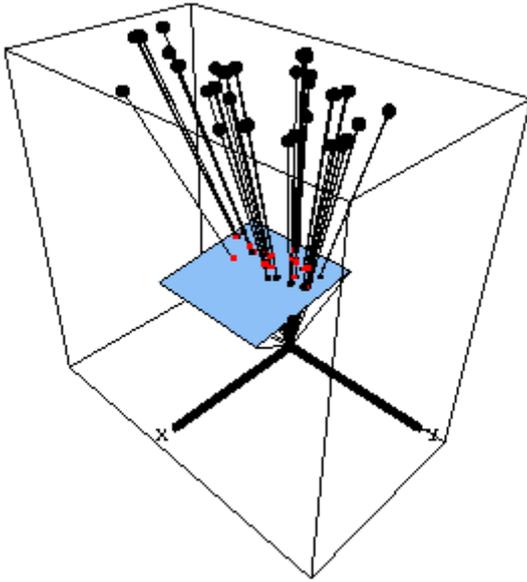




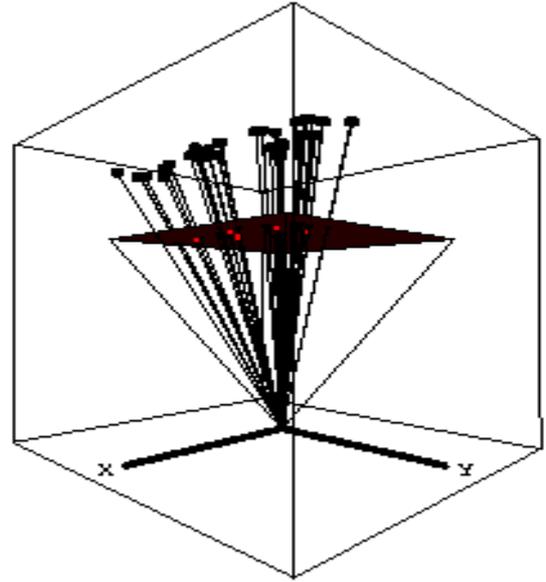
-g-



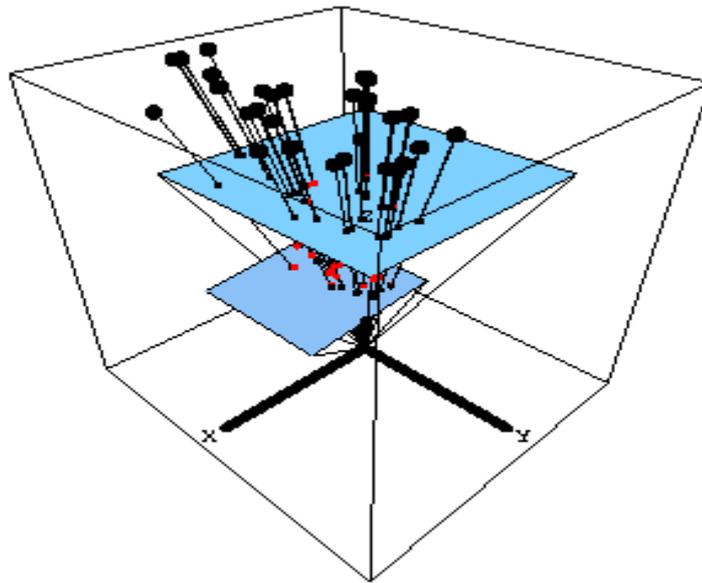
-h-



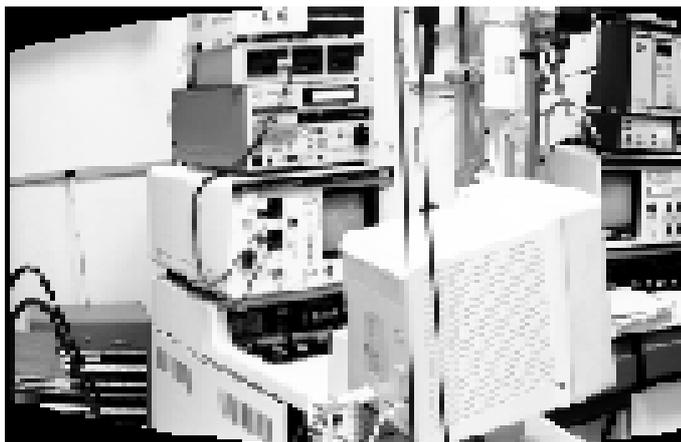
-i-



-j-



-k-

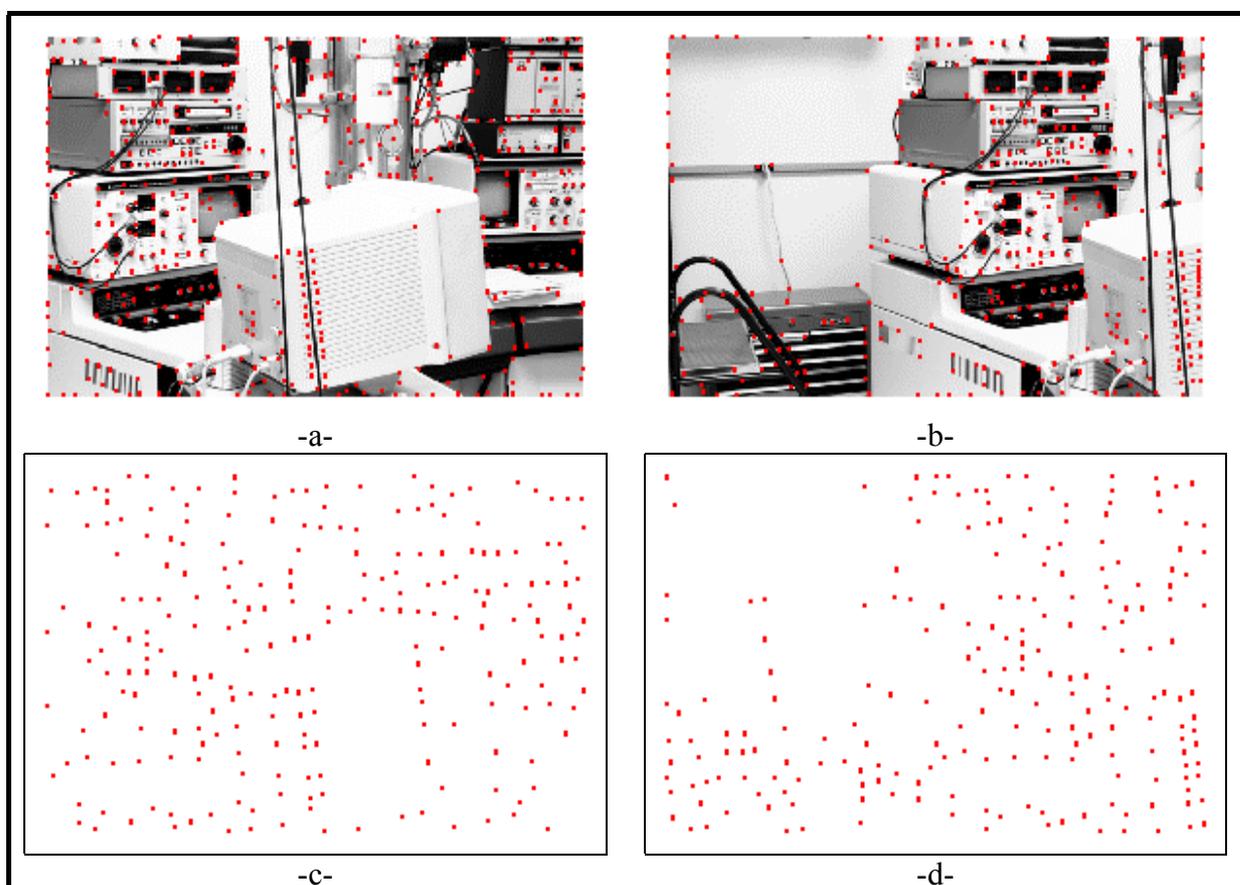


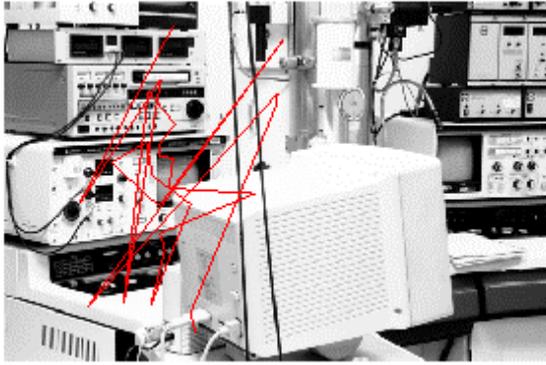
-|-

Figure 6.4: Exemple de construction de la mosaïque avec un seuil moyen

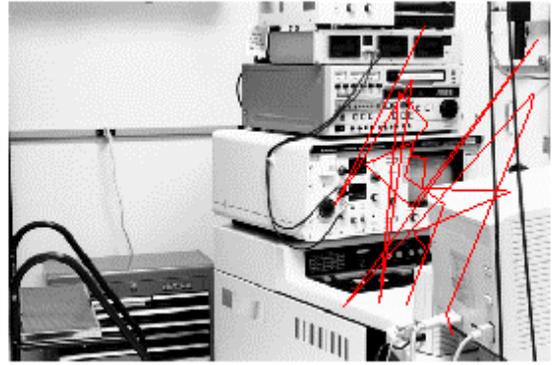
Exemple avec un seuil adaptatif

La figure 6.5 montre le même exemple que celui de la figure 6.4, mais en utilisant seuil adaptatif. Les figures 6.5-c et 6.5-d montrent l'uniformité des points d'intérêt obtenus. Cette propriété donne l'avantage du seuil adaptatif en comparant avec le seuil moyen. Les points correspondants (reliés par des lignes) sont illustrés aux figures 6.5-e, 6.5-f, 6.5-g et 6.5-h. On remarque qu'il y a moins de triangles en comparant avec la figure 6.4-c et 6.4-d puisque les points d'intérêt obtenus sont bien distribués dans les images ce qui donne moins des points d'intérêt dans la zone d'intersection de deux images et donc moins de triangles. Les figures 6.5-i, 6.5-j et 6.5-k montrent l'efficacité de la mise en correspondance puisque nous avons une ligne qui relie chaque point du premier plan (image A) à sa projection dans la scène en passant par sa correspondance se trouvant dans le deuxième plan (image B). La figure 6.5-l montre l'image mosaïque obtenue. La calibration obtenue par les deux méthodes (seuil adaptatif et seuil moyen) est équivalente. Pour mesurer l'impact exact de différentes méthodes sur la calibration il faudrait utiliser des images synthétiques, mais malheureusement l'image synthétique se prête mal à la détection de points d'intérêt à cause de leurs textures peu réaliste, nous préférons donc illustré la méthode sur des images réelles.

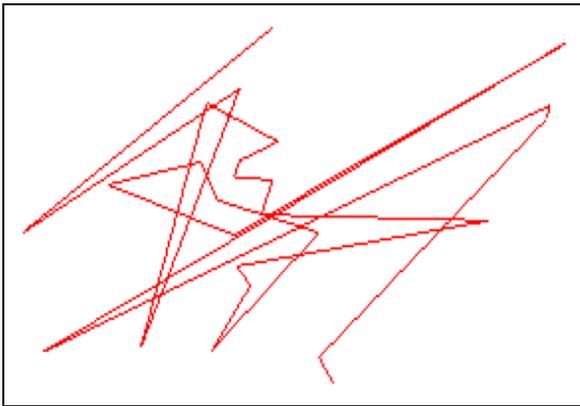




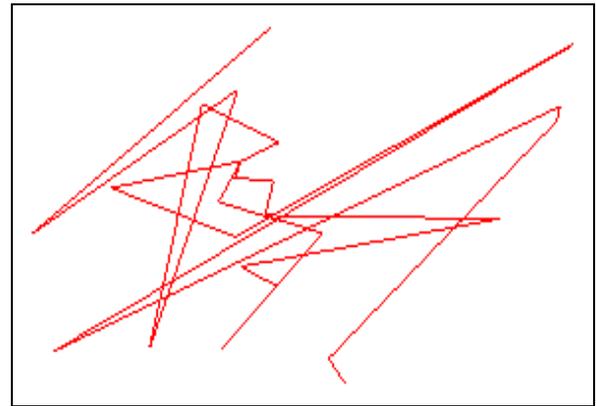
-e-



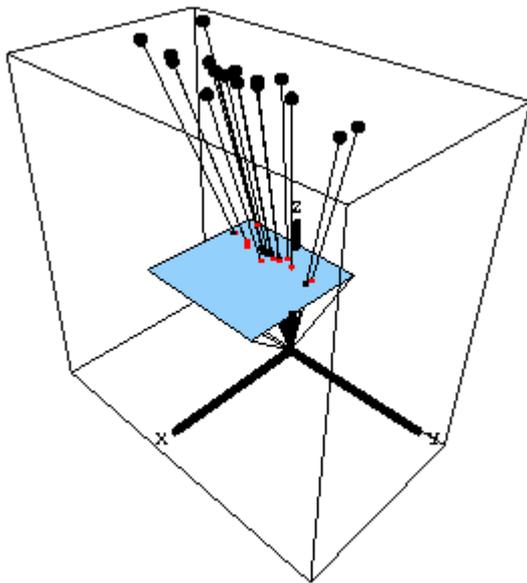
-f-



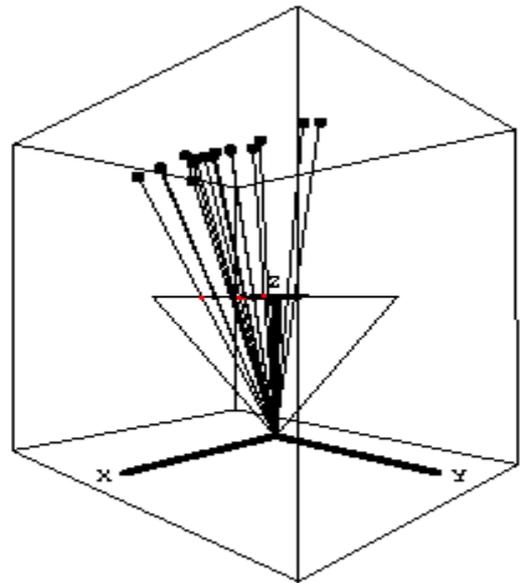
-g-



-h-



-i-



-j-

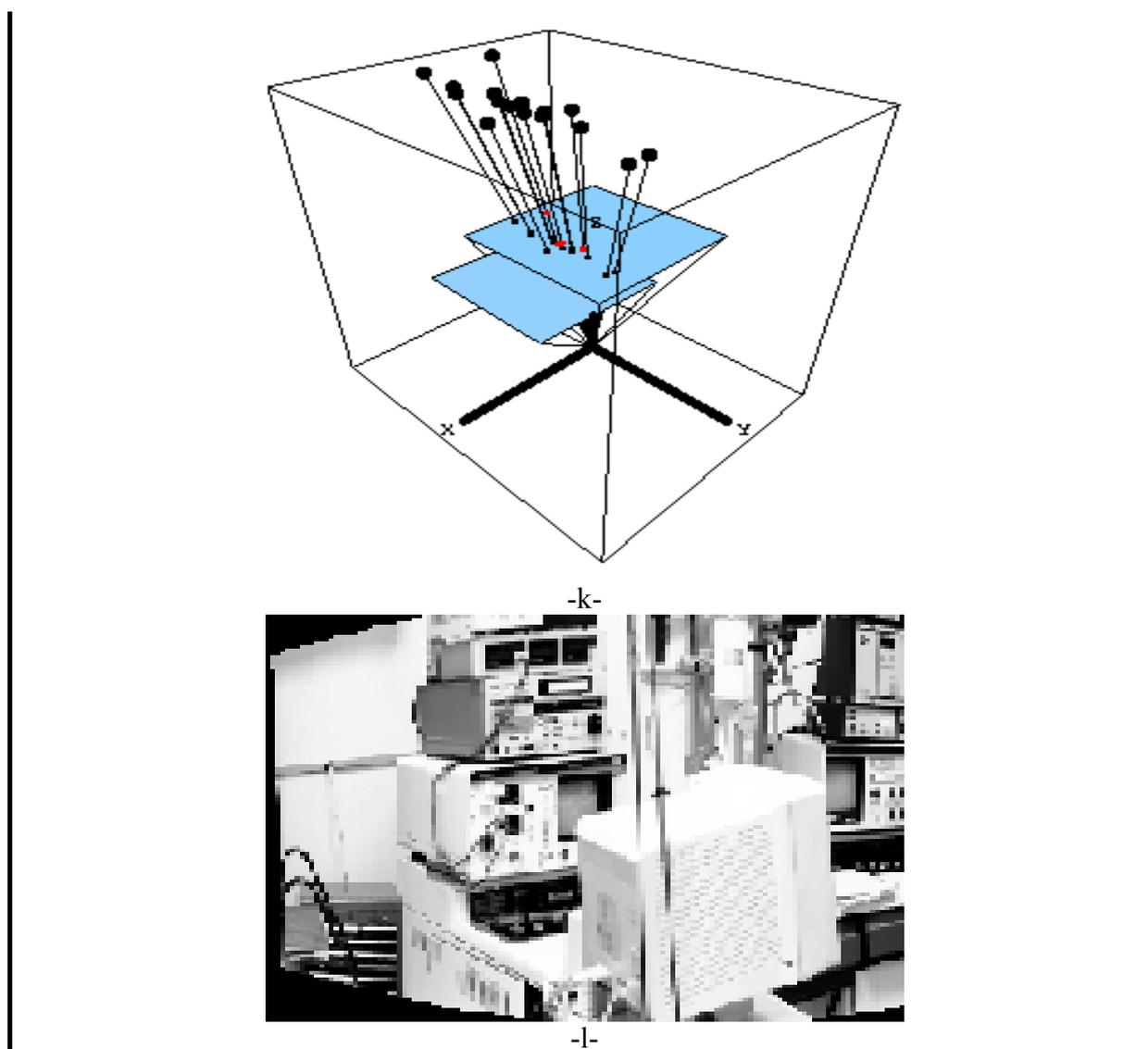


Figure 6.5: Exemple de la construction d'une mosaïque avec un seuil adaptatif

	Seuil	Points d'intérêt	V	Paires de mise en correspondance
Figure 6.1	Moyen	450	30	23
Figure 6.2	Moyen	265	53	48
Figure 6.3	Moyen	363	33	23
Figure 6.4	Moyen	361	55	45
Figure 6.5	Adaptatif	228	34	25

Tableau 6-1 :statique sur les résultats.

Chapitre 7:

7 Conclusion

La construction d'une mosaïque est un sujet important dans le domaine de la vision par ordinateur. Cette importance est vue clairement dans plusieurs situations où il est nécessaire d'avoir une image panoramique composée de plusieurs images distinctes. Une telle situation peut apparaître dans le cas où la taille d'un paysage dépasse le champ visuel de la caméra. Une mosaïque est construite en de quatre étapes que nous avons abordées dans notre mémoire. Ces étapes sont : la détection des points d'intérêt, la mise en correspondance, la calibration des caméras et la projection des images sur un plan cylindrique.

Les contributions principales de ce mémoire sont d'améliorer et automatiser la mise en correspondance par le recours à une méthode de triangulation et à un meilleur modèle de seuillage des points d'intérêt.

Dans le chapitre 3, nous avons traité la première étape qui consiste à détecter les points d'intérêt dans les différentes images. Cette détection est faite en utilisant le détecteur de Harris [Har88]. Nous avons utilisé plusieurs seuils pour améliorer la qualité de la distribution des points d'intérêt et ainsi obtenir le nombre de points nécessaires pour l'étape suivante de la mise en correspondance. De plus, nous proposons une méthode de seuillage « adaptatif » qui permet de générer un ensemble de points d'intérêt plus uniformément distribués dans l'image.

Après avoir détecté les points d'intérêt, la deuxième étape étudiée dans le chapitre 4 sert à utiliser ces points pour faire la mise en correspondance. Cette partie est cruciale dans le domaine de la vision par ordinateur. Une détection erronée d'une paire de points parmi plusieurs paires valides conduit à un échec complet dans la construction de la mosaïque. Cette caractéristique nous force à concevoir une méthode plus précise pour assurer la fiabilité des paires mises en correspondants. Pour atteindre ce but, nous avons élaboré une méthode efficace basée sur la comparaison de triangles, afin de détecter les liens parmi les points d'intérêt et assurer la véracité des paires obtenues.

La continuation de cette étape amène à la calibration des caméras. Dans cette partie, nous avons implémenté l'algorithme de la calibration affine pour calculer la matrice de transformation qui est ensuite utilisée dans l'étape finale. Cette dernière consiste à projeter les images en les utilisant sur la projection cylindrique.

Nous constatons que les étapes de la construction d'une mosaïque sont interdépendantes, dans le sens que la réussite ou l'échec d'une étape induit respectivement une amélioration ou une dégradation dans l'étape suivante. C'est pour cette raison que notre travail vise à accomplir une construction correcte d'une mosaïque en réduisant au maximum les erreurs de chaque étape.

Rappelons que l'objectif initial de notre projet était de réaliser une construction automatique et correcte d'une mosaïque à partir d'un ensemble d'images données. Nous avons atteint ce but dans ce mémoire en implantant différentes techniques pour le choix des points d'intérêt ainsi que pour la mise en correspondance. Les résultats que nous avons présentés dans le chapitre 6 montrent la qualité des mosaïques obtenues à partir de plusieurs tests empiriques de notre système. Il serait souhaitable d'ajouter assez tests de mesures quantitatives basées sur l'utilisation d'images synthétiques réalistes. De même il serait possible d'évaluer la robustesse de notre méthode en modifiant les caractéristiques de textures, par exemple en ajoutant de bruit.

Finalement, une direction pour des travaux futurs serait d'intégrer les étapes de choix de points de référence et la comparaison par triangulation en une seule étape, ce qui augmenterait la fiabilité.

Bibliographie

- [Asa86] H. Asada et M. Brady. The curvature primal sketch. IEEE Trans. On PAMI, 8(1): 2-14 , 1986.
- [Bar80] S. T. Barnard et W. B. Thompson. Disparity analysis of images. IEEE Trans. On PAMI 2(4):333-340, 1980.
- [Bau96] C. Bauchage and C. Schmid. Evaluation of keypoint detectors. Technical report, INIRIA, 1996.
- [Bea78] P.R. Beaudet. Rotationally invariant image operators. In proceedings of the quatrième International Joint Conference on pattern Recognition, pages 579-583, 1978.
- [Bea94] P. Beardsley, A. Zisserman et D. Murray. Sequential update of projective and affine structure from motion. Technical report, 2012/94, University of Oxford, 1994.
- [Bra87] M. Brady-Seeds of perception. in proceedings of the 3rd Alvey vision conference, pages 259-265,1987.
- [Bra94] P. Brand et R. Mohr. Accuracy in image measure. in proceedings of the SPIE conference on videometrics, volume 2350, pages 218-228, 1994.
- [Che84] J. K. Cheng and T.S. Huang. Image registration by matching relational structures. Pattern Recognition, 17(1): 149-159, 1984.
- [Der93a] R. Deriche et T. Blaszk. Recovering and characterizing image features using an efficient model based approach. In proceedings of the conference on computer vision and pattern recognition, pages 530-535, 1993.
- [Der93b] R. Deriche et G. Giraudon. A computational approach for corner and vertex detection. International Journal of Computer Vision, 10(2): 101-124, 1993.
- [Dre82] L. Dreschler et H. Nagel. Volumetric model and 3D trajectory of a moving car derived from monocular TV frame sequences of a street scene. Computer Graphics and Image Processing, 20: 99-228, 1982.
- [Esp00] F. Espiau. Extraction de points robustes dans des images naturelles complexes, INRIA, JJCR13 2000 Rennes.
- [For87] W. Forstner et Gulch. A fast operator of detection and precise location of distinct points, corners and circular features. In intercommission conference on fast processing of photogrammetric data, pages 281-305, 1987.
- [For94] W. Forstner. A framework for low level feature extraction. In j-o Eklundh editor, lecture notes in computer vision, computer vision- ECCV'94 volume 800, pages 383-394, 1994.
- [Har80] R.M. Haralick and G.L. Elliott. Increasing tree search efficiency for constraint satisfaction problems. Artif. Intell., 14 :263-313, 1980.
- [Har88] C. Harris et M. Stephens. A combined corner and edge detector. In proceedings of the 4th Alvey vision conference, pages 147-151, 1988.
- [Hei92] F. Heitger, L. Rosenthaler, R. von der Heydt, E. Peterhans et O. Kuebler,. Simulation of neural contour mechanism: from simple to end-stopped cells. Vision research, 32(5), pages, 963-981, 1992.

- [Hor90] R. Horaud, T. Skordas. Stereo correspondence through feature grouping and maximal cliques. IEEE transactions on PAMI, pages 374-384, 1990.
- [Hum83] R. A. Hummel and S. W. Zucker. On the foundation of relaxation labeling process. IEEE Trans. PAMI, 5(3): 267-286, 1983.
- [Kit82] L. Kitchen et A. Rosenfeld. Gray-level corner detection. Pattern Recognition letters, pages 95-102, 1982.
- [Med87] G. Medioni et Y. Yasumoto. Corner detection and curve representation using cubic B-splines. Computer vision, graphics and image processing, 39(1): 267-287, 1989.
- [Mok86] F. Mokhtarian et A. Mackworth. Scale-based description and recognition of planar curves and two dimensional shapes. IEEE transactions on pattern analysis and machine intelligence, 8(1) : 34-43, 1986.
- [Mok98] F. Mokhtarian and R. Suomela. Robust image corner detection through curvature scale space. IEEE Trans on Pattern Analysis and Machine Intelligence, 20(12):1376-1381, 1998.
- [Mor79] H. Moravec. Visual mapping by a robot rover. In proceedings of the 6th international joint conference on artificial intelligence, pages 598-600, 1979.
- [Mor81] H. Moravec. Rover obstacle avoidance. In proceedings of the 7th international joint conference on artificial intelligence, pages 785-790 1981.
- [Oga84] H. Ogawa. Labeled point pattern matching by fuzzy relaxation. Pattern Recog., 17(5) : 569-573, 1984.
- [Rad84] B. Rading. Image sequence analysis using relational structures. Pattern Recog., 17(1) :161-167, 1984.
- [Roh92] K. Roher. Recognizing corners by fitting parametric. International journal of computer vision, 9(3): 213-290, 1992.
- [Ros92] L. Rosenthaler, F. Heitger, O. Kuebler et R. von der Heydt. Detection of general edges and keypoints. In proceedings of the 2nd European conference on computer vision, pages 78-86, 1994.
- [Rou98] B. Rousso, S. Peleg, I. Finci et A. Rav-Acha. Universal Mosaicing using pipe projection ICC V'98, Bombay 1998 pp. 945-952.
- [Smi97] S. Smith et. Brady. A new approach to low-level image processing. International Journal of Computer Vision, 23(1): 45-78, 1997.
- [Zha95] Z. Zhang, R. Deriche, O. Faugeras et Q. Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. Artificial intelligence, pages 87-119, 1995.
- [Zog97] Imad Zoghlami, Rachid Deriche, Olivier Faugeras "Using geometric corners to build a 2D mosaic from a set of images" In proceedings of the international conference on computer vision and pattern recognition, Puerto Rico, 1997.