

République Algérienne Démocratique et Populaire
Ministère de l'Enseignement Supérieur et de la Recherche Scientifique
UNIVERSITE MOULOUD MAMMERI DE TIZI-OUZOU



FACULTE DU GENIE ELECTRIQUE ET D'INFORMATIQUE
DEPARTEMENT D'ELECTRONIQUE

Mémoire De Fin D'études
MASTER ACADEMIQUE
Domaine: Sciences et Technologies
Filière: Génie Electrique
Spécialité: Electronique Biomédicale

Présenté par :

BOUTRAA Kahina

Encadré par :

M^{elle} N.DRIF

Thème

Appariement entre deux images du visage

Soutenue publiquement le : 30 septembre 2015

Remerciement

Je remercie Allah le tout puissant de nous avoir donné le courage et la volonté de parvenir à la fin de notre parcours universitaire

Je tiens à remercier énormément et infiniment ma promotrice M^{elle} DRIF Naima pour avoir accepté de nous encadrer et diriger ce travail

Je tiens à remercier les membres de jury qui ont acceptés de juger et examiner notre travail.

Je tiens à remercier tous les enseignants du département d'électronique qui ont contribué à notre formation

Je tiens à remercier tous ce qui a contribué de près et de loin pour la réalisation de ce travail.

Dédicaces

Je dédie ce modeste travail aux personnes les plus chère au monde : mes parents qui mon toujours aidé et encourager dans mon parcours universitaire, sans oublier leurs sacrifices et amour.

À tous mes chers frères et mes sœurs.

A mon adorable mari Ali, pour son encouragement, son amour, son écoute et sa patience.

A ma cher et unique tante : Louiza

A mes très chères grands-mères.

A toute personne qui était derrière ma réussite et ma soutenait par tous les moyens.

Table des figures

Figure I_1 Exemple d'images comportant un seul visage.	09
Figure I_2 Exemple d'images comportant plusieurs visages.	09
Figure I_3 Exemple de caractéristiques pseudo-Haar.	12
Figure I_4 Architecture de cascade.	16
Figure II_1 Types de points d'intérêt : coins, jonction en Y.	18
Figure II_2 Classification chronologique des détecteurs des points qui existent.	20
Figure II_3 Détecteur Dog.	24
Figure II_4 Détecteur Harris-laplace.	25
Figure II_5 Le maximum de la fonction $F(I, X, s_n)$ dans l'espace d'échelle.	28
Figure II_6 Points détectés sur différents niveaux de résolution.	30
Figure III_1 Principe de la corrélation	37
Figure_ III_2 Exemple de mesures de corrélation	38
Figure IV_1 Images test d'un seul visage.	42
FigureIV_2 Images test de plusieurs visages.	43
FigureIV_3 Résultat d'application de la méthode sur des images d'un seul visage.	44
FigureIV_4 Résultat d'application de la méthode sur des images de plusieurs visages.	45
FigureIV_5 Détection des points d'intérêt sur le visage.	46
FigureIV_8 L'appariement entre deux images.	47
FigureIV_9 L'appariement entre deux images.	48
FigureIV_7 L'appariement entre deux visages.	49

Table des matières

Introduction générale.....	1
----------------------------	---

Chapitre I

Les différentes approches de la détection du visage

Introduction	3
La biométrie pour la reconnaissance de visage	3
La Détection de visage	4
Les Approches et les méthodes de détection du visage	4
Approches basées sur les connaissances acquises	4
Approches basées sur le Template-matching	5
Approches basées sur l'apparence	6
Approches basées sur des caractéristiques invariantes	7
Le principe de la détection de visages.....	8
Le détecteur de Viola et Jones	11
Fonctionnement des détecteurs du Viola et Jones	11
Apprentissage du Viola et Jones	12
Détection du Viola et Jones.....	13
Limites et Applications du Viola et Jones	13
Algorithme AdaBoost de Viola et Jones pour la détection de visages.....	13
• Algorithme Adaboost	14
• Détection en Cascade	15
Conclusion.....	17

Chapitre II

Détection des points d'intérêt

Introduction	18
Définition des points d'intérêt	18
Différentes approches.....	19
Détecteurs	19
Détecteur de Beaudet	20
Détecteur de Moravec	21
Détecteur de Dreschler et Nagel	21
Harris et Stephens	22
Détecteur de SUSAN	22
Détecteur DoG	23
Détecteur Harris-Laplace	24
Détecteur MSER	25
Détecteurs Harris-affine et Hessian-affine	25
Choix d'un algorithme de la détection des points d'intérêts	26
Points d'intérêt invariants à l'échelle	28
Conclusion.....	30

Chapitre III

Appariement d'image

Introduction	31
Caractérisation des points d'intérêts	31
Descripteur basé sur les moments	31
Les moments de Hu	32
Les moments de Zernike	32
Descripteur basé sur les transformées intégrales	33
Descripteur basé sur les histogrammes	33
Histogramme d'intensité lumineuse	34
Histogramme de gradients orientés	34

Appariement d'image	34
Principe de l'appariement	35
Discussion sur les méthodes d'appariement	36
L'appariement par corrélation	37
Difficultés liées à l'utilisation de la corrélation	39
Propriétés du la corrélation	40
Avantages et inconvénients	40
Cadre d'évaluation	41
Conclusion.....	41

Chapitre IV

Test et résultat

Introduction	42
Les images de tests	42
Détection du visage	43
Cas d'images à un seul visage	43
Cas d'images a plusieurs visage	44
Détection des points intérêt sur le visage	45
Appariement	46
L'appariement entre deux images	47
L'appariement entre deux visages	49
Conclusion.....	50
Conclusion générale	51
Bibliographie	

Le système visuel humain permet de détecter des objets tridimensionnels et d'estimer leur mouvement dans l'espace en une fraction de seconde.

Plusieurs travaux ont montré que le cerveau est capable de détecter des visages présentant un intérêt particulier en 200 ms dans le but de focaliser l'attention sur ces régions. Les primitives détectées sont par exemple des coins ou des points d'intérêts. C'est à partir de ces primitives et des informations contenues dans leur voisinage que, selon certains spécialistes, le cerveau détecte les visages et en estime le mouvement. Ceci représente les fondements mêmes du concept de description locale.

Un système de la reconnaissance faciale est une application logicielle basé sur la reconnaissance d'une personne grâce à l'image de son visage de manière automatique.

Nous pouvons citer la reconnaissance 3D, cette méthode est considérée comme une amélioration de la reconnaissance 2D. En effet, elle crée un modèle 3D depuis plusieurs photos 2D effectuées successivement sur un visage, cela permettant d'avoir différents points de vue de la personne à reconnaître afin de créer le modèle 3D.

Afin d'arriver à une reconstruction 3D du visage on a besoin de deux images stéréoscopiques sur lesquelles on fait d'abord la détection des points d'intérêts sur les deux images et ensuite on fait l'appariement entre celles-ci, ces deux étapes la qui sont très importante avant toute reconstruction 3D.

Dans ce mémoire le travail qu'on va réaliser est l'appariement entre deux images du visage pour ce faire le mémoire sera organiser comme suit :

Dans le premier chapitre nous présenterons les différentes approches et méthodes de détection du visage ainsi que la méthode de Viola et Jones.

Dans le deuxième chapitre nous présenterons la détection des points d'intérêts et les différents détecteurs qui existent, ensuite on présentera le détecteur de HARRIS.

Dans le troisième chapitre, nous parlerons sur la caractérisation des points d'intérêts ainsi que l'appariement entre deux images a comparés. On a choisi de travailler avec la méthode de corrélation pour l'appariement entre deux images du visage.

Dans le quatrième chapitre, Nous présentons les résultats expérimentaux obtenus par les méthodes de la détection du visage, détection de points d'intérêts et l'appariement.

Enfin, nous terminerons ce mémoire par une conclusion générale.

I_1 Introduction :

La détection de visage est un domaine de la vision par ordinateur consistant à détecter un visage humain dans une image numérique. C'est un cas spécifique de détection d'objet, où l'on cherche à détecter la présence et la localisation précise d'un ou plusieurs visages dans une image. C'est l'un des domaines de la vision par ordinateur parmi les plus étudiés, avec de très nombreuses publications et de conférences spécialisées. La forte activité de recherche en détection de visage a également permis de faire émerger des méthodes génériques de détection d'objet. Dans ce chapitre nous présenterons quelques notions sur la biométrie ainsi que quelques méthodes de détection du visage.

I_2 La biométrie pour la reconnaissance de visage:

La biométrie est l'étude quantitative des êtres vivants, plus précisément dans notre contexte: c'est la reconnaissance et l'identification des individus en utilisant des informations étroitement liés à leurs caractéristiques. C'est la science qu'on utilise pour différencier des personnes entre elles grâce à leur biologie, automatiquement reconnaissable et vérifiable. Certaines biométries utilisent des caractéristiques comportementales des individus comme l'écriture ou le signal de la parole. D'autres utilisent des caractéristiques physiques comme l'iris, les empreintes digitales ou la forme du visage.

La reconnaissance de visage est un domaine de la vision par ordinateur consistant à reconnaître automatiquement une personne à partir d'une image de son visage. C'est un sujet particulièrement étudié en vision par ordinateur.

L'efficacité des systèmes biométriques basés sur l'authentification de visage dépend essentiellement de la méthode utilisée pour localiser le visage dans l'image. Dans la littérature scientifique, le problème de localisation de visages est aussi désigné par la terminologie de la détection du visage. Plusieurs travaux de recherches ont été effectués dans ce domaine. Ils ont donné lieu au développement d'une multitude de techniques allant de la simple détection du visage, à la localisation précise des régions caractéristiques du visage, tels que les yeux, le nez, les narines, les sourcils, la bouche, les lèvres, les oreilles, etc.

Dans ce qui suit nous présenterons quelques-unes de ces méthodes.

I_3 la Détection de visage :

La détection de visage cherche à détecter la présence et la localisation précise d'un ou plusieurs visages dans une image numérique. C'est un sujet difficile, notamment dû à la grande variabilité d'apparence des visages dans des conditions non contraintes:

- Variabilité intrinsèque des visages humains (couleur, taille, forme),
- Présence ou absence de caractéristiques particulières (cheveux, moustache, barbe, lunettes...),
- Expressions faciales modifiant la géométrie du visage,
- Occultation par d'autres objets ou d'autres visages,
- Orientation et pose (de face, de profil),
- Conditions d'illumination et qualité de l'image.

I_3_1 Les Approches et les méthodes de détection du visage :

Des nombreuses méthodes de détection de visages sont apparues dans les deux dernières décennies et qui se diffèrent, aussi bien par les approches qu'elles emploient, que par les techniques d'apprentissage qu'elles utilisent. Yang et al. Classifient ces techniques en quatre classes :

- techniques descriptives basées sur la connaissance,
- techniques basées sur l'extraction de paramètres caractéristiques invariants,
- techniques basées sur la superposition de caractéristiques,
- techniques basées sur l'apparence.

Les techniques, basées sur l'apparence, utilisent l'analyse statistique et l'apprentissage automatique pour construire des machines capables de séparer les visages des non-visages. Les réseaux de neurones, les machines à vecteurs de support (SVM : Support Vector Machine).

I_3_1.a) Approches basées sur les connaissances acquises :

Cette méthodologie s'intéresse aux parties caractéristiques du visage comme le nez, la bouche et les yeux. Elle est basée sur la définition de règles strictes à partir des rapports entre les caractéristiques faciales. Ces méthodes sont conçues principalement pour la localisation de

visage. Dans [1], Kotropoulous et Pitas utilisent une méthode à base de règles. Tout d'abord, les caractéristiques du visage sont localisées à l'aide de la méthode de projection proposée par Kanade[2] pour détecter les contours d'un visage. Soit $I(x,y)$ l'intensité de la luminance du pixel (x,y) de l'image $m \times n$, les projections horizontale et verticale de cette image sont définies par :

$$HI(x)=\sum_{y=1}^n I(x, y) \text{ et } VI(x)=\sum_{x=1}^m I(x, y) .$$

Le profil horizontal de l'image originale est calculé en premier. Les deux minima locaux sont déterminés, ils correspondent aux bords gauche et droit du visage. Ensuite, le profil vertical est à son tour calculé. Les minima locaux de ce profil vertical correspondent aux positions de la bouche, du nez et des yeux. L'inconvénient de cette méthode est qu'elle n'arrive pas à détecter le visage lorsque ce dernier se trouve sur un arrière-plan complexe.

Yang et Huang [3] quant à eux, ont étudié les évolutions des caractéristiques du

visage en fonction de la résolution. Quand la résolution de l'image d'un visage est réduite progressivement, par sous-échantillonnage ou par moyenne, les traits macroscopiques du

visage disparaissent. Ainsi, pour une résolution faible, la région du visage devient uniforme.

Yang et Huang se sont basés sur cette observation pour proposer une méthode hiérarchique de détection de visages. En commençant par les images à faible résolution, un ensemble de candidats de visage est déterminé à l'aide d'un ensemble de règles permettant de rechercher les régions uniformes dans une image. Les candidats de visage sont ensuite vérifiés en cherchant l'existence de traits faciaux proéminents grâce au calcul des minima locaux à des résolutions supérieures. Une caractéristique intéressante de cette technique « descendante » de recherche de zone d'intérêt (informations globales vers des informations plus détaillées) est de réduire le temps de calcul nécessaire par l'utilisation d'images sous-échantillonnées. Malheureusement, cette technique occasionne de nombreuses fausses détections et un taux faible de détection.

I_3_1.b) Approches basées sur le Template-matching :

Les Templates peuvent être définis soit manuellement, soit paramétrés à l'aide de fonctions. L'idée est de calculer la corrélation entre l'image candidate et le Template. Ces méthodes rencontrent encore quelques problèmes de robustesse liés aux variations de lumière, d'échelle, etc. Sinha [4] [5] utilise un ensemble d'invariants décrivant le modèle du visage.

Afin de déterminer les invariants aux changements de luminosité permettant de caractériser les différentes parties du visage (telles que les yeux, les joues, et le front); cet algorithme calcule ainsi les rapports de luminance entre les régions du visage et retient les directions de ces rapports (par exemple, la région 1 est-elle plus claire ou plus sombre que la région 2). Une relation est vérifiée si le rapport entre les deux régions qui lui correspond dépasse un seuil. Le visage est localisé si le nombre de relations essentielles et de confirmation dépasse lui aussi un seuil.

I_3_1.c) Approches basées sur l'apparence :

Ces approches appliquent généralement des techniques d'apprentissage automatique. Ainsi, les modèles sont appris à partir d'un ensemble d'images représentatives de la variabilité de l'aspect facial. Ces modèles sont alors employées pour la détection. L'idée principale de ces méthodes est de considérer que le problème de la détection de visage est un problème de classification (visage, non-visage). Une des approches les plus connues de détection de visage est l'Eigenface [6], elle consiste à projeter l'image dans un espace et à calculer la distance euclidienne entre l'image et sa projection. En effet, en codant l'image dans un espace, on dégrade l'information contenue dans l'image, puis on calcule la perte d'information entre l'image et sa projection. Si cette perte d'information est grande, l'image n'est pas correctement représentée dans l'espace : elle ne contient pas de visage. Cette méthode donne des résultats assez encourageants, mais le temps de calcul est très important.

Dans Rowley et al. [7], les auteurs proposent un système de détection de visage basé sur la classification par des réseaux de neurones. Leur technique est divisée en deux étapes: la localisation des visages en utilisant un réseau de neurones et la vérification des résultats obtenus. Les auteurs ont construit un réseau de neurones qui, à partir d'une image prétraitée de 20x20 pixels, indique s'il s'agit d'un visage ou non. Le prétraitement consiste à égaliser l'histogramme. L'image est balayée avec des fenêtres de 20x20. Pour détecter les visages de différentes tailles, une analyse multi-résolutions est effectuée. L'extension a aussi été proposée afin de déterminer un réseau de neurones indiquant le degré de rotation d'un visage. Ainsi, le système est capable de détecter des visages ayant subi des rotations dans le plan et de différentes échelles. L'un des avantages des réseaux de neurones est leur robustesse au bruit. Malheureusement, les réseaux de neurones, sont souvent difficiles à construire. Leur structure (nombre de couches cachées pour les perceptrons par exemple) influe beaucoup sur les résultats et il n'existe pas de méthode pour déterminer automatiquement cette structure. La

phase d'apprentissage est difficile à mener puisque les exemples doivent être correctement choisis en nombre et en configuration. La méthode de Féraud et al [8] est basée sur un type particulier de réseau de neurones qualifié de génératif contraint et entraîné sur les valeurs des images globales. Le réseau est un Perceptron Multicouches auto-associatif et entièrement connecté. Il est conçu pour permettre de mettre en œuvre l'Analyse Composantes Principales (ACP) non-linéaire. Cette approche fournit des taux intéressants de détection de visage.

Une autre approche performante qui repose sur les réseaux de neurones est celle de Garcia et al [9], elle est basée sur l'utilisation des réseaux de neurone convolutionnel multicouche. Le réseau, entraîné de manière supervisée depuis une base d'apprentissage de visage et non visage, est capable de dériver automatiquement des extracteurs de caractéristiques (produit de convolution) spécialisées.

Une méthode bien connue de détection d'objets complexes tels que les visages est l'utilisation de « classifieurs de Haar » montés en cascade (boostés) au moyen d'un algorithme AdaBoost. Cette méthode est implémentée nativement dans la bibliothèque OpenCV [10] et a été présentée initialement dans Viola et Jones [11]. Le principe de cette méthode est obtenir un algorithme complexe de classification, composé de classifieurs élémentaires qui éliminent au fur et à mesure les zones de l'image qui ne sont pas compatibles avec l'objet recherché. Ces classifieurs binaires reposent sur des primitives visuelles qui dérivent des fonctions de Haar.

I_3_1.d) Approches basées sur des caractéristiques invariantes :

Ces approches sont utilisées principalement pour la localisation de visage. Les algorithmes développés visent à trouver les caractéristiques structurales existantes même si la pose, le point de vue, ou la condition d'éclairage changent. Puis ils emploient ces caractéristiques invariables pour localiser les visages. Nous pouvons citer deux familles de méthodes appartenant à cette approche :

- Les méthodes basées sur la couleur de la peau qu'on détaillera dans la section suivante,
- Les méthodes basées sur les caractéristiques de visage; elles consistent à localiser les cinq caractéristiques (deux yeux, deux narines, et la jonction nez/lèvre) pour décrire un visage typique.

L'algorithme développé par De Silva et al. [12] est un exemple typique des méthodes basées sur les caractéristiques. Leur algorithme utilise tout d'abord une hypothèse sur la position du haut du visage ensuite l'algorithme de recherche parcourt le visage de haut en bas afin de trouver l'axe des yeux caractérisé par une augmentation soudaine de la densité de contours (mesurée par le rapport noir/blanc le long des plans horizontaux).

La longueur entre le haut du visage et le plan de l'œil est alors utilisée comme une longueur de référence pour construire un « template » facial flexible. Ce « template » couvrant des caractéristiques telles que les yeux et la bouche est initialisé à partir de l'image d'entrée. La forme initiale du « template » est obtenue en utilisant la longueur anthropométrique en respectant la longueur de référence. Le template flexible est alors ajusté par rapport aux positions finales des caractéristiques en utilisant un algorithme de réglage fin qui emploie une fonction de coût basée contour. Les auteurs rapportent un taux de réussite pour la détection de toutes les caractéristiques faciales dans le cas de visages quasi-frontaux sur un fond homogène. Bien que l'algorithme réussisse à détecter les caractéristiques d'ethnies différentes puisqu'il ne se base pas sur les informations de niveaux de gris et de couleur, il n'arrive pas cependant à détecter correctement ces caractéristiques si l'image du visage contient des lunettes ou bien si les cheveux couvrent le front.

Par ailleurs, Yow et Cipolla [13] ont développé une méthode basée sur l'extraction de points d'intérêts. Ces points sont ensuite utilisés pour réaliser une détection de contours qui sont à leur tour groupés en régions. Le regroupement des contours est basé sur un critère de proximité et un critère de similarité en orientation et en épaisseur. A partir de chaque région, l'algorithme définit alors un vecteur pour lequel il calcule la moyenne et la matrice de covariance par rapport aux différents vecteurs des modèles. Le critère d'appartenance à un élément du visage s'appuie sur la distance de Mahalanobis: Les différents candidats sont alors groupés en se basant sur un modèle de connaissance indiquant leur position relative. Chaque composant du visage est enfin analysé avec un réseau bayésien. L'intérêt de cette méthode est qu'elle peut détecter des visages dans diverses poses. On a un taux de réussite de 85%.

I_3_2 Le principe de la détection de visages :

La détection automatique de visages prend son importance de point de vue qu'elle est à la base de la reconnaissance de visages dans une image ou dans une séquence vidéo.

Le concept de base de la détection de visages serait celui des K-PPV(K-Plus Proches Voisins), qui consiste à parcourir l'image avec une fenêtre (Figure I_1, Figure I_2), puis à comparer chaque imagerie extraite avec une série de visages types et de définir un Visage comme étant tout résultat dont la distance à l'une des images de la base soit suffisamment faible.

Partons de ce constat, il est nécessaire que la détection de la présence ou non d'un visage dans une image soit basée sur des éléments stables et relativement descriptifs du visage humain et qui permettent ensuite de le reconnaître. Parmi ces éléments, on peut citer la forme du visage, la couleur de la peau, le contour des yeux, la forme du nez ou de la bouche...



Figure I_1: Exemple d'images comportant un seul visage.



Figure I_2: Exemple d'images comportant plusieurs visages.

Mais en considérant des tailles, des orientations, des rotations et des éclairages différents il faudrait pouvoir comparer chaque imagette extraite à des centaines de références. Si l'on rajoute les expressions faciales, la détection de visages devient un problème difficile à traiter et c'est pour plusieurs raisons:

- Bien que la plupart des visages soient structurellement semblables avec des caractères morphologiques communs (yeux, bouche, nez,...) placés selon une certaine configuration spatiale, il existe de grandes différences entre deux visages (forme du nez, couleur des yeux, couleur de peau,...).
- Certains caractères morphologiques peuvent être présents ou non selon les visages comme par exemple la moustache, la barbe, ...

Certains caractères extérieurs peuvent déformer des caractères morphologiques comme par exemple le bronzage modifiant partiellement la couleur de la peau, l'âge peut modifier les rides du visage ou la couleur des cheveux, un éventuel accident peut laisser ses traces sur le visage, les lunettes ;

- Un visage peut avoir des orientations et des dimensions très différentes. Il s'y ajoute les conditions d'éclairage et la position dans l'image où certaines zones du visage peuvent être cachées soit par un objet soit par un autre visage.
- Les visages sont avant tout des structures 3D dans un espace 3D, de nombreux paramètres s'ajoutent encore au problème original : des contraintes de luminosité (dus soit à la position de la tête, soit au type d'éclairage choisi), de couleur, d'ombres et de rotations éventuelles de la tête ;

La détection de visages doit être en temps réel surtout lorsque le résultat de la détection demande une réaction en temps réel comme par exemple une personne s'introduisant dans une zone très dangereuse.

Par conséquent, la détection automatique de visages dans une image reste un domaine de recherche très vaste et très riche en termes d'approches et techniques utilisées.

Dans notre travail on va s'intéresser à des images du visage sans orientations et sans expressions faciales. Et la méthode qu'on a utilisée pour la détection du visage est la méthode de Viola et Jones.

I_4 Le détecteur de Viola et Jones :

Paul Viola et Michael Jones ont publié la première version de leur papier le 13 juillet 2001. Leur travail est inspiré de ceux de Papageorgiou, Oren et Poggio en 1998, qui utilisent eux-mêmes les ondelettes de Haar. Ils utilisent également les travaux de Paul Viola et Kinh Tieu, concernant l'utilisation de techniques de boost (inspiré d'AdaBoost). La méthode Viola et Jones est considérée comme l'une des plus efficaces en détection de visage, et est devenue un standard dans le domaine.

Le détecteur de Viola et Jones opère sur l'information de luminance uniquement. Il génère une boîte englobante rectangulaire du visage très fiable lorsque le visage est de face ou légèrement de profil. Cependant, il est défaillant en présence de fortes occultations.

I_4_1 Fonctionnement des détecteurs du Viola et Jones :

Cette méthode permet d'analyser une image, sans avoir besoin d'étudier chaque pixel la constituant. En effet, la notion d'image intégrale permet de définir plusieurs zones rectangulaires au sein d'une image. L'intérêt de cette technique réside dans le fait qu'elle offre la possibilité d'accéder à la valeur des autres zones à gauche et au-dessus de la zone sur laquelle nous sommes.

Ces zones permettent de créer des caractéristiques pseudo-Haar, qui sont en fait des masques permettant de déterminer plusieurs patterns :

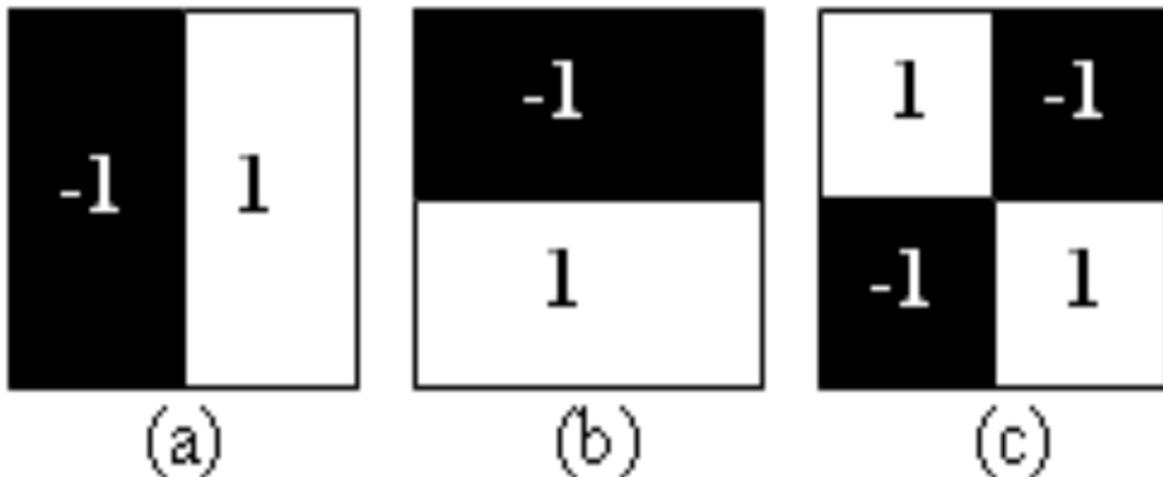


Figure I_3 : Exemple de caractéristiques pseudo-Haar.

Ces caractéristiques permettent de détecter des motifs. Par exemple, la reconnaissance des visages est rendue possible par:

- La variation de l'intensité de la lumière entre les yeux et le nez.
- Et la variation de l'intensité de la lumière entre les yeux et les pommettes.

La méthode repose donc sur l'utilisation de ce caractéristique pseudo Haar et des images intégrales, améliorant ainsi la vitesse de traitement. Deux autres mécanismes interviennent: le classifieur et le boosting.

I_4_2 Apprentissage du viola et Jones:

L'apprentissage consiste à analyser un très grand nombre d'images positives et négatives.

Plusieurs milliers d'exemples sont nécessaires, et leur analyse prend plusieurs semaines. L'apprentissage comprend :

- Le calcul des caractéristiques pseudos Haar sont positifs ou négatifs
- L'entraînement de la cascade : pour chaque étage, ont créé un classifieur fort, par ajout successif de classifieur faibles entraînés sur la même caractéristique. On ajoute ces classificateurs faibles jusqu'à obtenir un classifieur fort possédant la performance souhaité.

I_4_3 Détection du viola et jones:

La détection est l'étape qui s'applique sur l'image dans laquelle on souhaite déceler la présence d'objet.

On parcourt l'ensemble de l'image à toutes les positions et échelles, avec une sous fenêtre de 24x24 pixels. On applique la cascade à chaque sous fenêtre:

- Calcul du caractéristique pseudo Haar
- Calcul de la réponse du classifieur
- Passage à l'étage supérieur si la réponse est positive (sous fenêtre suivante sinon)
- Si on a traversé tous les étages successivement, l'exemple est déclaré positif

Enfin, on fusionne les détections multiples qui se chevauchent.

I_4_4 Limites et Applications du Viola & Jones :

Pour améliorer la méthode en détection de visage de côté, Viola et Jones ont proposés une amélioration, qui consiste à apprendre une cascade dédiée pour chaque orientation de vue.

L'utilisation d'arbre de décision permet de parcourir les espaces des cascades avec une bonne complexité algorithmique. Cependant, la génération de la cascade prend entre plusieurs jours et plusieurs semaines, ce qui limite fortement les possibilités de tests dans le choix des paramètres.

La méthode de Viola et Jones est principalement utilisés dans la détection de visage et la détection de personnes. Les systèmes de vidéosurveillance, d'IHM multimodales sont des exemples d'applications concrètes. Dans le grand public, les appareils photos numériques utilisent cette méthode pour détecter les visages sur une photographie, et effectuer un autofocus sur les visages.

I_4_5 Algorithme AdaBoost de Viola et Jones pour la détection de visages:

L'algorithme AdaBoost est un algorithme plus sophistiqué pour le boosting de la combinaison des hypothèses. Il est appelée Adaptive dans le sens que les exemples qui sont mal classifiés doivent obtenir les poids les plus élevés dans la prochaine itération, par

exemple, les exemples près de la frontière de décision sont plus difficiles à classer et à cet effet obtenir des poids élevés dans l'ensemble d'entrée après les premières itérations. Viola et Jones ont proposé un algorithme AdaBoost avec une détection en cascade et lui appliqué avec succès à la détection de visages. Dans AdaBoost de Viola, le classifieur faible a été établi par une caractéristique rectangulaire, l'apprentissage par AdaBoost est adopté pour combiner ces classificateurs faibles. Par conséquent, dans un certain sens, dans AdaBoost de Viola, le classifieur faible est un peu équivalent à une caractéristique faible. Pour le problème spécifique de la détection de visages.

I_4_5_1 Algorithme Adaboost :

Soient les images d'exemple

$$S = (x_1, y_1), \dots, (x_n, y_n) \quad (1)$$

ou $Y = 0.1$ pour les exemples négatifs et positifs respectivement, Ou $i = 1, \dots, n$

Initialiser les poids :

$$W_{t,i} = \frac{1}{2m}, \frac{1}{2l} \quad (2)$$

Pour $Y_i = 0.1$ respectivement, ou l et m sont des nombres des exemples négatifs et positifs respectivement.

Pour $t = 1, \dots, T$:

1. normaliser les poids.

$$W_{t,i} \leftarrow \frac{W_{t,i}}{\sum_{j=1}^n W_{t,j}} \quad (3)$$

Alors que, W_t est une loi de probabilité.

2. pour chaque attribut (caractéristique) j , lancer un apprentissage d'un classificateur h_j .

Le taux d'erreur est calculé par rapport à W_i ,

$$\varepsilon_j = \sum_i W_i |h_j(x_i) - Y_i| \quad (4)$$

3. choisir le classifieur h_t , avec la plus basse erreur ε_t
4. faire la mise à jour des pondérations:

$$W_{t+1,i} = W_{t,i} \beta_t^{1-e_i} \quad (5)$$

Ou $e_i = 0$ si l'image x_i est classifiée correctement sinon $e_i = 1$ et $\beta_t = \frac{\varepsilon_t}{1-\varepsilon_t}$

Le classificateur fort est construit en faisant la somme des classificateurs faibles de la façon suivante :

$$H(x) = \begin{cases} 1 & \text{si } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{sinon} \end{cases} \quad (6)$$

Ou $\alpha_t = \log \frac{1}{\beta_t}$.

L'itération suivante (et donc le choix de la prochaine fonction faible). L'erreur de h_t dans la nouvelle distribution W_{t+1} . Est exactement 1/2. La fonction de discrimination obtenue est une somme pondérée des fonctions faibles générées par l'algorithme. Le facteur β_t s'est réduit avec ε_t , ce qui incrémente la différence entre les distributions W_t et W_{t+1} . Un β_t plus petit, signifie que la fonction faible est plus précise et va avoir une influence plus importante dans le classifieur $H(x)$ à travers la pondération α_t .

L'erreur de G est bornée par :

$$\varepsilon_H \leq 2^T \prod_{t=1}^T \sqrt{\varepsilon_t (1 - \varepsilon_t)} \quad (7)$$

I_4_5_2 Détection en Cascade :

La cascade est la combinaison successive des classifieurs forts, dont la complexité est croissante tout au long de la structure. Au départ, les classifieurs simples, éliminent la plupart des fausses alarmes, avant que les classifieurs plus complexes 1 soient utilisés pour éliminer des fenêtres plus difficiles. L'attention du système se concentre, au fur et à mesure des étages de la cascade, sur des zones de plus en plus réduites et pertinentes. Viola et Jones [14] consiste en l'introduction d'un apprentissage basé sur la méthode de dopage qui permet d'obtenir une fonction de classification avec une architecture en cascade, qu'ils appellent Cascade Attentionnelle. L'architecture de cascade est très efficace, l'image d'entrée est passée par une séquence de classifieurs. Pour qu'une image soit classifiée en tant que visage, elle doit passer par tous les classifieurs. La classification s'arrête immédiatement si un quelconque de ces étapes échoue.

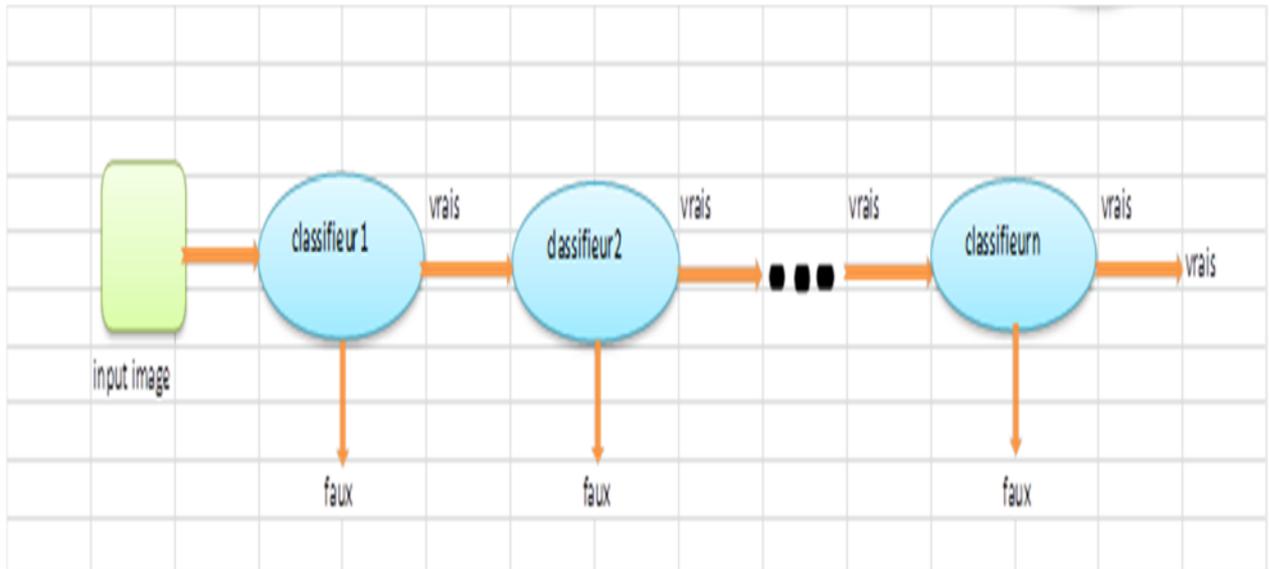


Figure I_4 Architecture de cascade.

La décision « fausse » stoppe le calcul et impose le détecteur de retourner « faux ». Une décision « vraie » passe l'image d'entrée au prochain classifieur dans la cascade. Si tous les classifieurs votent vraie alors l'entrée est classifiée comme un vrai exemple. Si n'importe quel classifieur vote « faux » alors le calcul se stoppe et l'entrée sera classifiée comme fausse.

La structure en cascade a trois paramètres principaux qu'il faut déterminer :

K : Le nombre total des classifieurs.

n_i : Nombre de caractéristiques pour chaque étape i .

θ_i : Seuil pour chaque étape i .

Pour une cascade, le taux total de faux positifs est le produit du taux de faux positifs de chaque étape :

$$F = \prod_{i=1}^K f_i \quad (8)$$

Et de même pour le taux total des détections correctes :

$$D = \prod_{i=1}^K d_i \quad (9)$$

I_5 Conclusion :

Nous avons étudié dans ce chapitre la détection de visages qui consiste à détecter des visages humains dans une image numérique. Il s'agit alors de repérer la présence dans l'image de caractéristiques de visages humains, réunis sur une surface suffisamment petite pour qu'on puisse considérer qu'ils appartiennent à la même personne.

On a présenté également dans ce chapitre le détecteur de Viola et Jones et son algorithme Adaboost.

II_1 Introduction :

La détection de point d'intérêt est une terminologie récente dans la vision par ordinateur qui fait référence à la détection de points d'intérêt pour un traitement ultérieur.

La détection de points d'intérêts peut cependant être gênée par plusieurs facteurs tels que la présence de bruit dans les images, les occultations partielles ou totales, les erreurs de détection de points, etc.

Dans ce chapitre nous allons présenter la définition des points d'intérêt et quelques détecteurs de points d'intérêts.

II_2 Définition des points d'intérêt :

La notion de points d'intérêt a été introduite pour la première fois par Moravec [15] : les points d'intérêts correspondent à des doubles discontinuités de la fonction d'intensité provoquées par des discontinuités de la fonction réflectance ou de profondeur, comme par exemple les coins, les jonctions en T, en Y ou des points de forte variation de texture (Figure II_1).

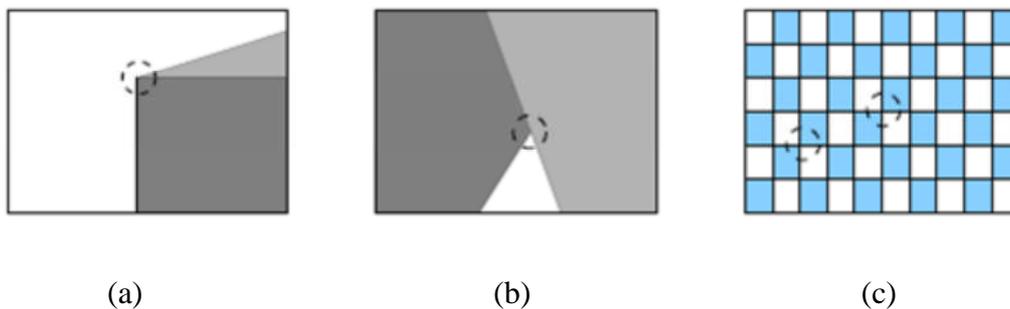


Figure II_1 Types de points d'intérêt : coins, jonction en Y et points de fortes variations de texture.

Historiquement, on a cherché à extraire les coins des objets dans une image, toutefois, un point d'intérêt est plus général qu'un coin, pourtant ils sont utilisés dans la littérature comme équivalents de schmid [16]

Les différentes applications des point d'intérêt sont : la robotique, l'indexation ou la reconnaissance d'objets et on peut extraire d'une image une suite de points caractéristiques d'un objet afin de pouvoir l'indexer dans une base de données.

II_2_1 Différentes approches :

De nombreuses méthodes ont été proposées pour détecter des points d'intérêts, Elles peuvent être classées en trois catégories :

- **Approches contours** : détecter les contours puis extraction des points d'intérêts le long des contours en considérant les points de courbures maximales ainsi que les intersections de contours.
- **Approches intensité** : à partir des niveaux de gris de l'image, trouver un opérateur qui est maximal aux points d'intérêt
- **Approches à base de modèles** : identification des points d'intérêts par mise en correspondance de la fonction d'intensité avec un modèle théorique de cette fonction des points d'intérêts considèrent.

Les approches de la deuxième catégorie sont celles utilisées généralement car :

- indépendance vis à vis de la détection de contours ;
- indépendance vis à vis du type de points d'intérêts.

II_3 Détecteurs :

Il existe une quantité importante de détecteurs de points d'intérêt. Nous n'en décrivons que quelques-uns. De façon générale, un détecteur de point d'intérêt consiste à calculer une valeur de réponse représentative de l'intérêt pour chaque pixel de l'image et ensuite à sélectionner les meilleurs. Le choix d'un détecteur de points d'intérêts repose essentiellement sur l'utilisation souhaitée. Il faut par conséquent les classer afin de faciliter ce choix. La figure II -2 propose une classification chronologique des différents détecteurs qui existent :

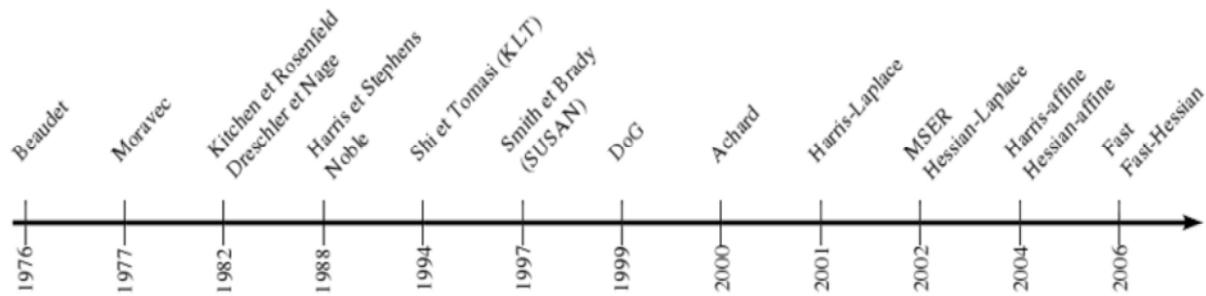


Figure II _2 : Classification chronologique des détecteurs des points qui existent.

En général, un détecteur de points d'intérêt réussi doit satisfaire les critères importants suivants :

- Tous les vrais points doivent être détectés.
- Pas de faux coins doivent être détectés.
- Les points doivent être bien localisés.
- Le détecteur de coins doit être robuste au bruit.
- Le détecteur de coins doit être rapide et efficace.

Dans ce qui suit on va présenter quelques détecteurs de point ;

II_3_1 Détecteur de Beaudet :

C'est le premier à proposer un détecteur de points d'intérêt. Cet opérateur utilise les dérivées secondes du signal pour calculer la mesure « DET » :

$$DET = I_{xx}I_{yy} - I_{xy}^2 \quad (1)$$

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 souillée. 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.

II_3_2 Détecteur de Moravec :

Moravec propose l'un des tout premiers détecteurs de points d'intérêt (Moravec en 1977), [15]. Le but était de sélectionner des points intéressants pour les mettre en correspondance de façon fiable sur des images prises par des caméras embarquées sur un robot afin de l'aider à se guider et à éviter les obstacles. Moravec définit un point d'intérêt comme étant un point où il y a une large variation de l'intensité dans certaines directions ; Son détecteur est connu comme étant un détecteur de coins. L'auteur propose d'utiliser la fonction d'autocorrélation (corrélation du signal avec lui-même décalé spatialement) pour détecter des variations de la fonction d'intensité dans un voisinage carré W . Même si le terme autocorrélation est clairement indiqué, la fonction utilisée pour détecter les coins est la somme du carré de la différence du signal avec lui-même décalé spatialement. Ainsi, étant donné un pixel (x, y) de l'image et un décalage spatial (dx, dy) , la fonction f calculée en (x, y) est définie par :

$$f(x, y, \delta_x, \delta_y) = \sum_W (I(x_k, y_k) - I(x_k + \delta_x, y_k + \delta_y))^2 \quad (2)$$

Où (x_k, y_k) sont les pixels du voisinage carré W centrée en (x, y) et I la fonction intensité. Le décalage spatial (δ_x, δ_y) . L'auteur définit une fonction C à partir de laquelle les points d'intérêt vont être détectés. La valeur de la fonction en chaque pixel correspond à la valeur minimale de la fonction $f(x, y, \delta_x, \delta_y)$ pour les 8 décalages possibles :

$$C(x, y) = \min f(x, y, \delta_x, \delta_y) \quad (3)$$

La détection des points d'intérêt est simplement réalisée les maxima locaux de C supérieurs à un seuil fixé. Le détecteur de Moravec a une faible complexité ce qui le rend intéressant pour des applications nécessitant une implémentation temps-réel.

II_3_3 Détecteur de Dreshler et Nagel [Dre82] :

Dreshler et Nagel ont trouvé un défaut à l'approche de Beaudet [17] 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.

II_3_4 Harris et Stephens :

Harris et Stephens [18] ont identifié certaines limitations au détecteur de Moravec. Ils ont alors amélioré l'approche en calculant une matrice liée à la fonction d'autocorrélation, 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 utilisée pour calculer la fonction d'autocorrélation, due au déplacement et aux directions choisies, n'est plus nécessaire. Par voie de conséquence, Harris et Stephen proposent l'opérateur suivant pour détecter les coins dans une image :

$$K(x, y) = \det(C) - \lambda \text{trace}(C) \quad (4)$$

Où λ est déterminé par l'expérience, en maximisant la réponse sur un modèle de coin idéal :

$$\lambda \approx 0,04.$$

La matrice C est définie par :

$$C = \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \quad (5)$$

Le terme trace (C) constitue l'information contour de l'opérateur, tandis que $\det(C)$ contient l'information de coin. Les valeurs de $K(x, y)$ sont positives au voisinage d'un coin, négatives au voisinage d'un contour et faibles dans une région d'intensité constante.

II_3_5 Détecteur de SUSAN :

Une autre approche, introduite par Smith et Brady [19], a conduit au détecteur SUSAN (Small est Univalve Segment Assimilating Nucleus). À la différence de toutes les méthodes déjà proposées, cette approche n'utilise pas les dérivées des intensités de l'image mais elle mesure, dans une fenêtre, le nombre de pixels ayant une intensité voisine de celle du pixel central. Une forte variation des intensités signifie l'existence d'un point d'intérêt. Ce détecteur permet de détecter les contours, les coins et aussi les jonctions.

Le principe est simple : un masque est placé sur chaque point de l'image et pour chaque point, la luminance de chaque pixel à l'intérieur du masque est comparée avec celle du centre nommé noyau. Une simple équation détermine cette comparaison :

$$C(\vec{r}, \vec{r}_0) = \begin{cases} 1 & \text{si } |I(\vec{r}) - I(\vec{r}_0)| \leq t \\ 0 & \text{si } |I(\vec{r}) - I(\vec{r}_0)| > t \end{cases} \quad (6)$$

Où \vec{r}_0 est la position du noyau dans l'image bidimensionnelle et \vec{r} est la position de n'importe quel autre pixel à l'intérieur du masque. $I(\vec{r})$ est la luminance de n'importe quel pixel, t est le seuil de variation de luminance et c est le résultat de la comparaison. Cette comparaison est faite pour chaque pixel à l'intérieur du masque et un total n des résultats de comparaison est ensuite calculé par :

$$n(\vec{r}_0) = \sum_{\vec{r}} c(\vec{r}, \vec{r}_0) \quad (7)$$

Cette valeur n donne le nombre de pixels dans l'USAN.

II_3_6 Détecteur DoG (Détecteur basé sur des différences de gaussiennes) :

Lowe [20] propose en 1999, d'intégrer la notion d'espace d'échelles dans le calcul de différences de gaussiennes, afin de rendre plus stable la détection de points d'intérêt, il définit la fonction $D(x; y; \theta)$ comme étant la différence de deux espaces d'échelles gaussiens consécutifs :

$$D(x, y, \theta) = L(x, y, k\theta) - L(x, y, \theta) \quad (8)$$

Avec k un facteur multiplicateur constant (généralement égal à $\sqrt{2}$). La Figure (3) permet de visualiser les différentes étapes de construction du DoG.

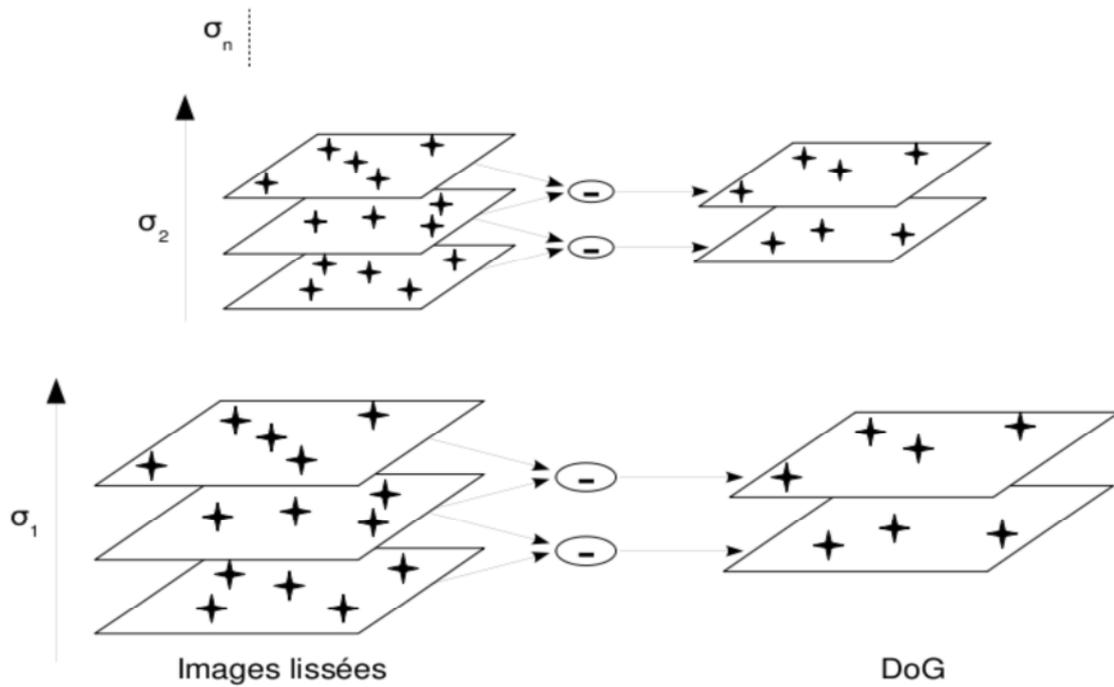


Figure II_3 Détecteur Dog.

Dans ce schéma présente des différences du gaussiennes :

- Partie gauche : ensemble des images lissées par une gaussienne.
- Partie droite : ensemble des images des différences de gaussiennes.

II_3_7 Détecteur Harris-Laplace :

Suite aux travaux de Harris et Stephens [21] et de Lindeberg, Mikolajczyk et Schmid un détecteur de points d'intérêts multi-échelles nommé Harris-Laplace utilisant les avantages de ces deux méthodes, à savoir la performance du détecteur de Harris couplée à l'aspect multi-échelles utilisé par Lindeberg avec le détecteur LoG.

Soit L_{σ_D} la représentation de la fonction d'intensité I à l'échelle δ_D . La matrice de structure A se définit par :

$$A(x,y,\delta_D, \delta_I) = \delta_D^2 \begin{bmatrix} (G_{\delta_1} * (\frac{\partial L_{\delta_D}}{\partial x})^2)(x,y) & (G_{\delta_1} * (\frac{\partial L_{\delta_D}}{\partial x} \cdot \frac{\partial L_{\delta_D}}{\partial y}))(x,y) \\ (G_{\delta_1} * (\frac{\partial L_{\delta_D}}{\partial x} \cdot \frac{\partial L_{\delta_D}}{\partial y}))(x,y) & (G_{\delta_1} * (\frac{\partial L_{\delta_D}}{\partial y})^2)(x,y) \end{bmatrix} \tag{9}$$

Où δ_I est appelé échelle d'intégration et δ_D échelle de dérivation.

Le détecteur de Harris-Laplace est considéré à l'heure actuelle comme étant le meilleur détecteur multi-échelles au sens des critères de précision.



Figure II _4 Détecteur Harris-laplace.

Les régions détectées sont, par construction, circulaires. Le détecteur Harris-Laplace en utilisant des voisinages ellipsoïdaux. L'appariement d'images est alors plus robuste aux changements de point de vue.

II_3_8 Détecteur MSER :

Matas et al. [22] proposent en 2002 une approche originale permettant de définir une région d'intérêt robuste aux transformations affines. Dans un premier temps, une étape de classification permet de regrouper en classe chaque pixel de l'image suivant sa valeur d'intensité. L'histogramme ainsi créé permet de déterminer une fonction d'intensité en se basant sur l'aire de chaque classe. En appliquant différents seuillages à cette fonction d'intensité, certaines régions de l'image classées vont varier et d'autres non. Les régions d'intérêt correspondent donc à celles qui restent robustes aux différents seuillages. Matas et al. Démontrent que les régions d'intérêt ainsi extraites, sont invariantes aux transformations affines, aussi bien photométriques que géométriques.

II_3_9 Détecteurs Harris-affine et Hessian-affine :

En 2004, Mikolajczyk et Schmid proposent de coupler un détecteur multi-échelles avec la méthode d'adaptation (ou de normalisation) affine décrite précédemment. Ils proposent ainsi deux nouveaux détecteurs [23] : le Harris-affine et le Hessian-affine.

Pour chacun d'entre eux, la méthode repose tout d'abord sur l'extraction multi-échelles des points d'intérêt, puis sur la détermination itérative d'une région locale circulaire.

La partie analyse multi-échelles de la scène est assurée par l'utilisation du détecteur

Harris-Laplace pour le Harris-affine, et celle du détecteur Hessian-Laplace pour le Hessian-affine.

II _4 Choix d'un algorithme de détection des points d'intérêts :

Pour la détection de points d'intérêts on a choisi de travailler avec l'algorithme de Harris et Stephens. Cet algorithme améliorées le détecteur d'angle de Moravec.

Nous supposons une image en deux dimensions en niveaux de gris est utilisé. Cette image est donnée par I . Envisagez de prendre une fenêtre d'image sur la zone (u, v) et le déplacement par (x, y) . Le pondérée somme des différences au carré entre ces deux fenêtres, noté S , est donné par :

$$S(x, y) = \sum_u \sum_v w(x, y) (I(u + x, v + y) - I(u, v))^2 \quad (10)$$

$I(U + X, V + Y)$ peut être approchée par un développement de Taylor . Laissez I_x et I_y être les partiels dérivés de I tels que :

$$I(u + x, v + y) \approx I(u, v) + I_x(u, v)x + I_y(u, v)y \quad (11)$$

On obtient ainsi le rapprochement :

$$S(x, y) \approx \sum_u \sum_v w(u, v) (I_x(u, v)x + I_y(u, v)y)^2 \quad (12)$$

qui peut être écrit sous forme matricielle :

$$S(x, y) \approx (x, y)A \begin{pmatrix} x \\ y \end{pmatrix} \quad (13)$$

Où A est le tenseur de structure ,

$$A = \sum_u \sum_v w(u, v) \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} = \begin{bmatrix} \langle I_x^2 \rangle & \langle I_x I_y \rangle \\ \langle I_x I_y \rangle & \langle I_y^2 \rangle \end{bmatrix} \quad (14)$$

avec :

I_x : Gradient X

I_y : Gradient Y

I_x^2 : est l'image convolutive 2 fois avec le gradient X

I_y^2 : est l'image convolutive 2 fois avec le gradient Y

Cette matrice est une matrice Harris et crochets désignent la moyenne. Si une fenêtre circulaire (ou fenêtre pondérée circulaire, par exemple une gaussienne) est utilisée, alors la réponse sera isotrope.

Un coin (ou en général un point d'intérêt) est caractérisé par une grande variation de S dans toutes les directions du vecteur (X, Y) . En analysant les valeurs propres de A , cette caractérisation peut être exprimée de la façon suivante : A devrait avoir deux valeurs propres (λ_1, λ_2) «grands» pour un point d'intérêt. Basé sur les amplitudes des valeurs propres, les conclusions suivantes peuvent être prises en fonction de cet argument :

1. Si $\lambda_1 \approx 0$ et $\lambda_2 \approx 0$ alors ce pixel (x, y) n'a pas de caractéristiques d'intérêt.
2. Si $\lambda_1 \approx 0$ et λ_2 a une certaine valeur positive, alors un bord est trouvé.
3. Si λ_1 et λ_2 ont de grandes valeurs positives, puis un coin est trouvé.

Harris et Stephens ont noté que le calcul exact des valeurs propres est coûteux en calcul, car il nécessite le calcul d'une racine carrée, et suggère plutôt la fonction suivante M_C , où k est un paramètre de sensibilité réglable :

$$M_C = \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 = \det(A) - k \text{trace}^2(A) \quad (15)$$

Par conséquent, l'algorithme montre que ce n'est pas nécessaire de calculer la décomposition en valeurs propres de la matrice A mais, il suffit d'évaluer le déterminant et la trace de A pour trouver les coins, ou plutôt des points d'intérêt en général.

Le Shi-Tomasi détecteur d'angle calcule directement $\min(\lambda_1, \lambda_2)$ parce que sous certaines hypothèses, les coins sont plus stables pour le suivi. Notez que cette méthode est également parfois désignée sous le détecteur d'angle Kanade-Tomasi.

La valeur de k doit être déterminée de manière empirique, et dans les valeurs de la littérature dans la gamme $[0,04 ; 0,15]$ ont été signalés que possible.

On peut éviter de fixer le paramètre k à l'aide de Noble la mesure de coin M'_C qui se élève à la moyenne harmonique des valeurs propres :

$$M'_C = 2 \frac{\det(A)}{\text{trace}(A) + \varepsilon'} \quad (16)$$

ε' Étant une petite constante positive.

La matrice de covariance de la position d'angle est A^{-1} , soit :

$$A^{-1} = \frac{1}{\langle I_x^2 \rangle \langle I_y^2 \rangle - \langle I_x I_y \rangle^2} \begin{bmatrix} \langle I_y^2 \rangle & -\langle I_x I_y \rangle \\ -\langle I_x I_y \rangle & \langle I_x^2 \rangle \end{bmatrix} \quad (17)$$

II_5 Points d'intérêt invariants à l'échelle :

Afin de trouver un ensemble de points plus stables, nous cherchons les maxima dans la représentation en échelle (tridimensionnelle) d'une image. Un point dans l'espace est considéré comme un point d'intérêt s'il est un maximum dans le voisinage le plus proche et si, de plus, sa valeur est plus grande qu'un certain seuil,

$$\forall x_n \in W \text{seuil}_a < F(I, X, s_n) > F(I, X_w, S_m), m \in \{n-1, n+1\}$$

où w est le voisinage immédiat du point x .

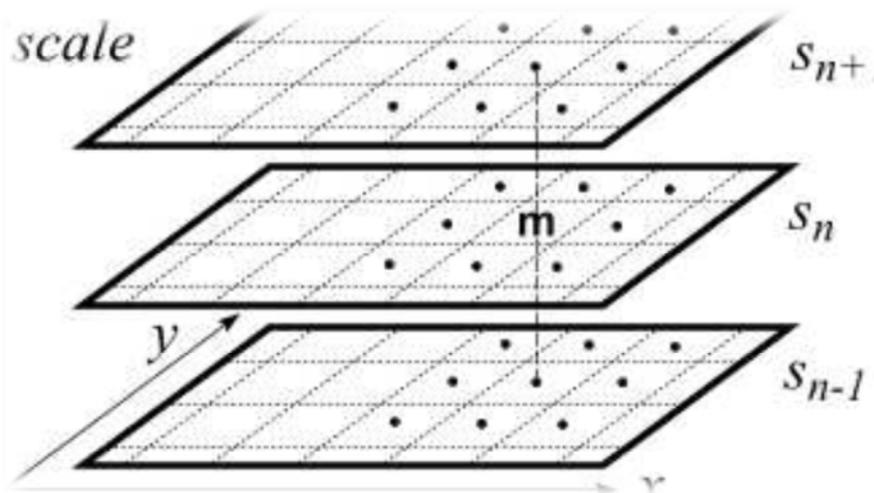


Figure II_5 Le maximum de la fonction $F(I, X, s_n)$ dans l'espace d'échelle.

La fonction de Harris atteint rarement des maxima dans l'espace 3D. Si trop peu de points sont détectés, la représentation de l'image n'est pas robuste aux différentes transformations. Pour résoudre ce problème, deux représentations en échelle sont construites

pour une image donnée. La première est créée par la fonction de Harris adaptée. On obtient un ensemble de points en détectant à chaque niveau de la pyramide des maxima dans la surface d'image ; La majorité des points (x,y) sont détectés à des endroits très proches sur plusieurs niveaux d'échelle consécutifs.

$$\forall x_n \in W \text{seuil}_h < F(I, X, s_n) > F(I, X_w, S_m),$$

Afin d'obtenir une représentation plus compacte, on vérifie pour chaque point détecté précédemment s'il constitue un maximum dans la dimension d'échelle dans la deuxième représentation. La deuxième représentation est construite avec le Laplacien :

$$F(I, X, S_{n-1}) < F(I, X, S_n) > F(I, X, S_{n+1}) \cap F(I, X, S_n) > \text{seuil}_l$$

Avec seuil_l et seuil_h choisis expérimentalement. Le détecteur de Harris modifié détermine la localisation de points d'intérêt dans la surface d'image pour un niveau de résolution donné. La fonction Laplacien détermine l'échelle caractéristique d'un point d'intérêt. Le choix de la fonction de Harris a été motivé par les résultats de la comparaison de différents détecteurs menée par [24], où le détecteur de Harris s'est avéré le plus fiable. Une extension de ces travaux proposée a montré la possibilité d'adapter ce détecteur aux changements d'échelle. Le choix de la fonction de recherche de maxima dans la direction d'échelle est moins critique. Nous proposons d'utiliser l'opérateur Laplacien. La figure 5 montre la représentation en échelle de deux images réelles avec des points détectés par la méthode de Harris- Laplacien. Pour les deux images présentant le même objet nous montrons les points d'intérêt sur des niveaux d'échelle consécutifs pour lesquels les points ont été détectés. Il y a beaucoup de points qui se correspondent sur les niveaux (indiqué par les flèches) pour lesquels la relation des échelles correspond au facteur d'échelle réel.

Les points sont alors caractéristiques sur la surface d'image et dans la dimension d'échelle.



Figure II _6 Points détectés sur différents niveaux de résolution par la méthode de Harris-Laplacien.

II _6 Conclusion :

Dans ce chapitre, nous avons présenté la définition des point intérêt et quelques détecteurs. Parmi l'ensemble des détecteurs qui existe on a choisi de travailler avec la méthode de Harris Stephen.

III_1 Introduction :

L'objectif de notre travail est l'appariement entre deux images. L'appariement des images s'intéresse prioritairement au cas où l'on dispose de deux images à comparer, Ces méthodes existantes sont soit basées sur des grandeurs géométriques soit sur des grandeurs photométriques.

Pour faire l'appariement entre deux images on a utilisé la méthode de la corrélation. Dans ce chapitre on va présenter quelques méthodes de caractérisation des points d'intérêts ainsi que la méthode de corrélation.

III_2 Caractérisation des points d'intérêts :

Pour la mise en relation des différents points détectés, l'utilisation d'une méthode de description est indispensable. Cette dernière permet de caractériser chaque point d'intérêt, et d'en extraire différentes composantes (intensités, informations sur le voisinage, échelles, gradients, ...).

Il existe une multitude de méthodes de description, chacune ayant ses avantages et ses inconvénients. Il est donc indispensable d'en connaître les caractéristiques afin de choisir la plus appropriée à la problématique. D'un point de vue historique, la première méthode à avoir été proposée se base uniquement sur le point d'intérêt, et plus particulièrement sur l'utilisation de ses coordonnées. Par la suite l'observation de son voisinage a permis d'améliorer fortement les résultats. Pour ce faire, un certain nombre de méthodes ont été créées se basant sur un résumé visuel plutôt que sur les pixels, nous pouvons citer les moments, les transformées, ou encore les histogrammes. Nous proposons donc de détailler ces différents descripteurs.

III_2_1 Descripteur basé sur les moments :

La description basée sur les moments a été initialement proposée pour de la reconnaissance d'objets. Ils étaient utilisés pour travailler sur des images contenant l'objet entier. L'avantage des moments, tels que les moments de Hu et de Zernike, est leur invariance aux translations, rotations et changements d'échelle isotrope. Néanmoins, un changement de point de vue, un changement d'échelle anisotrope ou l'ajout d'occultations, provoque une forte diminution de la qualité des résultats. En se basant sur les études comparatives proposées par Choksuriwong et al. En 2005 [25] et 2008 [26], nous proposons d'en détailler les principaux composants.

III 2 1 1 Les moments de Hu :

Introduits en 1962 par Hu [27], leur utilisation était tout d'abord globale, puis en limitant le calcul des moments à un voisinage d'un point d'intérêt, cette méthode est devenue locale. Cette dernière se base sur la détermination des moments centraux et de leur normalisation. Soit un moment d'ordre $(p + q)$ avec $(p, q > 0)$, cette l'équation est détaillé la détermination du moment central.

$$M_{p,q}(x,y) = \int_{\Omega} x_p y_q (x + x_0, y + y_0) dx dy \quad (1)$$

ou (x_0, y_0) représentent les coordonnées du point d'intérêt. L'équation suivante est caractérisé la normalisation de ce moment.

$$\mu_{p,q} = \frac{m_{p,q}}{m_{0,0}^{1+(p+q)/2}} \quad (2)$$

Hu propose donc d'utiliser une succession de polynômes, basés sur l'équation (2), afin de créer un descripteur local.

III 2 1 2 Les moments de Zernike :

Les polynômes de Zernike ont été introduits en 1934 avant d'être utilisés dans le domaine de l'optique, de la robotique, puis en vision par ordinateur, notamment en 2003 par Chong et al. [28]. D'un point de vue général, un polynôme de Zernike, noté $P(r;\theta)$ et caractérisé par son rayon r et son angle θ , est défini par :

$$P_{mn}(r;\theta) = R_{mn}(r)e^{-in\theta} \quad (3)$$

Ou m et n représentent l'ordre du moment, et $R_{mn}(r)$ le polynôme radial orthogonal défini par :

$$R_{mn}(r) = \sum_{s=0}^{m-n/2} \frac{(m-s)!}{s! \left(\frac{m+n}{2}-s\right)! \left(\frac{m-n}{2}-s\right)!} r^{m-2s} \quad (4)$$

Afin de décrire localement un point d'intérêt, l'auteur propose d'utiliser les moments de Zernike. Ces derniers se basent sur un ensemble de polynômes décrits dans l'équation (4), caractérisant ainsi un ensemble orthogonal défini sur un disque unité. Il est par conséquent possible de déterminer les moments de Zernike de la façon suivante :

$$M_{zmn} = \frac{m+1}{\pi} \int_{\Omega} Vi(x;\Omega) [Pmn(r,\theta)] dr d\theta \quad (5)$$

Finalement, le descripteur est composé de seize moments de Zernike.

III _2_2 Descripteur basé sur les transformées intégrales :

Il existe un certain nombre de méthodes utilisant les transformées intégrales pour décrire localement un point d'intérêt. Ghorbel en donne une description succincte en 1994, puis Derrode et al. S'intéressent à l'utilisation des transformées de Fourier en 1999[29]. Plus récemment Mennesson et al. [30] Présentent de nouveaux descripteurs utilisant notamment une généralisation des transformées notée T, la transformée de Fourier locale d'un point d'intérêt est égale a :

$$T_F(V_x, V_y) = \int_{R^2} V_I(X; \Omega) e^{-2i\pi(xV_x + yV_y)} dx dy \quad (6)$$

Il est également possible d'utiliser la transformée de Fourier circulaire, définie par l'équation suivante :

$$T_{FC}(r, V_\theta) = \frac{1}{2\pi} \int_0^{2\pi} V_I(r, \theta; \Omega) e^{-2i\pi\theta V_\theta} d\theta \quad (7)$$

Ou r et θ correspondent aux coordonnées polaires du point.

Derrode et al. Proposent également l'utilisation de la transformée de Fourier-Mellin définie par :

$$T_{FM}(V_x, V_y) = \int_0^{+\infty} \int_0^{+\infty} V_I(X; \Omega) x^{i(V_x-1)} y^{i(V_y-1)} dx dy \quad (8)$$

En couplant ces différents types de transformées, les auteurs obtiennent des invariances par translation (avec Fourier), par rotation (avec Fourier circulaire et Fourier-Mellin) et par changement d'échelles (avec Fourier-Mellin).

III _2_3 Descripteur basé sur les histogrammes :

On général, l'histogramme représente une estimation de la distribution des intensités de l'image. Swain et Ballard [31] utilisent en 1991 des histogrammes ayant pour objectif la reconnaissance d'objets. De nombreuses méthodes ont dès lors été proposées et il est possible de les classer en deux catégories, l'une s'appuyant sur des histogrammes d'intensités lumineuses et l'autre sur des histogrammes de gradients orientés. Nous proposons d'en donner une description afin de faciliter le choix de la méthode à utiliser.

- Histogramme d'intensité lumineuse,

- Histogramme de gradients orientés.

III_2_3_1 Histogramme d'intensité lumineuse (ou de couleur) :

L'histogramme de couleur proposé en 1991 par Swain et Ballard [32] puis repris par Schiele et Waibel [33] en 1995, est utilisé comme résumé visuel de l'image. Il présente l'avantage d'avoir une construction rapide et peu onéreuse en terme d'espace mémoire.

Appelé communément « estimateur de densité non-paramétrique du premier ordre », l'histogramme consiste en un graphique statistique permettant de représenter la distribution des intensités des pixels. Généralement appliqué à l'image entière, il est néanmoins possible de l'extraire dans le voisinage d'un point d'intérêt. L'histogramme h du voisinage d'un point est défini par :

$$h_x(i) = \frac{1}{\text{card}(\Omega)} \sum_{x \in \Omega} (I_x = i) \quad (9)$$

Avec $i \in [0,255]$

Pour résumer, l'histogramme consiste donc à comptabiliser le nombre de pixels présentant la même valeur d'intensité dans le voisinage considéré.

III_2_3_2 Histogramme de gradients orientés (HOG) :

Proposés en 2005 par Dalal et Triggs [34], les histogrammes de gradients orientés sont utilisés principalement en vision par ordinateur pour de la détection d'objets. Leurs utilisations se sont également révélées particulièrement efficaces pour la détection de personnes. D'un point de vue général, un gradient permet de calculer les variations d'une fonction par rapport aux changements de ses paramètres. Dans le cas d'une image, la détermination du gradient consiste à calculer la variation de l'intensité des pixels dans différentes directions. En effectuant une telle analyse sur une image entière ou sur un voisinage, nous construisons un histogramme de gradients orientés permettant d'étudier les orientations des gradients locaux.

III_3 Appariement d'image :

L'appariement des images est l'un des difficiles problèmes de base de la vision. Il apparaît dès que l'on veut utiliser plusieurs images. Prenons, par exemple, le cas de la reconstruction : on dispose de deux images et on veut retrouver la géométrie de cet objet. On doit résoudre deux problèmes :

- étant donné un point de l'objet, il faut retrouver sa projection dans chacune des deux images : c'est le problème de l'appariement ;
- une fois les deux projections connues, il faut calculer la position du point de l'objet correspondant.

On peut tout d'abord distinguer celles destinées aux images en niveaux de gris (ou de couleur mais bien plus rarement) et celles qui utilisent des primitives extraites des images. Dans les premières, on peut chercher à appairer soit tous les points, soit seulement quelques-uns, choisis par exemple à l'aide d'un extracteur de points ou répartis de manière uniforme dans l'image. À chacun de ces points est associée une valeur, dite niveau de gris, qui code l'intensité lumineuse reçue. En reprenant la classification proposée dans [38], on peut trier les méthodes en trois groupes:

1. Les méthodes de relaxation: dans ces méthodes, on commence par proposer quelques appariements. On utilise alors les contraintes déduites de ces premiers appariements pour en faire d'autres. Puis, on réitère le processus. Des algorithmes basés sur ce principe ont été proposés dans la littérature [35], [36], avec des améliorations ultérieures [37].
2. Les méthodes de programmation dynamique : dans ce cas, on pose le problème de la mise en correspondance sous la forme de minimisation d'une fonction de nombreuses variables discrètes.
3. Les méthodes de corrélation : C'est la ressemblance entre deux images centrées sur les points dont on veut savoir s'ils sont en correspondance. C'est la méthode qu'on a utilisés dans notre travail, et dans la suite on va le détailler plus.

II_3_1 Principe de l'appariement :

Les grandeurs géométriques s'avèrent instables et peu discriminantes pour des images réelles.

Leur instabilité provient donc des erreurs lors de l'extraction de la description symbolique et leur manque de discriminante du fait qu'une partie de l'information, l'information de niveaux de gris, n'est pas utilisée. Les méthodes basées sur des grandeurs photométriques sont par nature plus discriminantes car moins symboliques. Elles sont en effet plus proches du signal. Par contre leur défaut principal est qu'elles ne sont pas invariantes à des types de transformations qui peuvent être pris en compte par les méthodes géométriques.

L'algorithme d'appariement est basé sur la caractérisation locale du signal. Cette caractérisation est invariante aux transformations d'image. Ceci permet d'être à la fois discriminant et invariant aux transformations images. Pour pouvoir appairer les vecteurs d'invariants composant cette caractérisation, il faut d'abord introduire une distance qui permet de les comparer. Il est ensuite possible de choisir dans deux images les vecteurs correspondant au même point physique en utilisant le principe de l'appariement croisé. Pour augmenter la robustesse des contraintes semi-locales sont ajoutées.

III_3_2 Discussion sur les méthodes d'appariement :

Il existe deux grandes classes de méthodes d'appariement.

La première regroupe les méthodes utilisant des informations locales telles que les mises en correspondances de motifs [39], de primitives (coins, lignes) ainsi que les méthodes utilisant le flot optique. L'inconvénient des deux premières méthodes est le faible nombre de points mis en correspondance. De plus, les méthodes basées sur des informations locales sont peu robustes vis-à-vis des scènes ayant beaucoup de disparité (profondeur) et aux régions peu texturées.

La deuxième classe regroupe les méthodes utilisant des informations globales.

Appartiennent à ce groupe les méthodes utilisant la programmation dynamique, celles s'appuyant sur des caractéristiques invariantes aux transformations rigides ou encore des techniques utilisant la propagation de croyance [40] et la diffusion non linéaire [41].

Ces méthodes sont généralement moins sensibles aux problèmes des régions ambiguës mais elles sont plus coûteuses en temps de calcul et l'élaboration d'un modèle global n'est pas toujours possible.

La majeure partie des méthodes classiques de mise en correspondance utilisent une représentation statistique de l'erreur d'appariement.

On est en droit de se poser des questions sur l'utilisation des outils statistiques lorsque le nombre d'informations à mettre en correspondance ne dépasse pas deux. Les techniques de régularisation s'appuient sur l'hypothèse du bruit de mise en correspondance et utilisent le voisinage de chaque pixel pour évaluer une représentation de la statistique des variations de niveaux de gris.

Plus le voisinage utilisé est étendu, plus la statistique sera précise, mais moins elle sera fiable.

III_3_3 L'appariement par corrélation :

La méthode par corrélation est principalement utilisée dans l'analyse de l'information d'intensités pour la mise en correspondance. Le principe est de déterminer, pour le voisinage d'un point x_1 de la première image, la corrélation maximale (distance minimale) avec un voisinage issu de la seconde image. Ce calcul permet donc d'extraire le point x_2 formant ainsi le couple (x_1, x_2) présentant la meilleure ressemblance au sens de la corrélation. La figure III.1, résume une telle mise en correspondance.

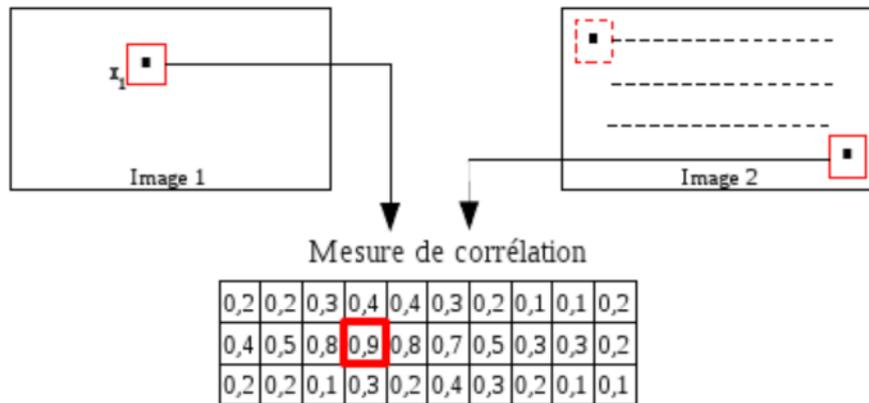


Figure III_1 Principe de la corrélation : recherche du point issu de la seconde image présentant la meilleure ressemblance.

Afin d'optimiser cette méthode, une estimation de la position de x_2 peut être introduite.

Nous déterminons alors les mesures de corrélation à l'intérieur d'une zone de recherche, et non plus sur l'image entière. La figure III.2 donne un aperçu de la zone de recherche.

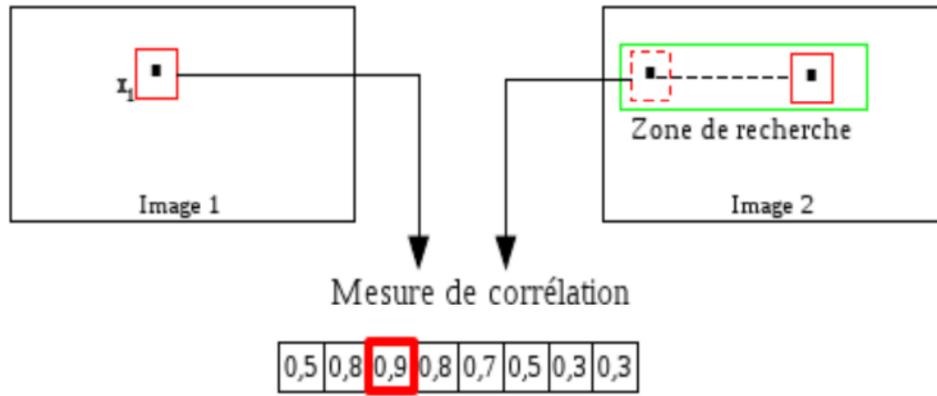


Figure III_2 Exemple de mesures de corrélation entre le point x_1 et

les points x_2 présents dans la zone de recherche.

La plus grande difficulté réside dans le choix de la mesure de corrélation. En effet il en existe un grand nombre, Ashwanden et Guggenbuhl [28] proposent en 1992 de les lister et d'en donner une courte description. Les plus classiques sont la somme des distances au carrée S est définie par :

$$S(x_1, x_2) = \sum_{i=-N}^N \sum_{j=-P}^P (I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j))^2 \quad (10)$$

et la somme des valeurs absolues des distances S_A est Définie par :

$$S_A(x_1, x_2) = \sum_{i=-N}^N \sum_{j=-P}^P |I_1(x_1 + i, y_1 + j) - I_2(x_2 + i, y_2 + j)| \quad (11)$$

En ajoutant des informations locales à ces deux mesures de corrélation, On présente Z une fiabilité supérieure aux autres et est définie par :

$$Z(x_1, x_2) = \frac{\sum_{i=-N}^N \sum_{j=-P}^P I_1(x_1 + i, y_1 + j) - \bar{I}_1)(I_2(x_2 + i, y_2 + j) - \bar{I}_2)}{\sqrt{\sum_{i=-N}^N \sum_{j=-P}^P I_1(x_1 + i, y_1 + j) - \bar{I}_1)^2 \sum_{i=-N}^N \sum_{j=-P}^P I_2(x_2 + i, y_2 + j) - \bar{I}_2)^2}} \quad (12)$$

Avec :

I_1 et I_2 sont respectivement les niveaux de gris des pixels dans les images gauche et droite,

i et j représentent les translations horizontales et verticales dans la zone d'agrégation

\bar{I}_1 et \bar{I}_2 sont les moyennes des intensités sur chaque fenêtre de similarité centrée en $(x ; y)$

I_k : Représentant la moyenne de l'image I_k , déterminée par :

$$I_k = \frac{1}{(2N+1)(2P+1)} \sum_{i=-N}^N \sum_{j=-P}^P I_k(x_k + i, y_k + j) \quad (13)$$

Il existe également des méthodes de corrélations utilisant un filtrage de l'image, Un dernier type de mesure peut être obtenu en s'appuyant sur l'information locale de l'orientation permettant ainsi d'accroître la robustesse aux rotations de l'image.

III_3_3_1 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.

III_3_3_2 Propriétés du la corrélation :

En exposant les mesures de corrélation, nous en présentons les caractéristiques suivantes:

- l'intervalle de variation des mesures formé d'un minorant et d'un majorant ;
- le type de la mesure : similarité ou dissimilarité ;
- les invariances par rapport aux transformations des niveaux de gris, c'est-à-dire par rapport aux changements de luminosité.

III_3_3_3 Avantages et inconvénients :

Les méthodes à base de corrélation possèdent les avantages suivants :

- Leur mise en œuvre est simple et rapide.
- Elles ne sont pas très coûteuses en temps de calcul (elles sont moins coûteuses que les méthodes globales).
- Dans de nombreux travaux, elles ont déjà prouvé leur efficacité : dans l'évaluation de Scharstein et Szeliski [42], la méthode proposée par [43]

Les difficultés qui peuvent être liées à ces méthodes sont les suivantes :

- Elles ne prennent pas toujours en compte certaines situations qui rendent difficile la mise en correspondance : les bruits, les raccourcissements, les zones non texturées, les textures répétitives, les changements de luminosité et les occultations.
- Elles ne fonctionnent pas dans le cas de déformation des objets ou de changement d'échelle (nous parlons ici des méthodes classiques qui n'utilisent pas des zones d'agrégation qui s'adaptent à la forme des objets).
- Elles n'autorisent pas des points de vue trop éloignés.
- Comme elles sont locales, les résultats peuvent être erronés car certaines caractéristiques globales de l'image ne sont pas prises en compte. Par exemple, certaines méthodes globales segmentent l'image en régions qui correspondent à des projections d'un même objet de la scène. Cette segmentation permet de réduire l'espace des disparités envisageables, et donc de réduire les erreurs.

II_3_5 Cadre d'évaluation :

Critère d'évaluation : Etant données deux images, on souhaite que tous les points détectés soient correctement appariés. Le critère d'évaluation choisi est donc le nombre d'appariements corrects par rapport au nombre d'appariements trouvés :

$$\frac{\text{nombre d'appariements corrects}}{\text{nombre d'appariements trouvés}}$$

Pour calculer le nombre d'appariements corrects, il est indispensable de connaître la relation entre les deux images à appairer.

II_4 Conclusion :

Nous avons présenté dans ce chapitre les méthodes de caractérisation des points d'intérêts et on a étudié l'appariement avec la méthode de corrélation. Il ne s'agit alors de mettre en correspondance un point d'une image avec une autre image pour faire la comparaison.

IV_1 Introduction :

Dans ce chapitre nous allons présenter les résultats de la programmation des méthodes qu'on a étudiées. Les programmes sont appliqués sur les images en niveau de gris et en couleur. Le langage de programmation utilisé dans notre travail est le logiciel MATLAB. Ce logiciel est très utilisé en traitement d'image car il offre de nombreuses possibilités de traitement avec une palette d'outils prêt à l'emploi.

IV_3 Les images de tests :

Figure IV_1 Images test contenant un seul visage.



Figure IV_2 Images test contenant plusieurs visages.

IV_4 Détection du visage :

Détection de visage par la méthode de Viola et Jones. On a obtenu les résultats présenté ci-dessous :

IV_4.1 cas d'images à un seul visage :



Figure IV_3 Résultat d'application de la méthode Viola Jones sur des images contenant un seul visage.

IV_4_2 : Cas d'images a plusieurs visage :

Figure IV_4 Résultat d'application de la méthode Viola et Jones sur des images contenant plusieurs visages.

IV_6 Détection des points intérêt sur le visage :

Pour faire la détection des points d'intérêt on a utilisé le détecteur de Harris et Stephens, elle consiste à détecter les points caractéristiques sur les images. Et dans ce qui suit on va présenter quelques résultats obtenus après test.

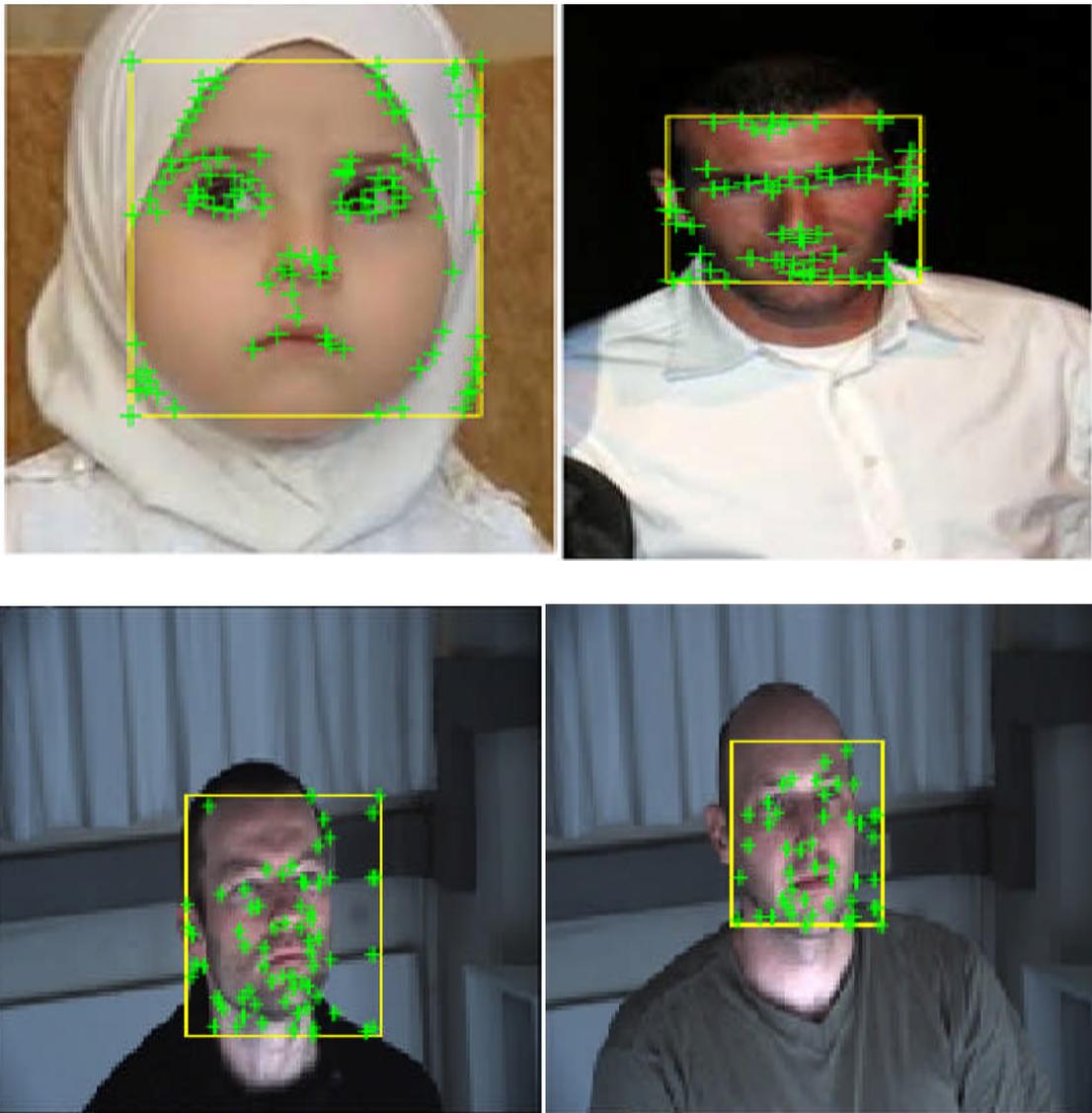


Figure IV_5 Détection des points d'intérêt sur le visage.

Presque tous les coins sont détectés, mais on constate qu'il y a une mauvaise localisation de certains points et des fausses détections.

IV_7 Appariement :

Pour éviter l'appariement manuel de points qui est une étape assez longue, nous proposons d'effectuer un appariement automatique des images. La première étape consiste à

trouver des points d'intérêts dans les images. Ensuite il faut créer des descripteurs pour chaque point d'intérêt et enfin chercher les meilleurs appariements possibles. Nous présentons ci-dessous quelques résultats obtenus en utilisant la méthode de corrélation.

IV_7_2 L'appariement entre deux images avec la méthode de corrélation :

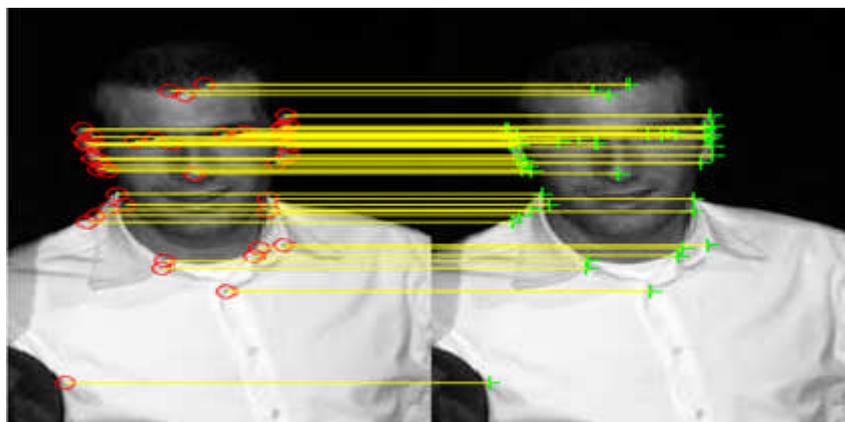
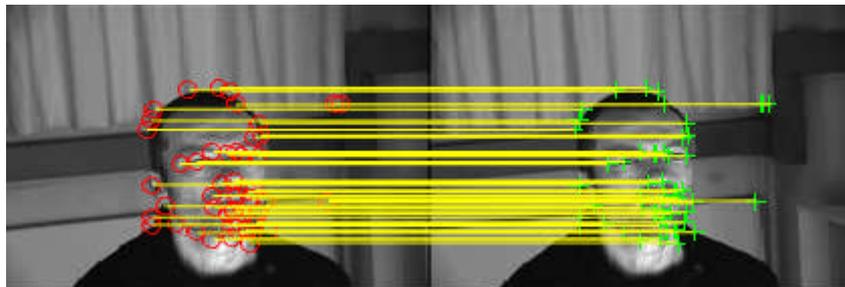
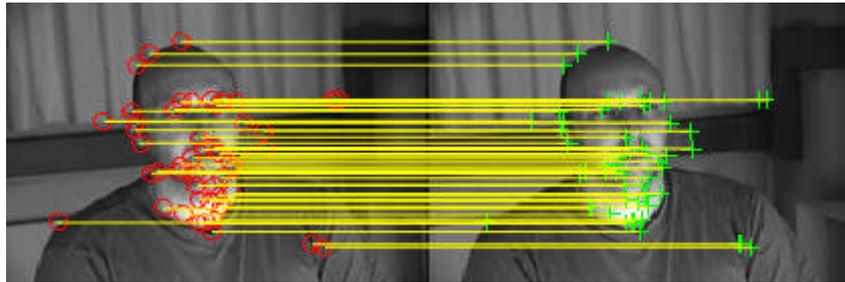




Figure IV_6 L'appariement entre deux images en utilisant la méthode de corrélation.

IV_7_1 L'appariement entre deux images du visage :

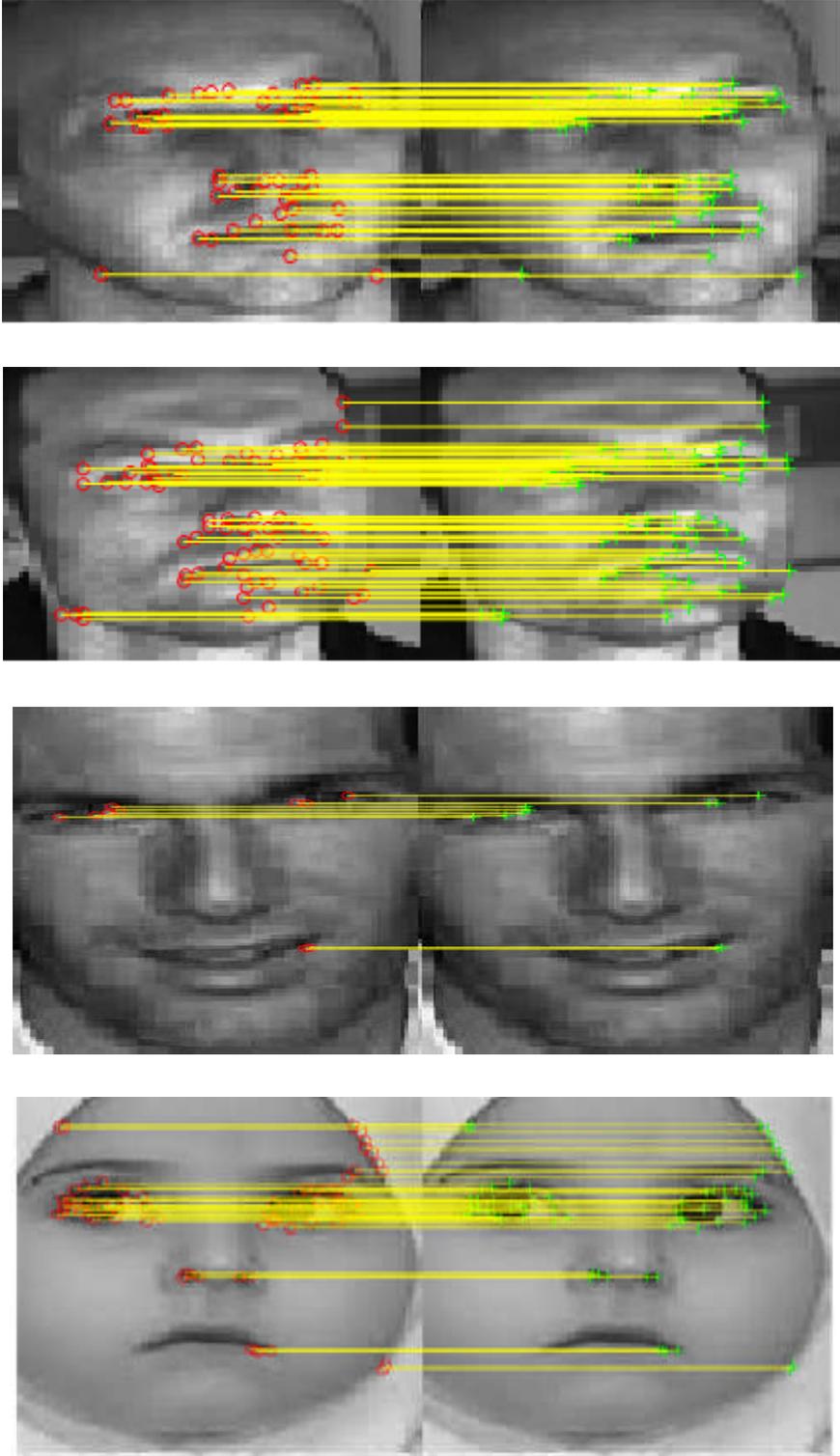


Figure IV_7 L'appariement entre deux images du visage en utilisant la méthode de corrélation.

Selon les résultats obtenus, nous avons réussi à faire l'appariement entre deux images du visage.

Les résultats obtenus montrent la capacité des méthodes appliqués. Avec la méthode Viola et Jones on arrive à détecter le visage sur toutes les images que ce soit elles contiennent un seul visage ou elles contiennent plusieurs visage.

Avec la méthode de Harris on arrive à détecter les points caractéristiques du visage sauf certains points inutiles qui apparaissent sur le visage.

La méthode de corrélation a donné de bon résultats de tel sort que tous les points détecter sont apparie chaque point sur l'image de droite au lui trouve le point correspondant sur l'image de gauche.

Conclusion :

Dans ce chapitre nous avons présenté les différents résultats obtenus. L'objectif de ce travail est de faire l'appariement entre deux images qui vient après l'étape de la détection des points d'intérêts sur celle-ci.

La méthode d'appariement utilisée a donné des résultats sur les deux types d'images : image à un seul visage et images à plusieurs visages. Comme on s'intéresse à la reconnaissance 3D du visage on a appliqué la méthode sur des images du visage.

Pour la méthode de Harris, en plus des points caractéristiques du visage, il détecte d'autres points qui sont inutiles.

L'image constitue l'un des moyens les plus importants qu'utilise l'homme pour communiquer avec autrui. C'est un moyen de communication universel dont la richesse du contenu permet aux êtres humains de tout âge et de toute culture de se comprendre.

C'est aussi le moyen le plus efficace pour communiquer, chacun peut analyser l'image à sa manière, pour en dégager une impression et d'en extraire des informations précises.

Notre travail est basé sur l'appariement entre deux images de visages. La reconnaissance 3D de visages constitue une alternative prometteuse pour surmonter les problèmes de la reconnaissance 2D. Cette technique conserve toutes les informations sur la géométrie de visage, ce qui permet d'avoir une représentation réelle de ce dernier.

La reconstruction 3D des visages repose sur les trois étapes développées dans notre travail:

La détection du visage et pour ce faire on a utilisé la méthode du Viola et Jones, avec cette méthode on a réussi à détecter les visages sur toutes les images qu'on a utilisées.

La détection des points d'intérêt, on a utilisé la méthode de HARRIS, son inconvénient est de donner des points inutiles.

L'appariement d'image, Nous avons présenté la méthode d'appariement entre deux images par corrélation, on voit que cette méthode a donné de très bons résultats.

Pour les travaux à venir il reste à améliorer la méthode de détection des points d'intérêts afin de ne faire détecter que les points caractéristiques du visage.

- [1] C. Kotropoulos and I. Pitas. Rule-Based Face Detection in Frontal Views. Proc. Int'l Conf. Acoustics, Speech and Signal Processing, vol. 4, pp. 2537- 2540, 1997.
- [2] T. Kanade. Picture Processing by Computer Complex and Recognition of Human Faces, PhD thesis, Kyoto Univ., 1973.
- [3] G. Yang and T. S. Huang. Human Face Detection in Complex Background, Pattern Recognition, vol. 27, no. 1, pp. 53-63, 1994.
- [4] P. Sinha, Object Recognition via Image Invariants: A Case Study, Investigative Ophthalmology and Visual Science, vol. 35, no. 4, pp. 1735-1740, 1994.
- [5] P. Sinha, Processing and Recognizing 3D Forms, PhD thesis, Massachusetts Inst. Of Technology, 1995.
- [6] M. Turk and A. Pentland. Eigenfaces for Recognition, J. Cognitive Neuroscience, vol. 3, no. 1, pp. 71-86, 1991.
- [7] H. A. Rowley, S. Baluja, et T. Kanade, Neural Network based Face Detection, IEEE Trans, Pattern Anal. Mach, Intell, 23-38, January 1998.
- [8] R. Féraud, O.J. Bernier, J.E. Viallet, M. Collobert. A Fast and Accurate Face Detector Based on Neural Networks, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 1, pp. 42-53, January, 2001.
- [9] C. Garcia, M. Delakis. Convolutional Face Finder: A Neural Architecture for Fast and Robust Face Detection IEEE Trans. PAMI, vol. 26, no. 11, Nov, 2004.
- [10] G. Bradski, A. Kaehler and V. Pisarevsky, Learning-Based Computer Vision with OpenCV. Intel Technology Journal, May 2005.
- [11] P. Viola, M. Jones. Rapid object detection using a boosted cascade of simple features. In Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 2001.
- [12] L. C. De Silva, K. Aizawa, and M. Hatori, Detection and tracking of facial features by using a facial feature model and deformable circular template, IEICE Trans. Inform. Systems E78–D(9), 1195–1207, 1995.
- [13] K.C. Yow and R. Cipolla, Feature-Based Human Face Detection, Image and Vision Computing, vol. 15, no. 9, pp. 713-735, 1997.
- [14] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In IEEE Conference on Computer Vision and Pattern Recognition, volume 1, pages 511–518, Decembre 2001.
- [15] H. Moravec. Towards automatic visual obstacle avoidance. In Proceedings of the 5th International Joint Conference on Artificial Intelligence, page 584, August 1977.

-
- [16] C. Schmid. Appariement d'images par invariants locaux de niveaux de gris. Doctoral thesis, Institut National Polytechnique de Grenoble, GRAVIR, July 1996.
- [17] P.R. Beaudet. Rotationally invariant image operators. In proceedings of the quatrième International Joint Conference on pattern Recognition, pages 579-583, 1978.
- [18] C. Harris and M. Stephens. A combined corner and edge detector. In Proceedings of The Fourth Alvey Vision Conference, pages 147–151, Manchester, UK, 1988.
- [19] S. M. Smith and J. M. Brady. Susan - a new approach to low level image processing. IEEE Transactions Pattern Analysis Machine Intelligence, 23(1) :45–78, 1997.
- [20] D. Lowe. Object recognition from local scale-invariant features. IEEE International Conference on Computer Vision, pages 1150{1157, 1999.
- [21] T. Lindeberg. Feature detection with automatic scale selection. IEEE Transactions Pattern Analysis Machine Intelligence, 30(2) :77–116, 1998.
- [22] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide baseline stereo from maximally stable extremal regions. British Machine Vision Conference, pages 384{396, 2002.
- [23] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. International Journal of Computer Vision, 1(60) :63{86, 2004.
- [24] R. Hummel and S. Zucker. On the foundations of relaxation labeling processes. IEEE Pattern Analysis and Machine Intelligence, 5(3) :267-287, 1983.
- [25] A. Choksuriwong, H. Laurent, and B. Emile. Object recognition using local caractérisation and zernike moments. Advanced Concepts for Intelligent Vision Système, 3708 :108-115, 2005.
- [26] A. Choksuriwong, H. Laurent, B. Emile, and C. Rosenberger. Comparative study of global invariant descriptors for object recognition. Journal of Electronic imaging, 17(2), 2008.
- [27] M. Hu. Visual pattern recognition by moment invariants. IEEE Transactions Information Theory, 8 :179{187, 1962.
- [28] P. Aschwanden and W. Guggenbuhl. Experimental Results from a Comparative Study on Correlation-Type Registration Algorithms. Robust computer vision, 1992.
- [29] J. Badri, C. Tilmant, J-M. Laveit, Q-C. Pham, and P. Sayd. Camera-to-camera mapping for hybrid pan-tilt-zoom sensors calibration. Lecture Notes in Computer Science, 2007.
- [30] J. Mennesson, C. Saint-Jean, and L. Mascarilla. De nouveaux descripteurs de fourier géométriques pour l'analyse d'images couleur. RFIA, pages 599{606, 2010.

-
- [31] H. Bay, T. Tuytelaars, and L. Van Gool. Surf : Speeded up robust features. European Conference on Computer Vision, pages 404-417, 2006.
- [32] M. Swain and D. Ballard. Color indexing. International Journal of Computer Vision, 7,1991.
- [33] B. Schiele and A. Waibel. Gaze tracking based on face-color. International Workshop on Automatic Face and Gesture Recognition, pages 344-349, 1995.
- [34] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. IEEE Conference on Computer Vision and Pattern Recognition, pages 886-893, 2005.
- [35] D. Marr and T. Poggio. Cooperative computation of stereo disparity. Science, 194:283-287, 1976.
- [36] D. Marr and T. Poggio. A computational theory of human stereo vision. Proceedings of the Royal Society of London, B 204:301-328, 1979.
- [37] W.E.L. Grimson. A computer implementation of the theory of human stereo vision. Philosophical Transactions of the Royal Society of London, B292(1058):217-253, 1981.
- [38] O. Faugeras. Three-Dimensional Computer Vision - A Geometric Viewpoint. Artificial intelligence. The MIT Press, Cambridge, MA, USA, Cambridge, MA, 1993.
- [39] D.N. Bhat-S.K. Nayar - Ordinal measures for visual correspondence. - IEEE Transactions on Pattern Analysis and Machine Intelligence - Num 20, 1998, pages. 415-423.
- [40] C. Tomasi-R. Manduchi - Stereo matching as a nearest-neighbor problem - IEEE Transactions on Pattern Analysis and Machine Intelligence, Num 20, 1998, pp 330-340
- [41] D. Scharstein-R. Szeliski - Stereo matching with non-linear diffusion - Int'l Joint Conf. Artificial Intelligence, Num 28, 1998, pages 155-174
- [42] D. Scharstein et R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision IJCV, 47(1):7- 42, avril 2002.
- [43] K.-J. Yoon et I.-S. Kweon. Locally Adaptive Support-Weight Approach for Visual Correspondence Search. Dans IEEE Conference Proceedings of Computer Vision and Pattern Recognition, CVPR, volume 2, pages 924-931, San Diego, Etats-Unis, juin 2005.