

الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique et Populaire
وزارة التعليم العالي و البحث العلمي
Ministère de l'enseignement Supérieur et de la Recherche scientifique



Université Mohamed Khider Biskra
Faculté des Sciences et de la Technologie
Département de Génie Electrique
Filière : Télécommunications
Option : Réseaux et Télécommunication

Réf:.....

Mémoire de Fin d'Etudes
En vue de l'obtention du diplôme:

MASTER

Thème

Localisation Indoor Par Vision Artificielle

Présenté par :

AMMARI Abd El Hafid

Soutenu le : 23 Juin 2018

Devant le jury composé de :

M^r BOUKERDINE Salah
M^{me} BELAHCENE Mébarka
M^{me} MEDOUAKH Sâadia

Univ. Biskra
Univ. Biskra
Univ. Biskra

Président
Encadreure
Examinatrice

Année universitaire : 2017 / 2018

DEDICACES

À MA TRÈS CHÈRE MÈRE

AUTANT DE PHRASES AUSSI EXPRESSIVES SOIENT-ELLES NE SAURAIENT MONTRER LE DEGRÉ D'AMOUR ET D'AFFECTION QUE J'ÉPROUVE POUR TOI. TU M'AS COMBLÉ AVEC TA TENDRESSE ET AFFECTION TOUT AU LONG DE MON PARCOURS. TU N'AS CESSÉ DE ME SOUTENIR ET DE M'ENCOURAGER DURANT TOUTES LES ANNÉES DE MES ÉTUDES, TU AS TOUJOURS ÉTÉ PRÉSENTE À MES CÔTÉS POUR ME CONSOLER QUAND IL FALLAIT. EN CE JOUR MEMORABLE, POUR MOI AINSI QUE POUR TOI, REÇOIS CE TRAVAIL EN SIGNE DE MA VIVE RECONNAISSANCE ET MON PROFOND ESTIME. PUISSE LE TOUT PUISSANT TE DONNER SANTÉ, BONHEUR ET LONGUE VIE AFIN QUE JE PUISSE TE COMBLER À MON TOUR.

Remerciements

Je tiens à exprimer toute ma reconnaissance à mon enseignante et encadreuse Madame BELAHCENE Mébarka. Je la remercie de m'avoir encadrée, orientée, aidée et conseillée.

J'adresse mes sincères remerciements à tous les enseignants, intervenants et toutes les personnes qui par leurs paroles, leurs écrits, leurs conseils et leurs critiques ont guidé mes réflexions et ont accepté à me rencontrer et répondre à mes questions durant mon projet de fin d'étude.

Je remercie très spécialement M^r ZOUBIRJ Bilal

Je remercie mes très chers parents, qui ont toujours été là pour moi, « Vous avez tous sacrifié pour vos enfants n'épargnant ni santé ni efforts. Vous m'avez donné un magnifique modèle de labeur et de persévérance. Je suis redevable d'une éducation dont je suis fier ».

Je remercie mes frères pour leurs encouragements.

Je remercie très spécialement qui ont toujours été là pour moi ;

Enfin, je remercie tous mes Ami(e)s

À tous ceux-ci, je vous présente mes Remerciements, mon Respect et ma Gratitude

Liste des Figures

FIGURE 1. 1 UN EXEMPLE DE STRUCTURE DE DETECTION INTERIEUR ET EXTERIEUR [GOOGLE].....	15
FIGURE 1. 2 EXEMPLE DE DETECTION DE VISAGE [GOOGLE].	16
FIGURE 1. 3 EXEMPLE D'APPLICATION GPS SUR TELEPHONE MOBILE[GOOGLE].....	19
FIGURE 1. 4 LOCALISATION PAR SATELLITE [GOOGLE]	20
FIGURE 1. 5 CAMERA IP [GOOGLE].....	21
FIGURE 1. 6 IMAGE DE DETECTION ET LOCALISATION DE PERSONNES [GOOGLE].....	21
FIGURE 1. 7 EXEMPLE D'ARCHITECTURE DE LOCALISATION PAR GSM [GOOGLE]	23
FIGURE 1. 8 EXEMPLE DE LOCALISATION PAR ADRESSE IP [GOOGLE]	24
FIGURE 1. 9 EXEMPLE DE LOCALISATION PAR CHAMP MAGNETIQUE [12]	27
FIGURE 1. 10 EXEMPLE ILLUSTRANT LE TRAVAIL D'UNE TECHNIQUE LOCALISATION PAR CHAMP MAGNETIQUE [12]	28
FIGURE 1. 11 ARCHITECTURE DU SUIVEUR VISUEL ROBUSTE [15]	29
FIGURE 2. 1 ENVIRONNEMENT SYSTEME POUR DES APPLICATIONS EN VIDEO SURVEILLANCE [GOOGLE].....	32
FIGURE 2. 2 DIAGRAMME BLOCS DU SYSTEME [16].....	32
FIGURE 2. 2. 1 COURBE SYMETRIQUE DE GAUSSIEN.....	32
FIGURE 2. 2. 2 LE MODELE DE MELANGE DE MIXTURE MODEL	32
FIGURE 2. 3 ORGANIGRAMME DE L'EXEMPLE SYSTEME DE DETECTION DE VOITURE [16]	34
FIGURE 2. 4 ROI DE SEGMENTATION. (A) IMAGE ORIGINALE, (B) REGION D'INTERET[16].....	35
FIGURE 2. 5 PROCESSUS MORPHOLOGIQUE POUR LA DETECTION DE PREMIER PLAN.....	36
FIGURE 2. 6 BOÎTE DE DÉLIMITATION[26]	36
FIGURE 2. 7 DETECTION D'OBJET [26]	37
FIGURE 2. 8 VEHICULE DETECTE DANS CHAQUE CADRE [26]	38
FIGURE 2. 9 EXEMPLE DES TYPES DE CARACTERISTIQUES UTILISEES PAR VIOLA ET JONES.....	40
FIGURE 2. 10 ILLUSTRATION DE L'ARCHITECTURE DE LA CASCADE.	41
FIGURE 2. 11 EXTENSION DES CARACTERISTIQUES PSEUDO-HAAR PROPOSEE PAR LIENHART	43
FIGURE 2. 12 DETECTION DE VISAGE AVEC LA METHODE DE VIOLA ET JONES	45
FIGURE 2. 13 EXEMPLE D'IMAGE D'UNE GRANDE COLLECTION DETECTEE PAR VIOLA JONES.....	46
FIGURE 2. 14 DETECTION DE VISAGE DE PLUSIEURS PERSONNES FRONTALES OUTDOOR	46
FIGURE 2. 15 DETECTION DE VISAGE DE PLUSIEURS PERSONNES FRONTALES INDOOR	46
FIGURE 2. 16 DETECTION DE VISAGE DE PLUSIEURS PERSONNES FRONTALES INDOOR A (LIEU DE TRAVAIL) ET B (UNE CLASSE).....	47
FIGURE 2. 17 DETECTION DE PERSONNE FRONTALE INDOOR	47
FIGURE 2. 18 DETECTION ET SUIVI DU VISAGE EN UTILISANT LE NUAGE DE POINTS CARACTERISTIQUES.....	48
FIGURE 2. 19 DETECTION DE DEUX/QUATRE PERSONNES INDOOR SUR LES DEUX IMAGES A ET B(LIEU DE TRAVAIL)	48
FIGURE 2. 20 DETECTION DE VISAGE CAS DE PERSONNES EN ACTION	49
FIGURE 2. 21 EXEMPLES DE DÉTECTIONS.....	50
FIGURE 2. 22 SYNOPTIQUE D'UN SYSTEME USUEL DE SUIVI PAR DETECTION[30].	51
FIGURE 2. 23 DETECTION PAR METHODE ACF (LUMIERE SOMBRE).....	52

FIGURE 2. 24 DETECTION PAR METHODE ACF (VARIANTE LUMIERE ET DISTANCE)	52
FIGURE 2. 25 DETECTION PAR METHODE ACF(VARIANTE DISTANCE ET LUMIERE)	52
FIGURE 2. 26 DETECTION PAR LA METHODE ACF UTILISATION CAM MOBILE	53
FIGURE 2. 27 DETECTION D'IMAGE STEREO DANS DIFFERENTES POSES.....	53
FIGURE 2. 28 DETECTION D'IMAGE STEREO AVEC ORIENTATION DE LA TETE	53
FIGURE 2. 29 DETECTION D'UNE PERSONNE INDOOR AVEC DIFFERENTES VARIANTES	57
FIGURE 2. 30 DETECTION EN PRESENCE DES VARIANTES LUMIERE, ROTATION DE LA TETE, PROFIL, DISTANCE. ...	58
FIGURE 2. 31 DETECTION D'UNE PERSONNE EN MOUVEMENT INDOOR AVEC DIFFERENTES VARIANTES(LUMIERE, PROFIL, DISTANCE)	58
FIGURE 3. 1 WEBCAM UTILISEE SES CARACTERISTIQUES.....	64
FIGURE 3. 2 LOGO D'APPLICATION	65
FIGURE 3. 3 UTILISATION DE L'APPLICATION IP WEBCAM [36]	65
FIGURE 3. 4 PHOTO DU PC ET SES CARACTERISTIQUES	66
FIGURE 3. 5 IMAGE MONTRANT LE CALCUL DE DIMENSION POUR LA CALIBRATION	66
FIGURE 3. 6 INSTALLATION ET MISE EN ŒUVRE DES CAMERAS BINOCULAIRES SUR PC	67
FIGURE 3. 7 PHOTOS DES DEUX SMARTPHONES UTILISES.....	67
FIGURE 3. 8 INSTALLATION ET MISE EN OEUVRE DE L'ACQUISITION PAR SMARTPHONE.....	68
FIGURE 3. 9 IMAGE MONTRANT LES ETAPES DE CALIBRATION	69
FIGURE 3. 10 CHOIX DE LA FORME DU MOTIF EN DAMIER (CARRE)	70
FIGURE 3. 11 CHOIX DES PRIMITIVES DU MOTIF EN DAMIER	70
FIGURE 3. 12 DAMIER A UN ANGLE INFERIEUR A 45 DEGRES PAR RAPPORT AU PLAN DE LA CAMERA.	71
FIGURE 3. 13 STATUT DES IMAGES PRISES POUR CALIBRATION	72
FIGURE 3. 14 REGLAGE DES PODITIONS DE LA PAIRE DE CAMERAS	72
FIGURE 3. 15 DEFINITION DE LA TAILLE DE LA BOITE D'ECHECS	73
FIGURE 3. 16 SERIE D'IMAGES ACQUISES POUR LA CALIBRATION (CAMERA GAUCHE)	73
FIGURE 3. 17 SERIE D'IMAGES ACQUISES POUR LA CALIBRATION (CAMERA DROITE).....	73
FIGURE 3. 18 EXEMPLE DE TEST SUR LA CALIBRATION	73
FIGURE 3. 19 AFFICHAGE DES IMAGES AVEC DES POINTS DETECTES	75
FIGURE 3. 20 SELECTION DES IMAGES PRISES A PARTIR DES CAMERAS 1 ET 2	76
FIGURE 3. 21 ÉVALUATION DES RESULTATS DE CALIBRATION	77
FIGURE 3. 22 VERIFICATION DES ERREURS DE PHOTO.....	77
FIGURE 3. 23 GRAPHIQUE A BARRES DES ERREURS DE PROJECTION	78
FIGURE 3. 24 REPRESENTATION EN 3D DE LA POSITION DES CAMERAS DES IMAGES ACQUISES EN REPRESENTATION (A) HORIZONTALE ET (B) VARTICALE	79
FIGURE 3. 25 AFFICHAGE DES IMAGES RECTIFIÉES	80
FIGURE 3. 26 ILLUSTRATION DU REGLAGE DU DAMIER POUR LA CALIBRATION	81
FIGURE 3. 27 RESULTAT DE CALIBRATION DE LA CAMERA.....	82
FIGURE 3. 28 LECTURE ET RECTIFICATION LES IMAGES VIDEO ET RECONSTRUCTION SCENE 3D	83
FIGURE 3. 29 CARTE DE DISPARITÉ	84
FIGURE 3. 30 RECONSTRUCTION SCÈNE 3D	84
FIGURE 3. 31 LOCALISATION DE LA 1 ^{ERE} SÉQUENCE VIDÉO.....	85
FIGURE 3. 32 EXEMPLE DE DÉTECTION, LOCALISATION ET AVERTISSEMENT.....	85

LISTE DES ABRÉVIATIONS

BSD	Berkeley Software Distribution
RFID	Radio Frequency Identification
Wi-Fi	Wireless Fidelity
GPS	Global Positioning System
GSM	Global System for Mobile Communication
UMTS	Universal Mobile Telecommunications System
ULB	Ultra Large Bande
SNR	Signal-To-Noise Ratio
EOTD	ECCMA Open Technical Dictionary
IP	Internet Protocol
API	Application Programming Interface
GMM	Gaussian Mixture Models
ROI	Region Of Interest
FPGA	Field-Programmable Gate Array
ASIC	Application-Specific Integrated Circuit
GPU	Graphics Processing Unit
ACF	Aggregate Channel Features
HOG	HistogramOrientedGradient
INRIA	Institut national De Recherche En Informatique Et en Automatique
RGB	Red Green And Blue
SVM	Support Vector Machine
IMH	Internal Motion Histograms
3D	Three-Dimensional
DVR	Digital Video Recorder
NVR	Network Video Recorder
ADSL	Asymmetric Digital Subscriber Line
USB	Universal Serial Bus

Tableaux des Matières

DEDICACES	I
REMERCIEMENTS	II
LISTE DES FIGURES	III
LISTE DES ABREVIATIONS	V
TABLEAUX DE MATIERE	VI
RESUMES	IX
INTRODUCTION GENERALE	10
CHAPITRE 1 DETECTION & LOCALISATION INDOOR	13
INTRODUCTION	14
PARTIE A LA DETECTION	15
<i>A.1 Méthodes de détections réparties en catégories</i>	15
PARTIE B LA LOCALISATION INDOOR	18
<i>B.1 Géolocalisation Indoor</i>	18
B.1.1 Contexte	18
B.1.2 Quelques applications de géolocalisation	18
B.1.3 Technologies de géolocalisation	19
B.1.4 Le système GPS en indoor	19
B.1.5 Les systèmes d'acquisition vidéo	21
B.1.6 Methodes de localisation indoor	21
CONCLUSION	30
CHAPITRE 2 : ETUDE ET CONCEPTION DE LA DETECTION	31
INTRODUCTION	32
2.1 METHODE GAUSSIAN MIXTURE MODELS (GMM)	32
2.1.1 Principe du GMM	33
2.1.2 Implémentation du détecteur GMM et application à une séquence véhicule	34
2.1.2.1 Extract Frame	34
2.1.2.2 ROI Segmentation	35
2.1.2.3 Gaussian Mixture Model (GMM)	35
2.1.2.4 DETECTION DE VOITURE	36
2.1.3 Expérience et Résultats sur la détection GMM	37
2.2 METHODE DE DETECTION VIOLA ET JONES	39
2.2.1 Éléments de la méthode	39
2.2.1.1 Caractéristiques	40
2.2.2 Étapes clés	42
2.2.2.1 Apprentissage :	42
2.2.2.2 Détection	42
2.2.3 Performances de la methode viola jones	43
2.2.4 Limitations et extensions de la methode viola jones	43

2.2.5 Applications la méthode de Viola et Jones.....	44
2.2.6 Notre implémentation du détecteur Viola Jones.....	45
Discussion.....	49
2.3 MÉTHODE AGGREGATE CHANNEL FEATURES (ACF).....	50
2.3.1 Détecteur ACF ajusté en tant que région	50
2.3.2 Synoptique général.....	51
2.3.3 Notre implémentation de la méthode ACF et Résultats.....	52
2.3.3.1 Détection indoor avec mono caméra (webcam)	52
2.3.3.2 Détection d'image stéréo (avec caméra binoculaire).....	53
Discussion :	53
2.4 METHODES DE DETECTION PAR HOG (HISTOGRAMME DE GRADIENTS ORIENTES)	54
2.4.1 Description générale de la méthode	54
2.4.2 Construction du descripteur	54
2.4.2.1 Calcul du gradient.....	54
2.4.2.2 Construction de l'histogramme.....	55
2.4.2.3 Formation des blocs	55
2.4.2.4 Normalisation des blocs	55
2.4.3 Classification.....	56
2.4.4 Extensions.....	56
2.4.5 Expériences et résultats de la détection par la méthode HOG.....	57
2.4.5.1 Détection avec mono caméra	57
Discussion.....	57
2.4.5.2 Détection avec caméra binoculaire.....	58
Discussion.....	58
CONCLUSION	59
CHAPITRE 3 : ETUDE ET CONCEPTION DE CALIBRATION ET LOCALISATION	60
INTRODUCTION	61
3.1 ETUDE DE LA CALIBRATION DES IMAGE	62
3.2 ETUDE DE LA LOCALISATION INDOOR	62
3.3 ACQUISITION DES IMAGES	63
3.3.1 Acquisition par webcam	63
3.3.2 Acquisition par caméra IP.....	64
3.3.2.1 Application utilisée de la caméra IP	65
3.3.2.2 Installation du toolbox caméras IP dans MATLAB	66
3.3.2.3 Photo du PC et ses caractéristiques.....	66
3.3.3 Acquisition stéréo binoculaire	66
3.3.3.1 Equation de disparité de base.....	66
3.3.4 Acquisition par smartphones.....	67
3.3.4.1 Photo et caractéristiques des smartphones utilisés	67
3.3.4.2 Installation et mise en œuvre de l'acquisition par smartphones	68
3.4 APPLICATION DE CALIBREUR DE CAMERA STEREO	68
3.4.1 Présentation du calibreur de caméra stéréo.....	68
3.4.1.1 Calibration de la Stéréo Caméra	69

3.4.1.2 Ouverture du calibrateur de caméra stéréo	69
3.4.1.3 Préparation du motif, une caméra et des images :.....	69
3.4.1.4 choix de la du motif en damier :	70
3.4.1.5 choix des primitives du motif en damier.....	70
3.4.1.6 Configuration de la caméra	71
3.4.1.7 Capture des images	71
3.4.2 Ajout des paires d'images	72
3.4.2.1 Chargement des images	73
3.4.2.2 Analyse des images	74
3.4.2.3 Visualisation des images et points détectés	75
3.4.2.4 Visualisation de la carte intrinsics (MATLAB).....	75
3.4.3 Mise en œuvre de la Calibration	76
3.4.4 Évaluation des résultats de calibration	77
3.4.4.1 Examen des erreurs de reprojection	77
3.4.4.2 Erreurs de reprojection (graphique a barres)	78
3.4.4.3 Visualisation des paramètres extrinseques	78
3.4.4.4 Affichage des images rectifiées	80
3.4.5 Amélioration de la calibration.....	80
3.4.6 Exporter les paramètres de la caméra	81
3.4.6.1 Exportation des paramètres de la caméra.....	81
3.4.6.2 Génération d'un script MATLAB.....	81
3.5 LOCALISATION	82
3.5.1 Chargement des paramètres de la caméra stéréo.....	82
3.5.2 Création des lecteurs de fichiers vidéo et le lecteur vidéo	82
3.5.3 Lecture et rectification d'images vidéo	83
3.5.4 Calcul de la disparité	83
3.5.5 Reconstruction de la scène 3D.....	84
3.5.6 Détection et avertissement	85
Resultats finaux sur séquence video.....	86
CONCLUSION	92
CONCLUSION GENERAL	93
REFERENCES BIBLIOGRAPHIQUE.....	97

Résumé

La démocratisation des terminaux mobiles telle que les téléphones cellulaires, les tablettes et les PC a rendu possible le déploiement de la réalité augmentée dans des environnements en extérieur à grande échelle. Cependant, afin de mettre en œuvre de tels systèmes, différentes problématiques doivent être traitées. Parmi elle, la localisation représente l'une des plus importantes. En effet, l'estimation de la position et de l'orientation (appelée pose) du point de vue (de la caméra ou de l'utilisateur) permet de recalibrer les objets virtuels sur les parties observées de la scène réelle. Dans nos travaux, nous présentons un système de localisation original destiné à des environnements à grande échelle qui utilise une approche basée vision sans marqueur pour l'estimation de la position de la personne de la caméra en indoor. Cette approche se base sur des points caractéristiques naturels extraits des images. Etant donné que ce type d'approche est sensible aux variations de luminosité, aux occultations et aux mouvements brusques de la caméra, qui sont susceptibles de survenir dans l'environnement intérieur, nous utilisons deux autres types de capteurs afin d'assister le processus de vision. Dans nos travaux, nous voulons démontrer la faisabilité d'un schéma de suppléance dans des environnements intérieurs indoor à large échelle en milieux incontrôlés. Le but est de fournir un système de détection et localisation indoor basé uniquement sur des algorithmes de traitement d'image permettant également de nous avertir en cas d'une distance limite d'une porte où autre objet à télésurveiller et déclencher une alarme. Le système de localisation vise à être autonome et adaptable aux différentes situations rencontrées.

ملخص

جعل الديمقراطية من الأجهزة المحمولة مثل الهواتف المحمولة والأجهزة اللوحية وأجهزة الكمبيوتر من الممكن نشر الواقع المعزز في بيئات خارجية واسعة النطاق. ومع ذلك، من أجل تنفيذ مثل هذه الأنظمة، يجب معالجة القضايا المختلفة. من بينها، الموقع هو واحد من أهمها. وبالفعل، فإن تقدير الموضع والاتجاه (الذي يطلق عليه بوز) من وجهة النظر (للكاميرا أو المستخدم) يجعل من الممكن إعادة معايرة الكائنات الافتراضية على الأجزاء المرصودة من المشهد الحقيقي. في عملنا، نقدم نظام الموقع الأصلي للبيئات واسعة النطاق التي تستخدم نهجاً قائماً على الرؤية بدون تحديد لتقدير الموقع الداخلي للكاميرا. ويستند هذا النهج على النقاط المميزة الطبيعية المستخرجة من الصور. نظراً لأن هذا النوع من الأساليب حساس للتغيرات في السطوح والاضطرابات والحركات المفاجئة للكاميرا، والتي من المرجح أن تحدث في البيئة الداخلية، فإننا نستخدم نوعين آخرين من أجهزة الاستشعار للمساعدة في عملية الرؤية. في عملنا، نريد أن نبرهن على جدوى مخطط الاستبدال في البيئات الداخلية المغلقة الكبيرة في البيئات غير المتحكم فيها. الهدف هو توفير نظام كشف وتوطين في الأماكن المغلقة يعتمد فقط على خوارزميات معالجة الصور التي يمكن أن تحذرنا أيضاً في حالة وجود مسافة محددة للباب أو أي كائن آخر لمراقبة عن بعد وإطلاق إنذار. يهدف نظام الموقع إلى أن يكون مستقلاً وقابلاً للتكيف مع الحالات المختلفة التي تمت مواجهتها.

Introduction générale

Il est facile pour un humain de reconnaître un autre homme autour de lui ou sur une image mais c'est un problème très complexe pour un système automatisé. Pourtant, beaucoup de systèmes ont besoin d'avoir des informations sur la présence ou l'absence de personnes dans leur environnement. Les applications potentielles d'un système de détection automatique de la présence humaine sont nombreuses. Nous pouvons par exemple citer les systèmes de transport intelligent, la robotique, la surveillance, la domotique intelligente, l'indexation d'images ou de vidéos . . . D'une manière générale, le principe de la détection de personnes dans une image ou une vidéo est le même. Cependant, pour une image, le détecteur ne dispose d'aucune information *a priori* et doit donc parcourir l'image entière avec une fenêtre de détection. Pour une vidéo, il est possible de simplifier le problème en se focalisant par exemple sur les zones de l'image où il y a eu un mouvement. Il est également possible de raisonner en fonction de la position des personnes présentes dans les images précédentes ou même de caractériser les personnes par leur mouvement. On trouve principalement deux groupes de méthodes dans la littérature pour détecter un humain dans une image. Il y a d'une part les méthodes qui utilisent un modèle explicite (2D ou 3D) de la forme du corps humain et d'autre part celles qui se basent sur des techniques d'apprentissage supervisé. À partir d'une base d'images, des caractéristiques de la forme du corps humain sont extraites et un modèle discriminant

est construit. Nous pouvons citer par exemple Papageorgiou *et al.* qui ont proposé un détecteur basé sur les ondelettes de Haar et les machines à vecteurs de support. Viola et Jones ont également proposé un détecteur basé sur les filtres de Haar et l'algorithme du boosting. Plus récemment, Dalal et Triggs ont utilisé avec succès les histogrammes de gradients orientés dans le cas de la détection de piétons. D'autres méthodes détectent séparément différentes parties du corps humain et fusionnent ensuite ces résultats.

Nos travaux s'inscrivent dans le cadre du projet de fin d'étude master 2 intitulé « localisation indoor ». Pour arriver à réaliser une bonne localisation d'un objet ou d'une personne, nous avons besoin de trois fonctionnalités efficaces : 1) une bonne détection ; 2) une bonne calibration (étalonnage de la caméra) et enfin 3) une bonne méthode de localisation.

L'étalonnage de la caméra est nécessaire pour relier les données de la caméra au monde physique et il est également important dans certains problèmes de vision traitables.

Introduction générale

La vision système, qui utilise la représentation basée sur les bords et la correspondance stéréo, repose sur l'étalonnage à la fois détermination des épi-polaires pour le processus d'appariement et le calcul de la position 3D à partir de la disparité. Dans le passé, les approches de l'étalonnage de la caméra se sont concentrées sur l'efficacité de calcul et le choix de modèle de caméra (par exemple: Tsai 1987). Souvent, la recherche de l'efficacité computationnelle a conduit à des solutions qui compromettent les exigences du problème original et entraînent des inexactitudes et des préjugés. Avec le développement continu de la technologie informatique, l'efficacité informatique devient de moins en moins importante dans la solution de ces problèmes. Sans l'exigence de la vitesse algorithmique de nombreux critères de jugement des algorithmes d'étalonnage vont simplement disparaître. Il nous reste la question: de quoi avons-nous besoin pour utiliser des algorithmes stéréoscopiques de vision par ordinateur dans un environnement réel?

L'étalonnage à cet égard est une procédure par laquelle un stimulus d'étalonnage est placé devant les caméras et les positions imagées des fonctions sur cet objet sont utilisées pour récupérer le modèle de caméra paramètres. Ce stimulus d'étalonnage est souvent un objet spécialement construit qui ne serait généralement pas visible au système en utilisation normale. Les paramètres d'étalonnage sont récupérés une seule fois sans souci de mise à jour de ce calibre à l'avenir par tout autre moyen autre que le remplacement total. Dans un système de vision pratique qui doit être en usage continue et peut-être avoir une certaine liberté de circulation, de telles méthodes d'étalonnage uniques sont presque inutiles. Pour les caméras fixes, le flux normal d'entrée doit être interrompu périodiquement pour être ré-étalonné. Afin d'avoir un système de vision stéréo pratique, le réétalonnage doit être une fonctionnalité intégrée fonctionnant avec les données disponibles. Idéalement, cette méthode peut également intégrer des informations provenant de différentes sources et pouvant être décrites comme "en ligne".

Avec les puissances de calcul actuelles, il est possible de construire un système de suivi visuel à une fréquence vidéo qui peut suivre de multiples cibles en utilisant simplement des processus de détection au niveau pixel. Les processus de détection au niveau pixel peuvent être définis en utilisant des algorithmes tels que la soustraction de fond adaptative, la différence d'image, et des tableaux d'indexation de couleurs. Le suivi permet de restreindre la détection aux régions où les cibles sont susceptibles d'apparaître, ce qui réduit considérablement les besoins en puissance de calcul. Le calcul robuste en temps réel peut

Introduction générale

être assuré par une régulation dynamique des paramètres de détection, incluant la résolution des cibles ainsi que le nombre de cibles suivies.

Le suivi visuel est un processus cyclique d'estimation récursif. De même que pour d'autres technologies d'estimation de localisation, le filtre de Kalman fournit un cadre classique pour la conception du suivi visuel. Le suivi conserve temporellement les informations sur les cibles, permettant une cohérence des objets. Cette cohérence des objets assure qu'un label affecté à une cible à un instant T1 peut être utilisé à un instant T2. Le suivi permet également de focaliser l'attention du système en appliquant le processus de détection approprié uniquement dans les régions de l'image où une cible sera potentiellement détectée. La position ainsi que la vitesse fournie par le suivi peuvent être très importantes afin de décrire des situations.

Les systèmes de suivi visuel sont généralement composés de quatre phases: la prédiction, l'observation, la détection et la mise à jour. La phase de prédiction actualise les attributs précédemment estimés pour un ensemble d'entités à une valeur estimée pour un instant donné. La phase d'observation applique la prédiction à la donnée courante afin d'actualiser l'état de chaque cible. La phase de détection détecte de nouvelles cibles. Enfin, la phase de mise à jour actualise la liste des cibles afin de tenir compte des nouvelles cibles ainsi que de celles qui ont disparu.

Notre document est structuré en **trois chapitres** :

Le **chapitre 1** présente des rappels, des généralités et un état de l'art sur la détection et localisation indoor.

Ensuite, nous consacrons le **chapitre 2** sur l'étude et conception de la détection ;

Dans le **chapitre 3** nous le présentons sous deux grandes parties : 1) **Partie A** : étude et conception de la calibration et 2) **Partie B** sur l'étude et la conception de la localisation .

Le mémoire se termine par une conclusion générale où nous résumons l'essentiel des expériences menées, les résultats les plus pertinents obtenus et les perspectives futures.

Chapitre 1

Detection & Localisation Indoor

Introduction

Dans ce présent nous présentons tout d'abord des généralités sur la détection et la localisation. Ensuite nous rappelons brièvement quelques méthodes utilisées dans la détection et aussi la localisation indoor qui fait objet de notre travail. Nous nous intéressons beaucoup plus à la détection/localisation des personnes (visages) et véhicules plus qu'autre chose.

La détection de visage dans l'image est un traitement indispensable et crucial avant la phase de reconnaissance. En effet, le processus de reconnaissance de visages ne pourra jamais devenir intégralement automatique s'il n'a pas été précédé par une étape de détection efficace. Le traitement consiste à rechercher dans une image la position des visages et de les extraire sous la forme d'un ensemble d'images dans le but de faciliter leur traitement ultérieur. Un visage est considéré correctement détecté si la taille d'image extraite ne dépasse pas 20% de la taille réelle de la région faciale, et qu'elle contient essentiellement les yeux, le nez et la bouche [1,2].

L'intérêt de la localisation faciale va au-delà de l'application de ce présent mémoire. Quelques appareils photos numériques récents emploient la détection de visages pour la mise au point automatique. Elle est également recherchée dans le domaine des économies d'énergie ; les télévisions et les ordinateurs peuvent économiser l'énergie en réduisant l'éclat. Le système peut identifier la direction de visage de l'utilisateur de TV. Quand l'utilisateur ne regarde pas l'écran, l'éclat de TV est abaissé, et quand le visage revient à l'écran, l'éclat est augmenté [3]. Les premières difficultés rencontrées par les méthodes de détections de visages sont les variations de pose (vue de profil ou de face), d'expression, de rotation du visage, d'âge et d'illumination. Ce dernier type de difficulté pouvant être surmonté par un prétraitement de normalisation et de compensation de l'illumination, ce sera présenté à la fin de ce chapitre.

La localisation permet à l'utilisateur de choisir une zone où il accepte d'être localisé. Cette zone est donc représentée par un périmètre virtuel dans lequel chaque sortie et entrée est notifiée. C'est le fait d'établir des barrières virtuelles à la façon d'un champ de bétail, et d'être notifié chaque fois qu'un utilisateur rentre dans cette zone. Les entreprises peuvent utiliser ce type de géolocalisation dans leur stratégie mobile afin de créer des lieux

« conscients » et ainsi avoir des interactions beaucoup plus pertinentes avec leurs utilisateurs. La localisation indoor est devenue très importante dans le marketing géolocalisé car elle permet d'assister le consommateur de son domicile au centre commercial.

Elle le guide à l'intérieur du lieu, déclenche l'envoi de messages et lui permet de régler ses achats à partir de son mobile. Un des objectifs principaux de cette nouvelle méthode de marketing est d'éviter la fuite des consommateurs vers des sites e-commerce. L'intérêt de la localisation indoor ne se limite pas au marketing. De nombreux aéroports, gares, centres de congrès et d'exposition s'équipent de ces technologies pour aider les utilisateurs en leur fournissant des informations utiles. Nous consacrons deux parties distinctes pour présenter la détection et la localisation.

Partie A : La détection

En vision par ordinateur on désigne par détection d'objet (ou classification d'objet) une méthode permettant de détecter la présence d'une instance (reconnaissance d'objet) ou d'une classe d'objets dans une image numérique. Une attention particulière est portée à la détection de visage et la détection de personne. Ces méthodes font souvent appel à l'apprentissage supervisé et ont des applications dans de multiples domaines, tels la recherche d'image par le contenu ou la vidéo surveillance.

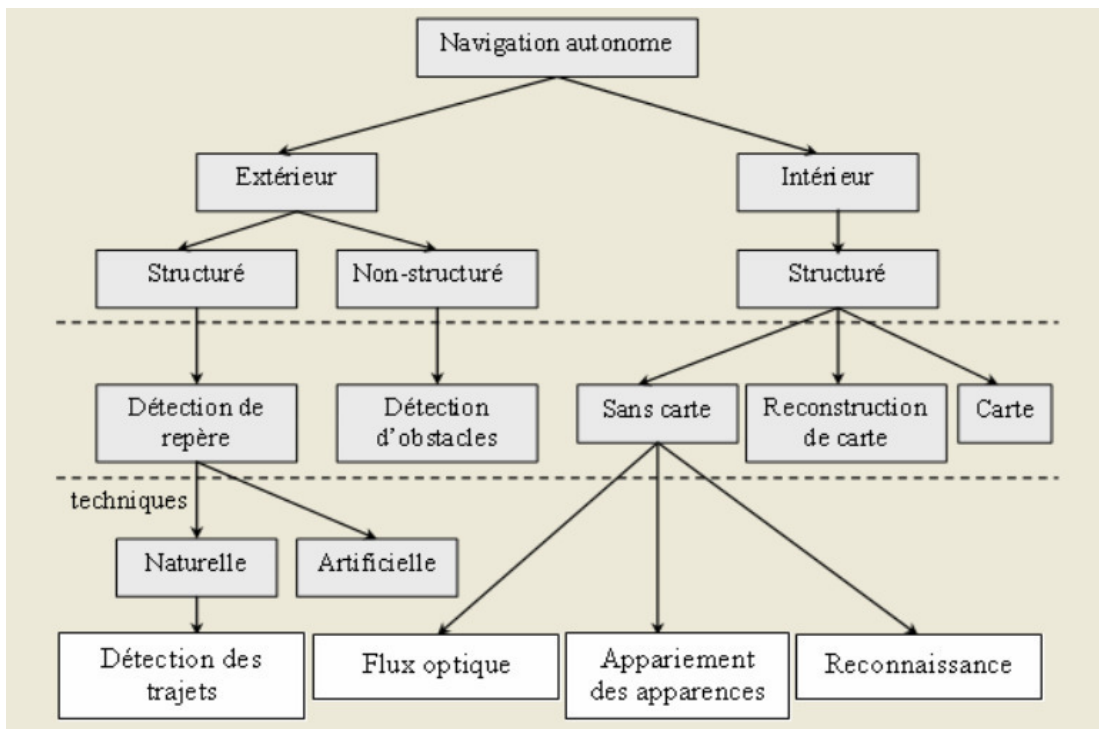


Figure 1.1 Exemple de structure de détection intérieur et extérieur [google]

A.1 Méthodes de détections réparties en catégories

Une classification des méthodes de localisation faciale a été proposée par Yang *et al*[2]. Les méthodes sont divisées en quatre catégories. Ces catégories peuvent se chevaucher

si un algorithme peut appartenir à deux ou plusieurs catégories. Cette classification peut être faite comme suit :

- **Knowledge-based methods** : Ces méthodes se basent sur la connaissance des différents éléments qui constituent un visage et des relations qui existent entre eux. Ainsi, les positions relatives de différents éléments clés tels que la bouche, le nez et les yeux sont mesurées pour servir ensuite à la classification 'visage' ou 'non visage' chez Chiang *et al.* [4]. Le problème dans ce type de méthode est qu'il est difficile de bien définir de manière unique un visage. Si la définition est trop détaillée, certains visages seront ratés tandis que si la description est trop générale, le taux de faux positifs montera en flèche.

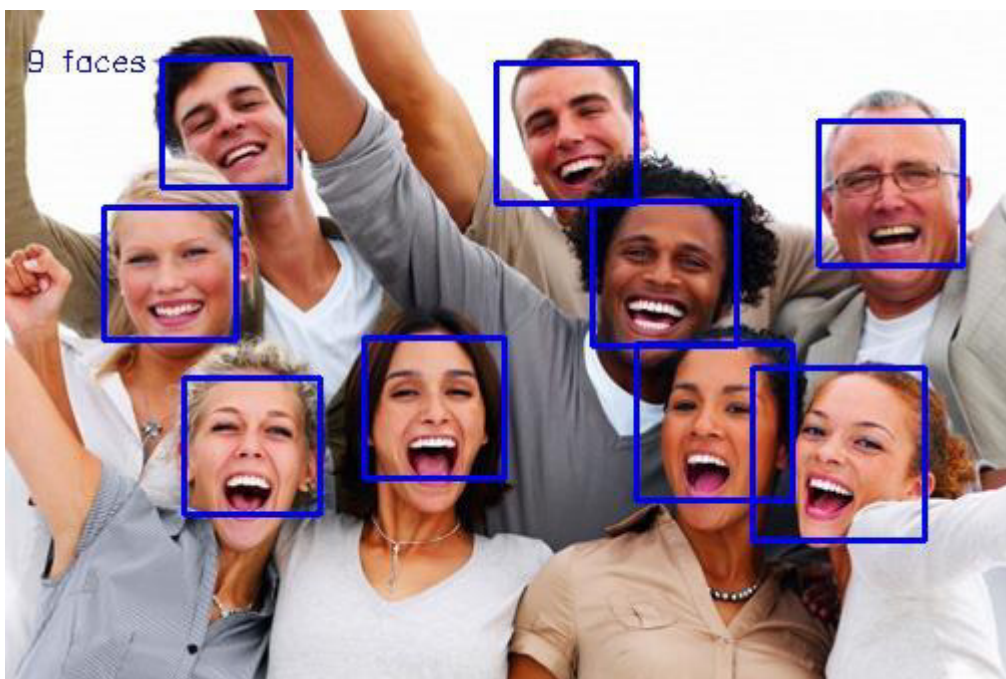


Figure 1. 2 Exemple de détection de visage [Google].

- **Feature invariant approaches** : Ces approches utilisent les éléments invariants aux variations d'illumination, d'orientation ou d'expression tels que la texture ou la signature de couleur de la peau pour la détection.
- **Template matching methods** : Des modèles caractéristiques d'un visage entier ou de sous-partie de visage (bouche, oeil, nez) sont créés. La localisation se fait ensuite sur base de la corrélation de ces modèles avec les candidats [9].
- **Appearance-based methods** : Ces méthodes utilisent le même principe que présenté au point précédent mais se basent sur des modèles appris à partir d'un ensemble d'essai. Ces méthodes présentent l'avantage de s'exécuter très rapidement mais demandent un long temps

d'entraînement. Les méthodes appartenant à cette catégorie ont montré de bons résultats par rapport aux trois autres types de méthodes [5]. On peut citer parmi celles-ci, la méthode basée sur les réseaux de neurones de Rowley *et al.* [6], la méthode de Schneiderman et Kanade [7] basée sur un classifieur de Bayes naïf ainsi que le fameux algorithme de Viola et Jones [8] fonctionnant en temps réel, et ce dernier sera détaillé ci-dessous

- **Détection de visage et suivi à l'aide de CAM Shift** : La détection et le suivi d'objets sont importants dans de nombreuses applications de vision par ordinateur, notamment la reconnaissance d'activité, la sécurité automobile et la surveillance. Dans ces méthodes, vous allez développer un système de suivi de visage simple en divisant le problème de suivi en trois problèmes distincts: 1. Détecter un visage à suivre ; 2. Identifier les traits du visage à suivre ; 3. Suivre le visage.

- **Algorithme de Viola et Jones** : Une avancée majeure dans le domaine a été réalisée par les chercheurs Paul Viola et Michael Jones en 2001 [8]. Ces derniers ont proposé une méthode basée sur l'apparence (*Appearance-based methods*). La méthode de Viola et Jones est une méthode de détection d'objet dans une image numérique, elle fait partie des toutes premières méthodes capables de détecter efficacement et en temps réel des objets dans une image. Inventée à l'origine pour détecter des visages, elle peut également être utilisée pour détecter d'autres types d'objets comme des voitures ou des avions. La méthode de Viola et Jones est l'une des méthodes les plus connues et les plus utilisées, en particulier pour la détection de visages et la détection de personnes. En tant que procédé d'apprentissage supervisé, la méthode de Viola et Jones nécessite de quelques centaines à plusieurs milliers d'exemples de l'objet que l'on souhaite détecter, pour entraîner un classifieur. Une fois son apprentissage réalisé, ce classifieur est utilisé pour détecter la présence éventuelle de l'objet dans une image en parcourant celle-ci de manière exhaustive, à toutes les positions et dans toutes les tailles possibles. Considérée comme étant l'une des plus importantes méthodes de détection d'objet, la méthode de Viola et Jones est notamment connue pour avoir introduit plusieurs notions reprises ensuite par de nombreux chercheurs en vision par ordinateur, à l'exemple de la notion d'image intégrale ou de la méthode de classification construite comme une cascade de classifieurs boostés. Cette méthode bénéficie d'une implémentation sous licence BSD dans OpenCV, la bibliothèque utilisée dans notre application.

Partie B : La localisation indoor

La localisation indoor ou la géolocalisation d'intérieur [10] est une technique qui permet de localiser en temps réel des biens ou des personnes dans des espaces fermés tels que les centres commerciaux, aéroports, hôpitaux, usines, complexes militaires ou industriels, etc...

Grâce aux nouveaux outils de mobilité : Smartphone professionnel, tablette tactile durcie... et aux nouvelles technologies : Wi-Fi, Bluetooth ou Ultra Wide Band et RFID (Radio Frequency Identification), les utilisateurs peuvent se repérer facilement à l'intérieur d'un bâtiment. Ces informations peuvent être utilisées également par les propriétaires de sites pour suivre le parcours des clients en temps réel et interagir avec eux en leur proposant des services et déclencher des actions marketing.

Les systèmes de géolocalisation indoor permettent de connaître avec une précision plus ou moins grande le positionnement d'une personne ou d'un objet dans un espace ou un lieu dans lequel l'accès aux satellites et les données GPS ne sont pas disponibles.

B.1 Géolocalisation Indoor

Quelques notions sur la géolocalisation indoor sont cités ci-dessous, pour plus de détails voir [11].

B.1.1 Contexte

▪ *Multiplications des applications et services demandeurs d'informations de localisation*

- ✓ Applications métier
- ✓ Réseaux sociaux (Facebook, Foursquare, Google, etc.)

▪ *Exemples de services*

- ✓ Sécurité : Protection du travailleur isolé
- ✓ Publicité géociblée
- ✓ Urgences
- ✓ Surveillance de personnes, de zones
- ✓ Tourisme/événementiel : guidage de personnes, diffusion de contenu

B.1.2 Quelques applications de géolocalisation

▪ *En milieu domestique*

- ✓ Détecter la présence de personnes
- ✓ Gérer la maison (gestion d'une alarme centrale)
- ✓ Déclencher des alarmes (enfants)
- ✓ Permettre le maintien à domicile

- ✓ Mesurer l'activité des personnes
- *Dans les bâtiments tertiaires*
 - ✓ Gestion de l'énergie
 - ✓ Travailleur isolé
 - ✓ Milieu hospitalier
 - ⇒ Gestions d'alarmes, obtention de statistiques sur des parcours
- *En milieu urbain*
 - ✓ Proposer des services urbains disponibles en indoor et outdoor
 - Ex : gestion de places de parking

B.1.3 Technologies de géolocalisation

- *Les solutions techniques existantes*
 - ✓ Exploitation de systèmes satellitaires
 - ✓ GPS, GLONASS, Galileo, Compass, IRNSS
 - ✓ Exploitation de réseaux terrestres de communications
 - ✓ GSM / UMTS, LORAN, TNT
 - ✓ Exploitation de réseaux locaux
 - ✓ WiFi, Bluetooth, Zigbee, Infra rouge, ultra son, ULB, vidéo
- *Constat*
 - ✓ Le GPS s'est "imposé" en outdoor
 - ✓ Limitations dans les environnements où les satellites et le récepteur ne sont pas en visibilité directe
 - ✓ Précision de 5 à 10m
 - ✓ Environnements indoor → les choses sont plus ouvertes

B.1.4 Le système GPS en indoor

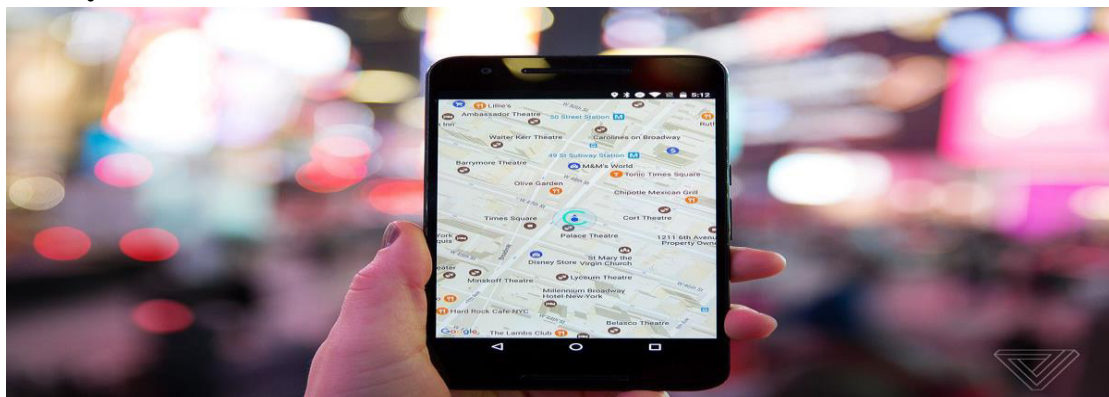


Figure 1. 3 Exemple d'application GPS sur téléphone mobile [Google]

▪ **Explosion du nombre de GPS**

- ✓ Smartphones intégrant un récepteur GPS

▪ **Intérêt**

- ✓ Avec son smartphone on peut bénéficier de services de géolocalisation en outdoor et en indoor

▪ **Limitation actuelle du GPS**

- ✓ En outdoor : Signaux des satellites GPS proches du seuil de bruit deschipset GPS
- ✓ Murs des bâtiments \Rightarrow Atténuation des signaux
- ✓ Amélioration des récepteurs (microélectronique)
- ✓ Augmentation du temps d'acquisition \rightarrow Augmentation du SNR \Rightarrow Corrélacion massive
- ✓ Localisation possible en "light indoor" (ex : proche d'une fenêtre)

▪ **Idée : Amplifier le signal \Rightarrow Pénétration dans les bâtiments**

▪ **Exploitation de pseudolites (pseudo satellites)**

- Amplification des signaux existants
- Génération de signaux de satellites GPS

▪ **Amplification des signaux existants**

- ✓ Signaux reçus à proximité dubâtiment \rightarrow Brouillage pour unrécepteur à proximité du bâtiment
- ✓ Mise en place d'un système decyclage entre les pseudolit

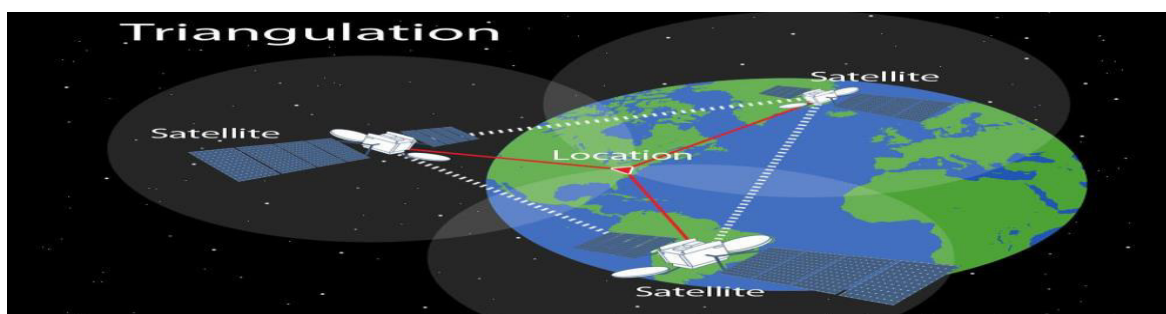


Figure 1. 4 Localisation par satellite [Google]

▪ Emission de signaux GPS

- ✓ Détection des satellites visibles depuis l'extérieur du bâtiment
 - ✓ Exploitation des séquences PN de satellites non visibles
 - ✓ Emission de signaux GPS construits par les pseudolites
 - ✓ Pas de brouillage pour un récepteur situé à l'extérieur
- ⇒ Localisation dans des environnements indoor ouverts

▪ Contraintes

- ✓ Eblouissement du récepteur GPS
- ✓ Prévu pour recevoir des signaux GPS autour -130 dBm
- ✓ Synchroniser les pseudolites
- ✓ Exploitation d'un réseau câblé

B.1.5 Les systèmes d'acquisition vidéo

Pour l'acquisition d'images et de séquences de nombreuses caméras sont positionnées dans nos environnements (Magasins, métro, aéroport, entreprises...)



Figure 1. 5 Caméra IP [Google]

On devrait exploiter les images provenant de ces caméras ensuite procéder à :

- **une calibration du système de vision**
 - ✓ Référencement de certaines positions des images
 - ✓ Projection du mobile dans le plan de l'image référencée
- **Précision de la localisation**
 - ✓ centimétrique



Figure 1. 6 Image de détection et localisation de personne [Google]

▪ Limitations

- ✓ Calibration des caméras
- ✓ Projection en fonction de l'angle de vision
- ✓ Complexité à assurer un suivi
- ✓ Champ de vision d'une caméra limité
- ✓ Variations de la luminosité
- ✓ Champ de vision limité → ne traverse pas les murs/obstacles
- ✓ Pas d'identification des cibles présentes dans le champ de la caméra
- ✓ Acceptabilité de la caméra difficile
- ✓ Services déployés à domicile

▪ Applications

- ✓ Comptage d'objets/personnes
- ✓ Détection de flux
- ✓ Statistiques sur des parcours
- ✓ Détection d'évènements (ex : passager s'éloignant de son bagage)

B.1.6 Méthodes de Localisation Indoor

Certains composants GPS peuvent recevoir un signal suffisamment élevé pour déterminer une position à l'intérieur d'un bâtiment. Cependant, le résultat n'est généralement pas assez précis pour être utilisable. De plus le GPS ne permet pas de se localiser dans un bâtiment à plusieurs étages.

La géolocalisation *indoor* est pourtant devenue le Saint Graal du marketing géolocalisé car elle permet d'assister le consommateur de son domicile au centre commercial. Elle le guide à l'intérieur du lieu, déclenche l'envoi de messages et lui permet de régler ses achats à partir de son mobile. Un des objectifs principaux de cette nouvelle méthode de marketing est d'éviter la fuite des consommateurs vers des sites e-commerce [12].

L'intérêt de la géolocalisation *indoor* ne se limite pas au marketing. De nombreux aéroports, gares, centres de congrès et d'exposition s'équipent de ces technologies pour aider les utilisateurs en leur fournissant des informations utiles.

Le terme générique de localisation indoor couvre 3 principaux niveaux de services :

- **La géolocalisation *indoor*** : véritable GPS d'intérieur qui aide les utilisateurs à trouver leur chemin, leur permet d'optimiser leur visite, de localiser leurs amis et collègues; ainsi que de fournir des analyses comportementales aux commerçants sur le parcours des visiteurs.

- **La micro-localisation** permet aux voyageurs, visiteurs ou consommateurs d'interagir avec un élément spécifique : un produit sur une étagère en magasin, une œuvre d'art exposée dans un musée... La présence du consommateur est identifiée seulement quand il est à proximité d'une balise. Sa géolocalisation est perdue lorsqu'il s'en éloigne.
- **Le Geofencing** envoie des informations spécifiques lorsque un utilisateur entre ou sort d'une zone prédéfinie. Le géomarketing ou les programmes de fidélisation comptent parmi les cas d'utilisation les plus courants de cette méthode de localisation.

Les méthodes utilisées et quels services proposent-elles ?

▪ Localisation par satellite

La localisation par satellite consiste à calculer, grâce aux signaux émis par une constellation de satellites prévue à cet effet, Cette position est alors traduite en termes de latitude, longitude et parfois altitude (ex : 43° 5494 N - 1° E) et peut alors être représentée physiquement sur une carte. Dans le cas du GPS, pour que le repérage spatial fonctionne, un immense réseau constitué de 27 satellites tournant autour de la Terre à une altitude de km et répartis sur 6 orbites différentes est nécessaire. Ces satellites constituent un maillage du ciel et servent de repères aux navigateurs GPS dans leur processus de calcul de position. Ce système de satellites est conçu de façon à ce qu'il y en ait toujours au moins quatre « visibles » par les navigateurs GPS, sans quoi la position ne peut pas être déterminée.



Figure 1. 7 Exemple d'architecture de localisation par GSM [Google]

• Localisation par GSM

Cette technique permet le positionnement d'un terminal GSM en se basant sur certaines informations relatives aux antennes GSM auxquelles le terminal est connecté. Plusieurs techniques existent :

- **Différence de temps observée ou EOTD** : le terminal calcule le temps écoulé entre l'émission et la réception de la requête envoyée à l'antenne, il peut alors calculer sa distance par rapport à celle-ci
- **Temps d'arrivée**
- **Angle d'arrivée**
- **Cell ID**

Aujourd'hui, la méthode GSM la plus utilisée est celle du CellID .Cette méthode consiste à récupérer les identifiants des antennes GSM auxquelles le terminal est connecté. Par la suite, grâce à une base de données faisant le lien entre les identifiants des cellules et les positions géographiques des antennes.

- **Localisation par adresse IP**

Cette méthode permet de déterminer la position géographique d'un ordinateur ou de n'importe quel terminal connecté à internet en se basant sur son adresse IP [13]. Les adresses IP sont gérées par l'IANA, une organisation qui s'occupe de découper les blocs d'adresses IP disponibles et de les distribuer de façon très contrôlée aux pays qui en demandent. On peut même obtenir un niveau de précision de l'ordre de la ville en se basant sur la distribution des adresses IP faite par les fournisseurs d'accès à internet. Pour une adresse IP associée à une date et une heure données, les forces de police peuvent obtenir auprès du fournisseur d'accès à internet la position géographique exacte, le nom et l'adresse communiquées par le titulaire. De même pour un espace commercial d'accès à Internet ("cybercafé"), selon les pays le commerçant peut être tenu de garder les dates, heures et documents d'identité scannés pendant une certaine période.

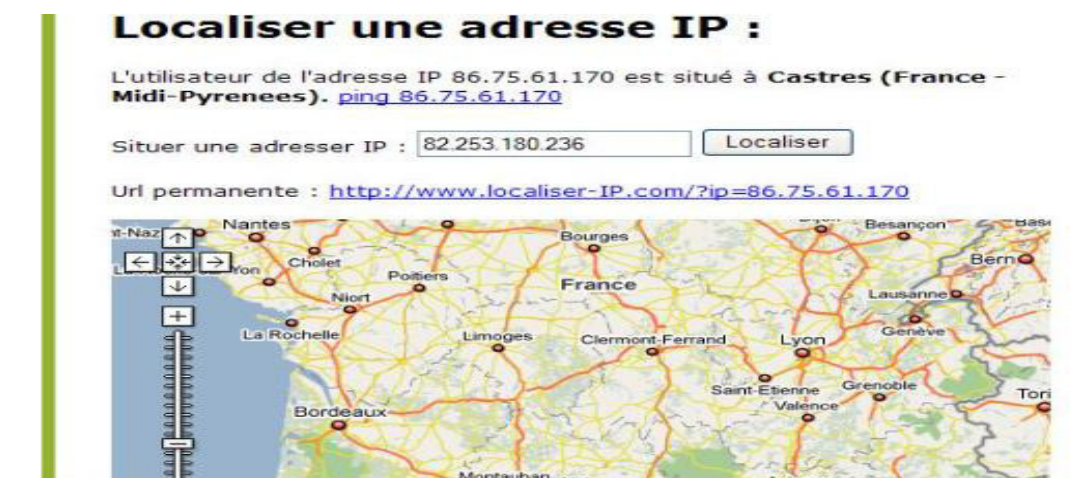


Figure 1. 8 Exemple de localisation par adresse IP [Google]

○ WI-FI POSITIONING SYSTEM : LE PIONNIER



Le **Wi-Fi positioning system (WPS)** est utilisé lorsque le GPS n'est pas adéquat. Il permet de localiser un appareil (*smartphone* par exemple) en utilisant la détection d'un réseau Wi-Fi.

○

○ Principe de fonctionnement

La technique de positionnement par point d'accès Wi-Fi est basée sur la mesure de l'intensité du signal reçu (*Received Signal Strength* ou RSS) et sur la méthode d'empreinte. Une empreinte est constituée du RSS, du SSID du point d'accès, et de l'adresse MAC du routeur. Il n'est pas nécessaire de se connecter au réseau. Un *ping* est suffisant pour déterminer la force du signal. L'appareil consulte ensuite une base de donnée distante pour faire l'association entre l'empreinte et la position. La précision du positionnement dépend du nombre de positions enregistrées dans la base de données. Pour collecter les données de localisation Wi-Fi, Google, Apple et Microsoft utilisent les *smatphones* et tablettes. Les appareils consultent périodiquement leur position en envoyant leur identifiant, la position GPS et l'empreinte Wi-Fi.

○ **Applications :** Le WPS est principalement utilisé pour du positionnement dans des lieux publics (gare, centre commerciaux, musées).

○ **Géolocalisation par RFID [14]:** La technologie RFID peut être utilisée pour la géolocalisation en intérieur. Pour ce faire, une série de lecteurs RFID équipés de différents types d'antennes sont positionnés de façon à couvrir l'ensemble de la zone souhaitée. Lorsqu'une personne équipée d'un tag RFID actif sera dans ces zones là, le système sera



capable de calculer sa position en se basant sur le nombre de lecteurs qui détectent le tag et de déduire la position approximative de l'individu en se référant au schéma de découpage établi.

En temps réel, cette technique reste néanmoins très approximative et sa précision permettra uniquement de déterminer la pièce ou le couloir dans lequel se trouve la personne géolocalisée. En effet, une fois tous les déplacements enregistrés, des systèmes informatiques peuvent réaliser toute une série de calculs probabilistes en se basant sur les lecteurs RFID. Les difficultés des localisations en intérieur en temps réel proviennent de l'environnement en constante évolution. Ces structures altèrent la puissance et la portée des signaux et rendent très difficiles l'utilisation de la triangulation avec la technologie RFID, c'est pourquoi une méthode de prédécoupage en grilles est généralement utilisée.

- **Limitations** : La localisation RFID a pour principale limitation d'être discrète par nature, fournissant une précision liée à la nature discrète de cette information discrète. Ceci provient de l'utilisation du modèle ensembliste de l'espace pour exprimer l'information de localisation par cette technologie. Afin d'affiner cette information de localisation ensembliste, la position ainsi que l'environnement entourant le lecteur tel que les murs, les portes, les tables ou les objets métalliques doivent être pris en compte. Dans l'objectif d'obtenir de cette manière une information de localisation complète il serait nécessaire de faire un pavage de l'environnement avec de multiples lecteurs dont les régions de détection formeraient dans l'idéal un recouvrement complet. Ceci est envisageable pour des applications spécifiques de gestion d'inventaire comme pour la localisation d'objets sur des étagères où chaque étagère est équipée d'une antenne. Une possibilité afin de contourner la nature discrète des RFID ainsi que d'améliorer indirectement la précision, et d'utiliser un modèle flou ou statistique pour les transitions entre l'intérieur et l'extérieur du champ de l'antenne: l'intégration temporelle de l'oscillation discrète des valeurs de détection à la frontière du champ de l'antenne peut définir une fonction caractéristique floue avec une transition lissée [14], alors qu'en tenant compte d'un modèle de déplacement humain il est possible d'identifier une distribution statistique continue. En résumé, nous devons composer avec une information de localisation RFID lacunaire, discrète et limitée, et déterminer comment nous pouvons l'améliorer cela en combinant cette information avec d'autres technologies.

- Localisation par champ magnétique : l'avenir de la géolocalisation *indoor*



Figure 1. 9 Exemple de localisation par champ magnétique [12]

Une autre solution pour se géolocaliser lorsqu'on ne capte pas de signal GPS est d'écouter les champs magnétiques, pour dessiner la carte du bâtiment dans lequel on se trouve. Le béton armé, l'acier renforcé, etc., autant de matériaux modernes qui laissent une empreinte magnétique caractéristique. Les animaux, on pense notamment au pigeon voyageur et au homard, sont capables de capter ces variations magnétiques pour retrouver leur chemin. Ainsi donc, chaque immeuble, étage, couloir, cage d'ascenseur, produit une perturbation qui lui est propre. Le système, en les lisant, peut ainsi générer une carte du lieu. Une technologie intéressante s'il en est puisqu'elle ne requiert aucun périphérique externe et offre une précision de 10 cm. Les chercheurs à l'origine du projet ont monté une entreprise IndoorAtlas et ont d'ores et déjà publié une API pour smartphones.

Principe de fonctionnement : Dans un premier temps, pour permettre la géolocalisation, il est nécessaire de cartographier le champ magnétique du bâtiment à l'aide d'un *smartphone* équipé de l'application MapCreator. Lorsque l'utilisateur se déplace, l'application va utiliser la boussole pour créer une carte du champ magnétique du bâtiment.

L'application va rechercher l'empreinte magnétique pour en déduire la position.

La précision varie de 0.1 à 2 mètres.

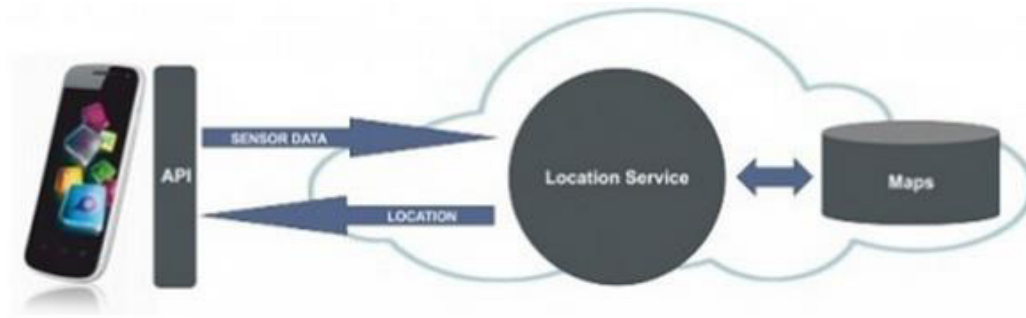


Figure 1. 10 Exemple illustrant le travail d'une technique de localisation par champ magnétique [12]

Remarque: Cette méthode de géolocalisation ne se limite pas au *Indoor*. Il est possible d'obtenir une position en pleine mer en utilisant la référence géomagnétique internationale. Cette méthode est employée par des centres de recherche pour géolocaliser la migration de bancs de poissons.

- **Applications :** L'utilisation principale du champ magnétique est la localisation dans les lieux publics (aéroports, gares, musées...).
- **Localisation basée vision**

Avec les puissances de calcul actuelles, il est possible de construire un système de suivi visuel à une fréquence vidéo qui peut suivre de multiples cibles en utilisant simplement des processus de détection au niveau pixel. Les processus de détection au niveau pixel peuvent être définis en utilisant des algorithmes tels que la soustraction de fond adaptative, la différence d'image, et des tableaux d'indexation de couleurs. Le suivi permet de restreindre la détection aux régions où les cibles sont susceptibles d'apparaître, ce qui réduit considérablement les besoins en puissance de calcul.

Le calcul robuste en temps réel [15] peut être assuré par une régulation dynamique des paramètres de détection, incluant la résolution des cibles ainsi que le nombre de cibles suivies. Le suivi visuel est un processus cyclique d'estimation récursif. De même que pour d'autres technologies d'estimation de localisation, le filtre de Kalman fournit un cadre classique pour la conception du suivi visuel. Le suivi conserve temporellement les informations sur les cibles, permettant une cohérence des objets. Cette cohérence des objets assure qu'un label affecté à une cible à un instant T1 peut être utilisé à un instant T2. Le suivi permet également de focaliser l'attention du système en appliquant le processus de détection

approprié uniquement dans les régions de l'image où une cible sera potentiellement détectée. La position ainsi que la vitesse fournies par le suivi peuvent être très importantes afin de décrire des situations.

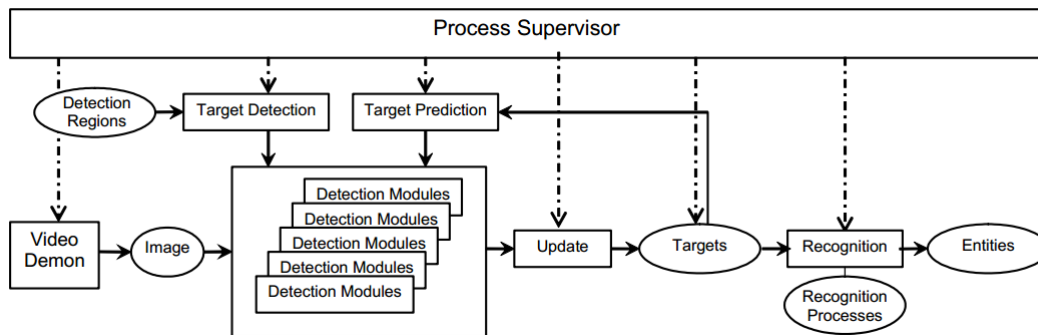


Figure 1. 11 Architecture du suiveur visuel robuste [15]

Les systèmes de suivi visuel sont généralement composés de quatre phases: la prédiction, l'observation, la détection et la mise à jour. La phase de prédiction actualise les attributs précédemment estimés pour un ensemble d'entités à une valeur estimée pour un instant donné. La phase d'observation applique la prédiction à la donnée courante afin d'actualiser l'état de chaque cible. La phase de détection détecte de nouvelles cibles. Enfin, la phase de mise à jour actualise la liste des cibles afin de tenir compte des nouvelles cibles ainsi que de celles qui ont disparu. Les expériences décrites ci-dessous utilisent un processus robuste de suivi visuel contenant plusieurs caractéristiques innovantes telles que présentées dans la Figure 1. Notre système de suivi est contrôlé par un processus superviseur intégrant des phases pour la reconnaissance, une auto régulation et la communication dans un cycle d'exécution de base. Dans la phase de reconnaissance, le suiveur interprète des méthodes de reconnaissances qui peuvent générer des événements ou des flux en sortie. La phase d'auto régulation détermine la qualité des mesures du service tel que la durée d'un cycle entier et adapte la liste des cibles ainsi que leurs paramètres afin de maintenir la qualité désirée. Durant la phase de communication, le superviseur répond aux demandes provenant d'autres processus. Ces demandes peuvent concerner des descriptions de l'état ou des capacités des processus, ou peuvent fournir des spécifications de nouvelles méthodes de reconnaissance.

Le composant de supervision du système de vision fournit des interprétations de commandes, un ordonnancement des exécutions, le traitement des événements, la régulation des paramètres, et une description réflexive. Le superviseur agit comme un interpréteur programmable recevant des morceaux de code de scripts qui détermine la composition et la

nature du cycle du processus d'exécution et la manière dont le processus réagit aux événements. Le superviseur agit comme un planificateur, sollicitant l'exécution de modules de manière asynchrone. Le superviseur régule les paramètres des modules à partir des résultats de l'exécution. Des rapports auto-critiques provenant des modules permettent au superviseur d'adapter dynamiquement l'exécution. Enfin, le superviseur répond à des requêtes externes avec une description de l'état courant et des capacités.

Conclusion

Dans ce chapitre nous avons présenté ce que nous jugeons essentiel pour notre travail sur le système qui rend possible l'intégration des technologies de localisation à différents niveaux d'abstraction. Le cas particulier de l'intégration multi-technologies présenté ici peut paraître ad hoc et spécifique aux technologies proposées, mais il peut en fait être généralisé aux technologies fournissant des données avec des propriétés similaires, correspondant à des modèles homologues de l'espace. Finalement, les technologies de localisation devraient être vues comme des services logiciels composables, en cohérence avec le paradigme des architectures orientées services. Les applications basées sur la localisation pourraient donc utiliser, dans un périmètre donné, une combinaison de solutions centralisées à large échelle comme les systèmes satellitaires ou cellulaires terrestres, avec, potentiellement, d'autres services de localisation indépendants qui pourraient leur être composés dynamiquement. Une infrastructure logicielle locale de localisation telle que décrite ci-dessus est une solution possible pour consolider cette intégration au niveau d'abstraction le plus élevé, leur combinaison aux niveaux inférieurs étant réalisable suivant le modèle décrit dans ce chapitre.

Nous ne pouvons concevoir une localisation sans une bonne détection, pour cela, nous consacrons le chapitre 2 pour l'étude, le choix et la conception d'une méthode de détection efficace, nous permettant la localisation indoor.

Chapitre 2

Etude et Conception de la Détection

Introduction

La détection de personnes est un domaine de la vision par ordinateur consistant à détecter un humain dans une image numérique. C'est un cas particulier de détection d'objet, où l'on cherche à détecter la présence et la localisation précise, dans une image, d'une ou plusieurs personnes, en général dans une posture proche de celle de la station debout ou de la marche. On parle également de détection de piéton, en raison de l'importance des applications en vidéosurveillance et pour les systèmes de vision embarqués dans des véhicules. Étudiée à partir de la fin des années 1990, la détection de personnes s'est révélée être un sujet assez difficile, en raison de la grande variété d'apparences des personnes, de l'articulation du corps humain et des phénomènes d'occultations.

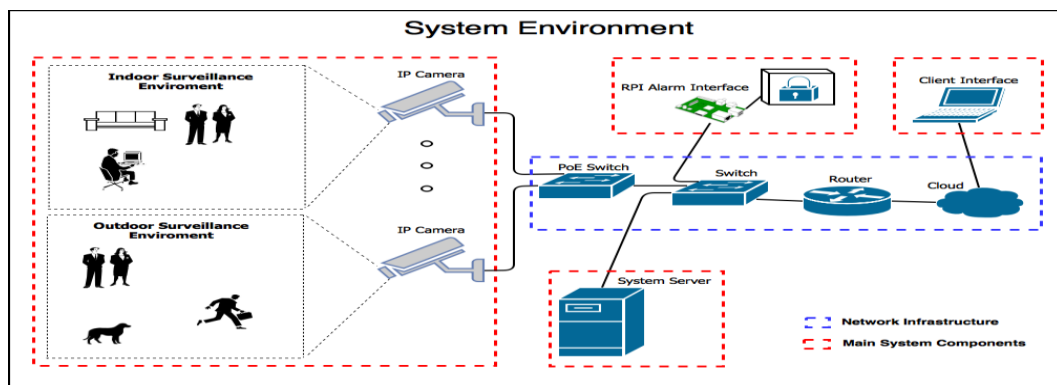


Figure 2. 1 Environnement système pour des applications en vidéosurveillance [Google]

Bénéficiant des progrès méthodologiques réalisés en détection de visage, la détection de personnes a inspiré des méthodes spécifiques, comme les histogrammes de gradient orienté, particulièrement performants. Les méthodes les plus efficaces construisent des modèles statistiques par apprentissage supervisé, à partir de caractéristiques de forme ou d'apparence, calculées sur de nombreux exemples d'images de personnes.

2.1 Méthode Gaussian Mixture Models (GMM)

Par exemple une recherche dans [16] qui a utilisé la vidéo au format (.mov) en entrée avec une fréquence d'images de 25 ips et une résolution de 640 x 480. Les données ont été prises du haut d'un pont pour piétons avec une position de caméra statique.

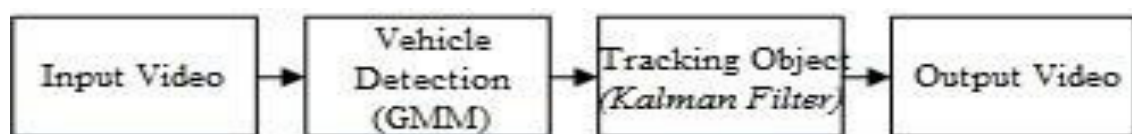


Figure 2. 2 Diagramme blocs du système [16]

Les étapes de la recherche ont été décrites dans le diagramme suivant.

Sur la base de la figure 2.2, la première étape est la préparation vidéo en entrée du système. La prochaine étape est la détection de véhicule en utilisant la méthode GMM. Dans cette étape, les objets de premier plan et les objets d'arrière-plan ont été séparés. L'objet détecté comme véhicule a été marqué avec une boîte englobante. La dernière étape consistait à suivre l'objet détecté dans chaque image à l'aide du filtre de Kalman.

2.1.1 Principe du GMM

Le GMM (Gaussian Mixture Model) est un modèle de distribution de probabilité de mélange fini. Bien que les deux modèles aient une relation étroite, ils sont toujours discutés indépendamment et séparément.

Définition de Gaussian Mixture Model :

Gaussian "Gaussien" est une courbe symétrique caractéristique "courbe en cloche" qui tombe rapidement vers 0 (pratiquement) "

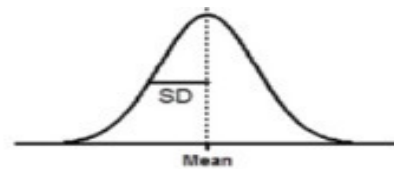


Figure 2.2.1 courbe symétrique de gaussien

Mixture Model "Le modèle de mélange est un modèle probabiliste qui suppose que les données sous-jacentes appartiennent à une distribution de mélange"

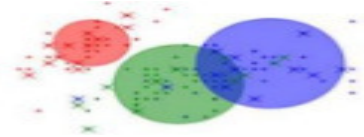


Figure 2.2.2 Le modèle de mélange de Mixture Model

GMM est une fonction de densité de probabilité décrite par une combinaison linéaire convexe de fonctions de densité Gaussienne de forme [27]:

$$f(x, \theta) = \sum_{k=1}^K \pi_k g(x; \mu_k, \Sigma_k) \quad (x \in \mathbb{R}^M), \quad (\text{Equ 2.1})$$

$\mu_k \in \mathbb{R}^M$ et Σ_k : les matrices moyenne et covariance respectivement. Les coefficients π_k sont dénotés pour mixer les probabilités

$$\pi_k > 0 \quad \text{and} \quad \sum_{k=1}^K \pi_k = 1. \quad (\text{Equ 2.2})$$

$$\Theta = ((\theta_1, \pi_1), \dots, (\theta_K, \pi_K)), \quad \text{où} \quad \theta_k = (\mu_k, \Sigma_k).$$

Le modèle est initialisé comme suit :

$$\begin{aligned} \pi_k^{(0)} &= \frac{1}{K_{max}}, \\ \mu_k^{(0)} &= p_n \rightarrow n = \lfloor (k-1)(N-1)/(K_{max}-1) \rfloor + 1, \\ \Sigma_k^{(0)} &= \frac{1}{N} \sum_{n=1}^N p_n p_n^T. \end{aligned}$$

$$MDL(K, \theta) = -\log p(p|K, \theta) + \frac{1}{2}L \log(N \cdot M). \tag{Equ 2.3}$$

2.1.2 Implémentation du détecteur GMM et application à une séquence véhicule

2.1.2.1 Extract Frame : La première étape consistait à traiter la vidéo du véhicule. La vidéo est une collection de plusieurs images. Plus la durée d'une vidéo est longue, plus le nombre d'images qu'elle contient est important. La vidéo est ensuite extraite pour devenir plusieurs images et traitée une par une jusqu'à la dernière image de la vidéo. La figure 2.3 montre l'organigramme complet du système de détection du véhicule. Les étapes des émissions du système de détection de véhicule dans l'organigramme sont discutées comme suit :

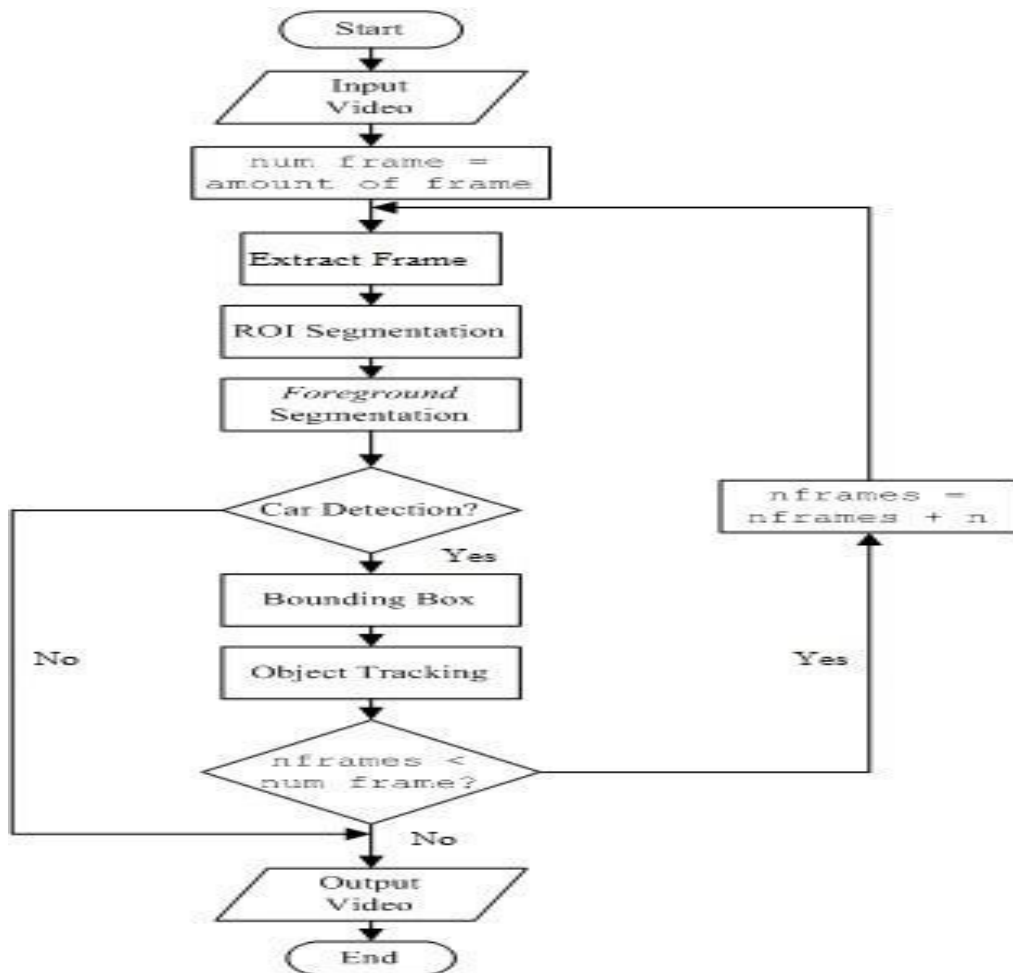


Figure 2. 3 Organigramme de l'exemple de détection de voiture [16]

Comme le montre la figure 2.3, chaque image a été traitée séquentiellement jusqu'à la dernière image de la vidéo.

2.1.2.2 ROI Segmentation : La région d'intérêt (ROI) est la zone qui contient l'objet à détecter [21]. La segmentation du ROI est nécessaire pour limiter la zone à traiter. La figure 2.4 montre la segmentation de la région d'intérêt sur la trame de la vidéo d'entrée. La figure 2.4 (a) est l'image originale avant donnée la zone limitée. La Figure 2.4 (b) montre la limite des zones ROI et non-ROI

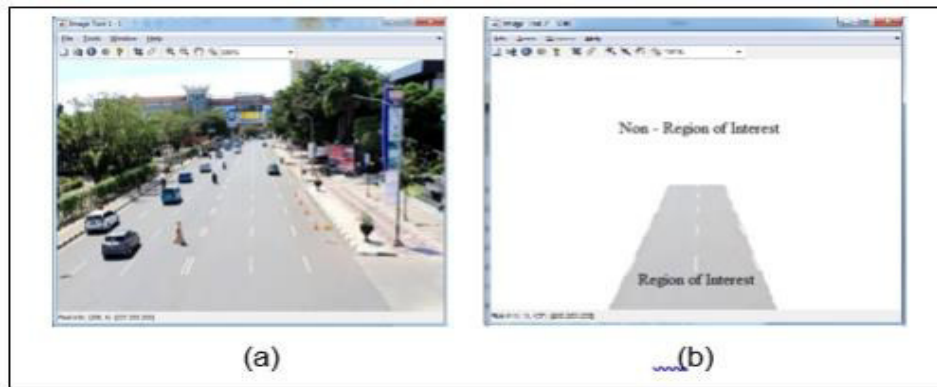


Figure 2. 4 ROI de segmentation. (a) image originale, (b) région d'intérêt [16]

2.1.2.3 Gaussian Mixture Model (GMM) : GMM est un modèle de densité constitué de plusieurs fonctions composantes gaussiennes. *Cette méthode peut bien fonctionner lorsqu'elle est utilisée pour le processus d'extraction d'arrière-plan car sa fiabilité contre les changements de lumière et d'état lors de la détection répétée d'objets* [18]. Le pixel dans la scène vidéo est modélisé en distribution gaussienne. Chaque pixel de la trame a été comparé avec le modèle formé à partir de GMM. Les pixels ayant des valeurs de similarité sous l'écart-type et le facteur de pondération le plus élevé ont été considérés comme arrière-plan, tandis que les pixels avec un écart-type plus élevé et un facteur de pondération inférieur ont été considérés comme de premier plan [22]. Le pixel est ensuite classé dans l'un des modèles candidats GMM. Si la couleur du pixel est catégorisée comme un modèle d'arrière-plan, alors le pixel recevra zéro (0) ou la couleur noire. Alors que le pixel est non catégorisé dans le modèle de fond, il sera considéré comme premier plan et donné une (1) ou couleur blanche. Ensuite, l'image binaire résultante sera traitée plus loin. Le premier plan est un objet en mouvement et change de position dans chaque image de la vidéo (dynamique), tandis que le fond est un objet dont la position est inchangée dans toutes les images vidéo (statique) [18]. Après que l'objet de premier plan a été détecté, le processus de filtrage est effectué pour remplir le trou sur l'objet de premier plan. Cette recherche utilise un processus de morphologie pour filtrer le bruit et remplir le trou sur l'objet détecté. La figure 2.5 (a) montre l'objet de premier plan et l'arrière-plan détecté. Au point (a) montre qu'il y a des trous sur l'objet de premier plan. Afin de clarifier

l'objet de premier plan a été détecté alors le processus de morphologie a été effectuée. L'opération de morphologie est un filtre qui combine le processus d'érosion et de dilatation dans une image binaire ou en niveaux de gris. Le filtrage sur le processus de morphologie est montré sur la figure 2.5 (b).

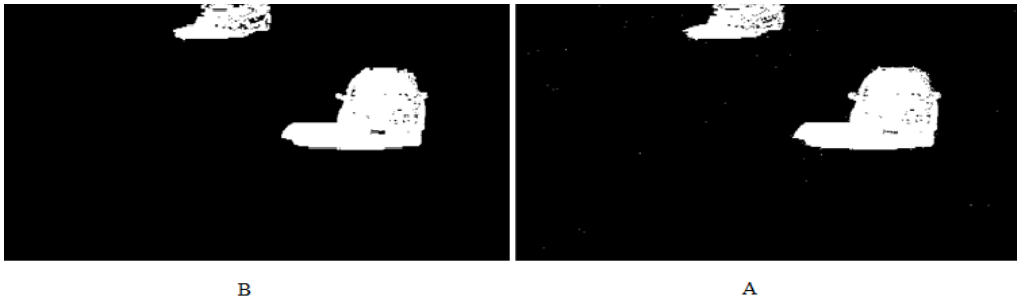


Figure 2. 5 Processus morphologique pour la détection de premier plan.

2.1.2.4 Détection de voiture : L'objet de premier plan détecté est adapté avec la zone de blob. L'objet correspondant à la zone de blob est détecté en tant qu'objet du véhicule et marqué par un cadre de délimitation. Alors que l'objet détecté en premier plan mais ne correspondant pas à la zone de blob sera ignoré et n'est pas marqué par un cadre englobant [26]. La figure 2.6 montre que l'objet du véhicule est détecté et marqué par un cadre de délimitation

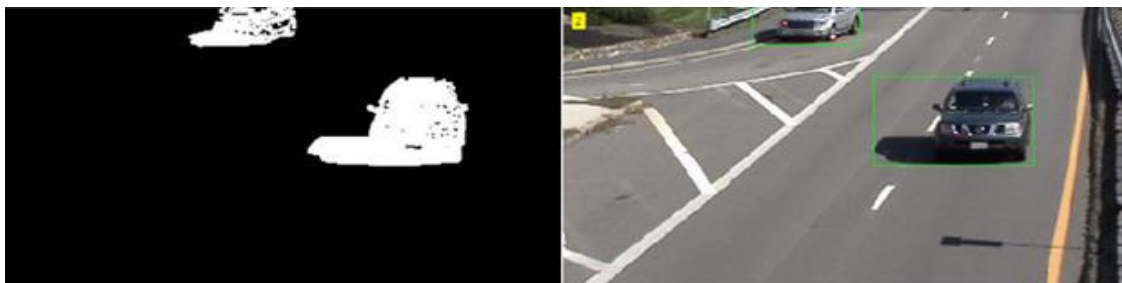


Figure 2. 6 Boîte de de limitation

2.1.2.5 KalmanFilter : Une fois l'objet du véhicule détecté, le système poursuit le suivi des objets. Le suivi d'objet est une méthode dans la vision par ordinateur pour trouver l'emplacement de l'objet détecté [23]. Dans cette recherche, la méthode de filtre de Kalman a été utilisée pour le suivi d'objet. Le filtre de Kalman est une méthode récursive bien exécutée utilisée pour suivre l'objet dans le cadre vidéo [19], [24]. Le filtre de Kalman utilise les informations de l'objet détecté sur la trame précédente et fournit la nouvelle estimation de position de l'objet. Le filtre de Kalman se compose de deux étapes à savoir la prédiction et la correction [21], [24]. L'étape de prédiction est responsable de la projection de la condition future et de la position actuelle de l'objet. Alors que l'étape de correction fournit la

réciprocité, à savoir combine la mesure réelle avec l'estimation préalable pour obtenir une estimation postérieure améliorée [21].

2.1.3 Expérience et Résultats sur la détection GMM

Ce travail utilise des données vidéo au format .mov et une résolution de 640 x 480 pixels. La détection d'objets dans la vidéo est déterminée en fonction de la taille du premier plan. Les causes de détection d'erreur sur les méthodes GMM comprennent la détection d'un véhicule fantôme en tant qu'objet et l'examen de deux véhicules adjacents comme un seul objet [25]. Dans ces conditions, la collecte de données a été effectuée pendant la journée en tenant compte de la lumière. L'objet en mouvement dans le domaine de non-retour sur investissement sera ignoré et considéré comme un arrière-plan. La ROI de segmentation déterminera les positions des pixels limites de la zone hors ROI.



(a) Cas de trafic léger (b) Cas de trafic intense

Figure 2. 7 Détection d'objet [26]

La figure 2.7.(a) montre la détection d'un objet du véhicule en utilisant GMM pour un état de trafic léger. La méthode GMM détecte l'objet en mouvement en fonction de la taille du blob du premier plan détecté. Lorsque le premier plan remplit la taille du blob, il sera traité comme un objet et une boîte englobante donnée. Dans le cas contraire, l'objet sera ignoré comme le montre la figure 2.7.(a) Dans des conditions de trafic léger, les véhicules ont été vus clairement séparés de sorte qu'aucune erreur ne soit détectée. La figure 2.7 (b) montre la détection du véhicule dans des conditions de trafic intense. La figure 3.8 (a) indique que le véhicule se chevauche avec un autre véhicule, de sorte que deux véhicules adjacents sont détectés en tant qu'objet unique, tel que montré sur la figure 3.8 (b).

La performance du système de détection est mesurée par l'analyse des caractéristiques

de fonctionnement du récepteur (ROC). Les paramètres de l'analyse ROC sont TP (Vrai Positif), FN (Faux Négatif), FP (Faux Positif) et TN (Vrai Négatif). Les performances du système sont déterminées par l'équation ci-dessous.

Précision / valeur prédictive positive (PPV):

$$Precision(PPV) = \frac{TP}{TP + FP} \quad (\text{Equ 2.4})$$

Spécificité / Vrai débit négatif (TNR): $Specificity(TNR) = \frac{TN}{N} = \frac{TN}{FP + TN}$ (Equ 2.5)

Sensibilité / Rappel / Taux Vrai Positif (TPR): $Recall(TPR) = \frac{TP}{P} = \frac{TP}{TP + FN}$ (Equ 2.6)

Précision: $Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$

On montre que la valeur de précision pour le trafic léger est de 100% alors que pour le trafic lourd, elle est de 75,79%. La sensibilité est de 94,44% pour le trafic léger et de 88,89% pour le trafic lourd. La valeur de spécificité pour le trafic léger est de 100% alors que pour le trafic lourd, elle est de 70,37%. La précision du système pour le trafic léger est de 97,22% alors que pour le trafic lourd, elle est de 79,63%. Cela prouve que la méthode GMM est meilleure pour les conditions de trafic léger.

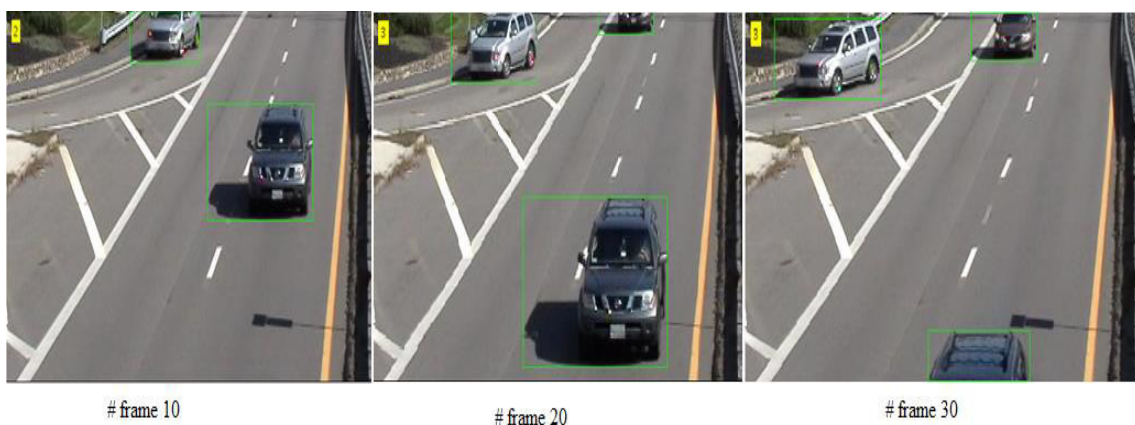


Figure 2. 8 Véhicule détecté dans chaque cadre [26]

La figure 2.9 montre le véhicule détecté dans chaque trame. Le véhicule a commencé à être détecté sur l'image 10 jusqu'à l'image 30. Avec l'utilisation de l'objet de suivi, on sait donc que l'objet détecté sur la première image est le même que l'objet détecté sur l'image suivante. La validation de l'objet de suivi est calculée par la cohérence de la prédiction d'ID d'objet de suivi en pourcentage calculée à l'aide de la formule suivante :

$$\% \text{ consistency} = \frac{\sum_{n=1}^N \left(\frac{P_n}{D_n} \times 100\% \right)}{N} \quad (\text{Equ 2.7})$$

Où : P_n = Nombre de n^{th} cohérence de prédiction d'ID

D_n = Numéro de données de n^{th} ID

N = ID total en vidéo

Le résultat de test du système d'objet de suivi pour voir la prédiction de cohérence en utilisant la méthode de filtre de Kalman montre que le pourcentage de prédiction de cohérence utilisant la méthode du filtre kalman atteint 100%. Cela prouve que cette méthode est appropriée à utiliser pour le suivi de l'objet.

2.2 Méthode de détection Viola et Jones

La méthode de Viola et Jones est une méthode de détection d'objet dans une image numérique, proposée par les chercheurs Paul Viola et Michael Jones en 2001[28]. Elle fait partie des toutes premières méthodes capables de détecter efficacement et en temps réel des objets dans une image. Inventée à l'origine pour détecter des visages, elle peut également être utilisée pour détecter d'autres types d'objets comme des voitures ou des avions. *La méthode de Viola et Jones est l'une des méthodes les plus connues et les plus utilisées, en particulier pour la détection de visages et la détection de personnes.* En tant que procédé d'apprentissage supervisé, la méthode de Viola et Jones nécessite de quelques centaines à plusieurs milliers d'exemples de l'objet que l'on souhaite détecter, pour entraîner un classifieur. Une fois son apprentissage réalisé, ce classifieur est utilisé pour détecter la présence éventuelle de l'objet dans une image en parcourant celle-ci de manière exhaustive, à toutes les positions et dans toutes les tailles possibles. *Considérée comme étant l'une des plus importantes méthodes de détection d'objet, la méthode de Viola et Jones est notamment connue pour avoir introduit plusieurs notions reprises ensuite par de nombreux chercheurs en vision par ordinateur, à l'exemple de la notion d'image intégrale ou de la méthode de classification construite comme une cascade de classifieurs boostés.*

2.2.1 Éléments de la méthode

La méthode de Viola et Jones est une approche basée sur l'apparence, qui consiste à parcourir l'ensemble de l'image en calculant un certain nombre de caractéristiques dans des zones rectangulaires qui se chevauchent. Elle a la particularité d'utiliser des caractéristiques très simples mais très nombreuses. Une première innovation de la méthode est l'introduction

des images intégrales, qui permettent le calcul rapide de ces caractéristiques. Une deuxième innovation importante est la sélection de ces caractéristiques par boosting, en interprétant les caractéristiques comme des classifieurs. Enfin, la méthode propose une architecture pour combiner les classifieurs boostés en un processus en cascade, ce qui apporte un net gain en temps de détection. La méthode, en tant que méthode d'apprentissage supervisé, est divisée en deux étapes : une étape d'apprentissage du classifieur basé sur un grand nombre d'exemples positifs (c'est-à-dire les objets d'intérêt, par exemple des visages) et d'exemples négatifs, et une phase de détection par application de ce classifieur à des images inconnues.

2.2.1.1 Caractéristiques

a) Description : Plutôt que de travailler directement sur les valeurs de pixels, et pour être à la fois plus efficace et plus rapide, Viola et Jones proposent d'utiliser des caractéristiques, c'est-à-dire une représentation synthétique et informative, calculée à partir des valeurs des pixels. Viola et Jones définissent des caractéristiques très simples, les caractéristiques pseudo-Haar, qui sont calculées par la différence des sommes de pixels de deux ou plusieurs zones rectangulaires adjacentes. La figure 2.10 donne des exemples des caractéristiques proposées par Viola et Jones à 2, 3 ou 4 rectangles, dans lesquelles la somme de pixels sombres est soustraite de la somme des pixels blancs. Leur nom vient de leur similitude avec les ondelettes de Haar. Plus formellement, l'image intégrale ii au point (x, y) est définie à partir de l'image i par : $ii(x, y) = \sum_{(x' \leq x, y' \leq y)} i(x', y')$. Grâce à cette représentation, une caractéristique formée de deux zones rectangulaires peut être calculée en seulement 6 accès à l'image intégrale, et donc en un temps constant quelle que soit la taille de la caractéristique.



Figure 2. 9 Exemple des types de caractéristiques utilisées par Viola et Jones

b) Calcul : Les caractéristiques sont calculées à toutes les positions et à toutes les échelles dans une fenêtre de détection de petite taille, typiquement de 24×24 pixels ou de 20×15 pixels. Un très grand nombre de caractéristiques par fenêtre est ainsi généré, Viola et Jones donnant l'exemple d'une fenêtre de taille 24×24 qui génère environ 160 000 caractéristiques.

c) Sélection de caractéristiques par boosting : Le deuxième élément clé de la méthode de Viola et Jones est l'utilisation d'une méthode de boosting afin de sélectionner les meilleures caractéristiques. Le boosting est un principe qui consiste à construire un classifieur « fort » à

partir d'une combinaison pondérée de classifieurs « faibles. Le classifieur se réduit alors à un couple (caractéristique, seuil). L'algorithme de boosting utilisé est en pratique une version modifiée d'AdaBoost. Plus formellement, on considère un ensemble de $n(x, \dots, xn)$ images et leurs étiquettes associées (y, \dots, yn) , qui sont telles que $yi = 0$ si l'image xi est un exemple négatif et que $yi = 1$ si xi est un exemple de l'objet à détecter. L'algorithme de boosting est constitué d'un nombre T d'itérations, et pour chaque itération t et chaque caractéristique j , on construit un classifieur faible hj . Idéalement, le but est d'obtenir un classifieur h qui prédise exactement les étiquettes pour chaque échantillon, c'est-à-dire $yi = h(xi) \forall i \in \{1 \dots n\}$. En pratique, le classifieur n'est pas parfait et l'erreur engendrée par ce classifieur est donnée par :

$$\epsilon_j = \sum_{i=1}^n w_i |h_j(x_i) - y_i|, \tag{Equ 2.8}$$

w_i étant les poids associés à chaque exemple et mis à jour à chaque itération en fonction de l'erreur obtenue à l'itération précédente. L'itération t et le classifieur ht présentant l'erreur la plus faible :

$$\epsilon_t = \min_j(\epsilon_j) \tag{Equ 2.9}$$

Le classifieur fort final $h(x)$ est construit par seuillage de la somme pondérée des classifieurs faibles sélectionnés :

$$h(x) = \begin{cases} 1 & \text{si } \sum_{t=1}^T \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t \\ 0 & \text{sinon} \end{cases} \tag{Equ 2.10}$$

Les α_t sont des coefficients calculés à partir de l'erreur ϵ_t

d) Cascade de classifieurs : L'une des idées-clés de la méthode pour réduire ce coût réside dans l'organisation de l'algorithme de détection en une cascade de classifieurs. Appliqués séquentiellement, ces classifieurs prennent une décision d'acceptation la fenêtre contient l'objet et l'exemple est alors passé au classifieur suivant, ou de rejet la fenêtre ne contient pas l'objet et dans ce cas l'exemple est définitivement écarté.

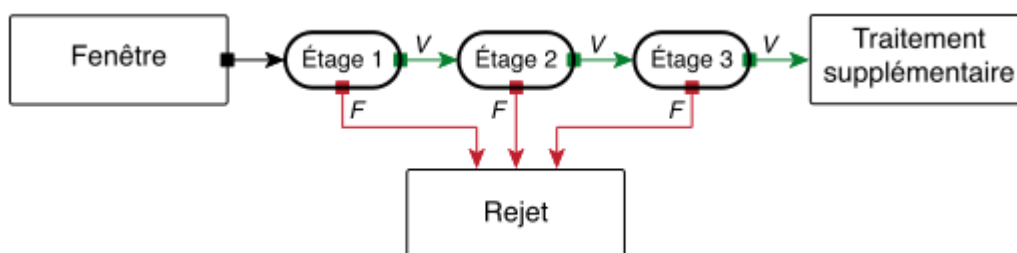


Figure 2. 10 Illustration de l'architecture de la cascade

L'idée est que l'immense majorité des fenêtres testées étant négatives (c.-à-d. ne contiennent pas l'objet), il est avantageux de pouvoir les rejeter avec le moins possible de calculs. Ici, les classifieurs les plus simples, donc les plus rapides, sont situés au début de la cascade, et rejettent très rapidement la grande majorité des exemples négatifs. Les itérations d'AdaBoost continuent jusqu'à ce que le taux de fausse alarme cible soit atteint.

Des caractéristiques/classifieurs faibles sont ajoutés jusqu'à ce que les taux cibles soient atteints, avant de passer ensuite à l'étage suivant.

Plusieurs chercheurs font remarquer que cette idée de filtrer rapidement les exemples négatifs les plus simples n'est pas nouvelle. Elle existe dans d'autres méthodes sous forme d'heuristiques, comme la détection de la couleur chair ou une étape de pré-classification.

2.2.2 Étapes clés

2.2.2.1 Apprentissage : L'apprentissage est réalisé sur un très large ensemble d'images positives (c'est-à-dire contenant l'objet) et négatives (ne contenant pas l'objet). Plusieurs milliers d'exemples sont en général nécessaires. Cet apprentissage comprend :

- *Le calcul des caractéristiques pseudo-Haar sur les exemples positifs et négatifs*
- *L'entraînement de la cascade : à chaque étage de la cascade, un classifieur fort est entraîné par AdaBoost. Il est construit par ajouts successifs de classifieurs faibles entraînés sur une seule caractéristique, jusqu'à l'obtention de performances conformes aux taux de détection et de fausse alarme souhaités pour l'étage.*

2.2.2.2 Détection : La détection s'applique sur une image de test, dans laquelle on souhaite déceler la présence et la localisation d'un objet. En voici les étapes :

- *Parcours de l'ensemble de l'image à toutes les positions et échelles, avec une fenêtre de taille 24×24 pixels, et application de la cascade à chaque sous-fenêtre, en commençant par le premier étage :*
 - *calcul des caractéristiques pseudo-Haar utilisées par le classifieur de l'étage courant,*
 - *puis calcul de la réponse du classifieur,*
 - *passage ensuite à l'étage supérieur si la réponse est positive, à la sous-fenêtre suivante sinon,*
 - *et enfin l'exemple est déclaré positif si tous les étages répondent positivement ;*
- *Fusion des détections multiples : l'objet peut en effet générer plusieurs détections, à différentes positions et échelles ; cette dernière étape fusionne les détections qui se chevauchent pour ne retourner qu'un seul résultat.*

2.2.3 Performances de la méthode Viola Jones

Viola et Jones ont testé leur algorithme sur la base de visages MIT+CMU, constituée de 130 images contenant 507 visages de face. Ils présentent leur résultat sous la forme d'une courbe *Receiver Operating Characteristic* (ROC), qui donne le taux de détection correct en fonction du nombre de fausses alarmes total sur toutes les images du corpus. À titre d'exemple, pour 50 fausses alarmes, ils obtiennent un taux de détection de 88,8 %.

2.2.4 Limitations et extensions de la méthode Viola Jones

De très nombreuses améliorations ont été proposées par la suite, visant à améliorer le paramétrage de la méthode, ou à en combler un certain nombre de limitations. L'une des premières améliorations est apportée par Lienhart et Maydt en 2002. Ils proposent d'étendre l'ensemble de caractéristiques pseudo-Haar utilisé de 4 à 14 caractéristiques. De même, ils introduisent des caractéristiques « de biais » (tournées de 45°), ainsi qu'une méthode pour les calculer basée sur une extension des images intégrales. D'autres types de caractéristiques ont également été utilisées en remplacement des caractéristiques de Haar : les histogrammes de gradients orientés, les motifs binaires locaux ou la covariance de région. Les chercheurs ont également proposé d'utiliser des variantes de l'algorithme de boosting, Notamment RealBoost, qui produit un indice de confiance à valeurs réelles, en plus de la classification. Plusieurs travaux ont ainsi montré la supériorité de RealBoost sur AdaBoost dans le cadre de l'algorithme de Viola et Jones. Viola et Jones étendent en 2003 leur système à la détection de piétons dans des vidéos, en incluant une information de mouvement en plus de l'information d'apparence.

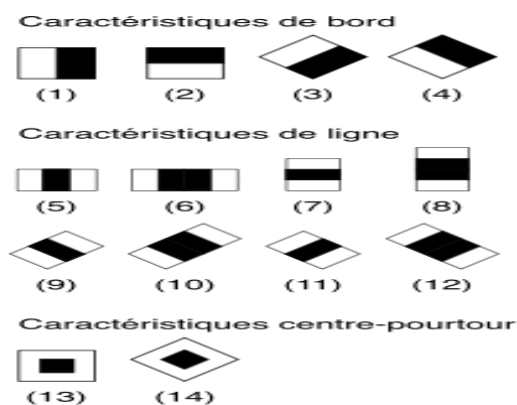


Figure 2.11 Extension des caractéristiques pseudo-Haar proposée par Lienhart

Une des limitations de la méthode est son manque de robustesse à la rotation, et sa difficulté à apprendre plusieurs vues d'un même objet. En particulier, il est difficile d'obtenir un

classifieur capable de détecter à la fois des visages de face et de profil. Viola et Jones ont proposé une amélioration qui permet de corriger ce défaut, qui consiste à apprendre une cascade dédiée à chaque orientation ou vue, et à utiliser lors de la détection un arbre de décision pour sélectionner la bonne cascade à appliquer. Plusieurs autres améliorations ont été proposées par la suite pour apporter une solution à ce problème. Une autre limitation importante de la méthode de Viola et Jones concerne le temps d'apprentissage de la cascade, compris généralement entre plusieurs jours et plusieurs semaines de calcul, ce qui limite énormément les possibilités de tests et de choix des paramètres. Un des problèmes majeurs de la méthode proposée par Viola et Jones est qu'il n'existe pas de méthode optimale pour choisir les différents paramètres régissant l'algorithme : le nombre d'étages, leur ordre ou les taux de détection et de fausses alarmes pour chaque étage doivent être choisis par essais et erreurs. Plusieurs méthodes sont proposées pour déterminer certains de ces seuils de manière automatique. Un reproche également fait à la méthode concerne la perte d'information subie au passage d'un étage à l'autre de la cascade, perte due à l'effet coupeur des décisions d'acceptation ou de rejet prises à chaque étage. Certains chercheurs proposent la solution de garder l'information contenue dans la somme pondérée des classifieurs faibles, par exemple le « boostingchain » de Xiao. Une modification plus radicale de structure est proposée par Bourdev et sa notion de cascade souple, qui consiste essentiellement à supprimer le concept d'étages, en formant un seul classifieur fort, donc une seule somme, et en permettant de prendre une décision à chaque évaluation de classifieur faible et de s'affranchir de la contrainte des taux de détection et de fausses alarmes cibles.

2.2.5 Applications la méthode de Viola et Jones

La méthode de Viola et Jones a essentiellement été appliquée à la détection de visage et à la détection de personne, principalement en raison des nombreuses applications pratiques qu'offrent ces deux domaines, notamment en vidéosurveillance, en indexation d'images et de vidéo ou pour les interfaces homme-machine multimodales. Un exemple d'application grand public de la méthode est donné par les appareils photographiques numériques, où elle sert à effectuer la mise au point automatique sur les visages. Combinée avec le standard JPEG 2000, la méthode peut également servir à compresser les visages avec un taux de compression plus faible que le reste de l'image, afin de préserver les détails des visages. Les constructeurs automobiles s'intéressent également à la méthode pour concevoir des systèmes de sécurité capables de détecter automatiquement les autres usagers de la route, en particulier les piétons. Des recherches ont également montré que l'efficacité de la méthode ne se limite pas au

domaine visible, mais qu'elle s'étend également au domaine infrarouge. La méthode de Viola et Jones a également été utilisée pour détecter d'autres types d'objets, par exemple des mains, pour la commande gestuelle d'une interface homme-machine, des voitures dans des images satellites pour la création de systèmes d'information géographique débarrassés de toute présence visuelle d'automobiles, ou pour l'évaluation et le suivi du trafic routier. La méthode a également été évaluée pour la détection d'avions dans des images de basse résolution à partir d'une caméra embarquée dans un véhicule aérien, pour l'évitement de collisions.

Des applications militaires existent aussi pour la détection de cibles (chars, avions) dans des images aériennes ou satellitaires.

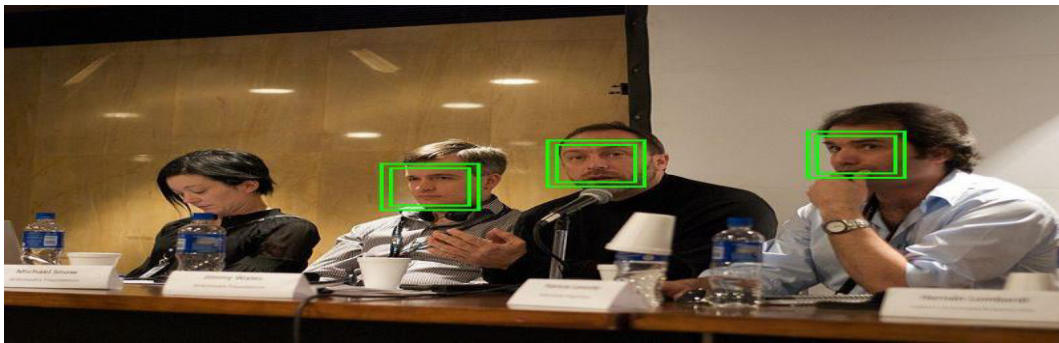


Figure 2.12 Détection de visage avec la méthode de Viola et Jones

2.2.6 Notre implémentation du détecteur Viola Jones

Il existe de nombreuses implémentations du détecteur de Viola et Jones, la plus utilisée étant celle en C++ présente dans la librairie de vision par ordinateur OpenCV, publiée sous licence BSD. Des implémentations ont été développées pour des environnements ou plates-formes spécifiques, notamment pour une exécution dans des navigateurs Web en utilisant le langage de script ActionScript du logiciel multimédia Flash. Des implémentations matérielles ont également été développées sur ASIC, FPGA et sur GPU. L'utilisation de l'architecture parallèle de ces derniers permet un net gain de temps de détection par rapport à l'implémentation OpenCV traditionnelle. Enfin, les implémentations les plus courantes sont celles rencontrées dans les appareils photographiques numériques pour la mise au point automatique par la détection de visage. Elles nécessitent des optimisations particulières pour faire face à la faible puissance de calcul de ce type de matériel. La première série d'expérience est menée sur des images issues du net .

1) *Expérience 1* : Exemple sur image de plusieurs personnes visages frontales indoor



Figure 2. 13 Exemple d'image d'une grande collection détectée par Viola Jones

2) *Expérience 2* : Exemple sur image de plusieurs personnes ((lunettes, expressions, pose) outdoor.

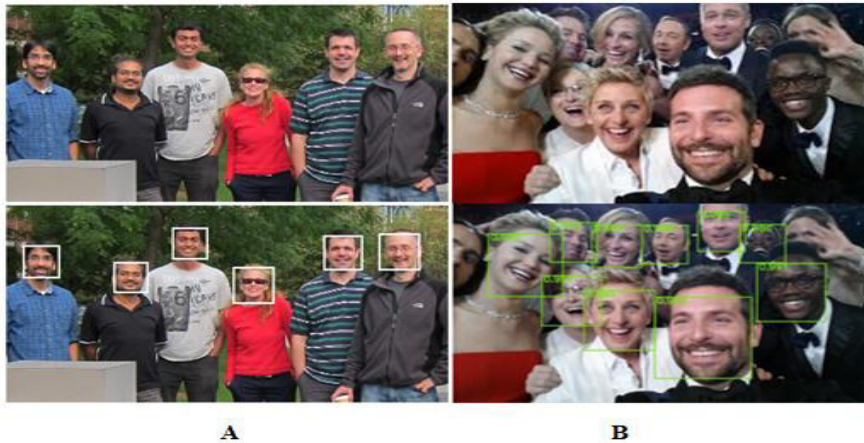


Figure 2. 14 Détection de visages de plusieurs personnes frontales outdoor

2) *Expérience 3* : Exemple sur image de plusieurs personnes frontales debout indoor.



Figure 2. 15 Détection de visages de plusieurs personnes frontales indoor

4) *Expérience 4* : Exemple sur image de plusieurs personnes frontales indoor



Figure 2. 16 Détection de visages de plusieurs personnes frontales indoor
A (lieu de travail) et B (une classe)

5) *Expérience 5* : Exemple sur image d'une personne visage frontale indoor



Figure 2. 17 Détection de personne frontale indoor

- 6) *Expérience 6* : Exemple sur image de personne visage indoor avec inclinaison de la tête.



Figure 2. 18 Détection et suivi du visage utilisant le nuage de points caractéristiques

- 7) *Expérience 7* : Exemple sur image de plusieurs personnes indoor avec différentes variantes du visage (profil, expressions, occlusions, pose)



Figure 2. 19 Détection de deux/quatre personnes indoor sur les deux images A et B(lieu de travail)

8) *Expérience 8* : Exemple sur image de plusieurs personnes debout outdoor en action sous diverses variantes (expressions, rotation de la tête, action)

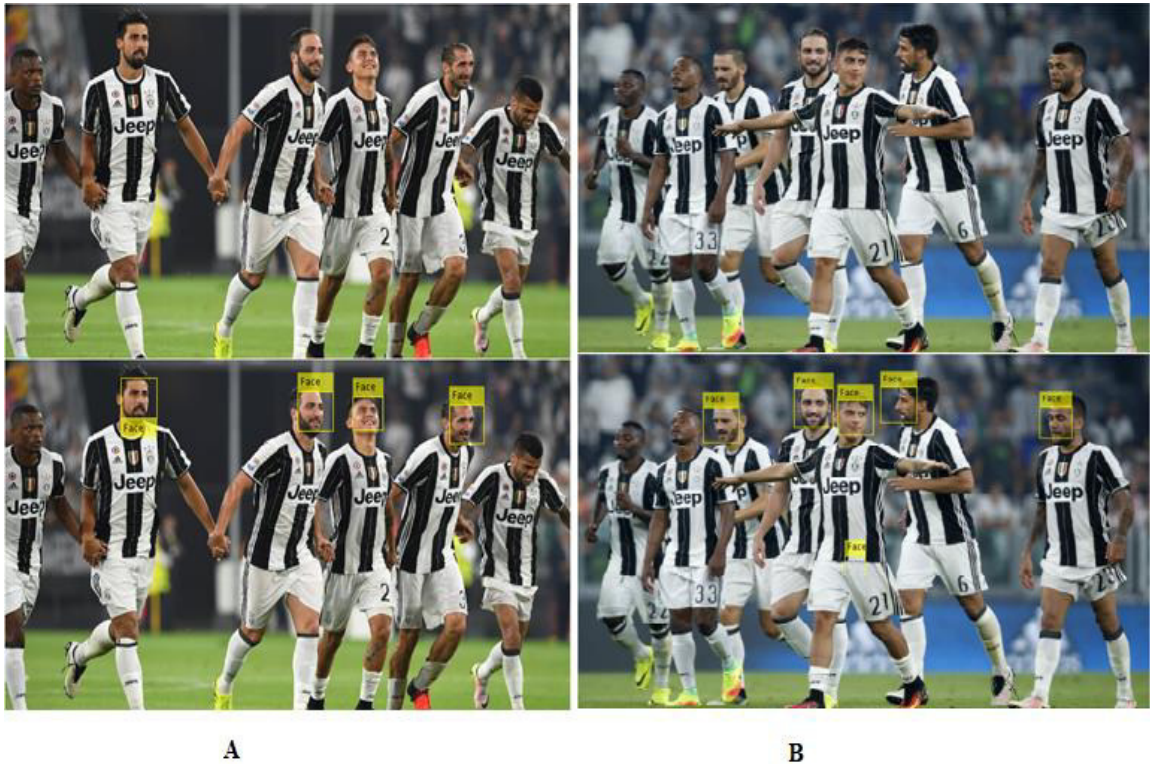


Figure 2. 20 Détection de visage cas de personnes en action

Dans l'exemple des images de la figure (footballeurs en action sur le terrain). Nous obtenons les résultats de la détection par Viola Jones suivants : quatre/six joueurs dans l'image A ont été détectés et cinq/sept joueurs dans image

Discussion

Notons que de l'expérience 1 à l'expérience 6 le processus de détection des personnes dans l'environnement interne et externe a été bien exécuté lorsque le taux d'erreur atteint 0%, ce qui signifie que la méthode de détection Viola Jones est efficace à 100% . Notons aussi que de l'expérience 7 à l'expérience 8 il y a une erreur dans le processus de détection des personnes dans l'environnement interne et externe où le taux d'erreur atteint 50% dans les images A et B pour l'expérience 7. Par contre dans l'expérience 8, quatre joueurs sur six étaient détectés sur la photo A et cinq sur sept sur l'image B. Nous pouvons dire en observant les deux faits, que les erreurs de détection peuvent être attribuées aux expressions faciales ou aux orientations. Cela nous pousse à dire que la détection de la méthode Viola Jones présente des limites dans les milieux incontrôlés et rencontre des difficultés à certaines variantes liées aux mouvements dans le visage notamment l'expression et la rotation de la tête ainsi que la présence d'arrière plan.

2.3 Méthode Aggregate Channel Features (ACF)

Aggregate Channel Features (ACF) ACF est un détecteur de personnes rapide basé sur la notion de canaux de descripteurs (Channel Features) qui a grandement amélioré les performances en détection sur de nombreux jeux de données en comparaison avec détecteurs précédemment disponible. Il s'appuie sur des agrégats de descripteurs hétérogènes que l'on représente comme des canaux, un classifieur appris par boosting, et un modèle d'apparence holistique. ACF s'appuie sur dix canaux : valeur du gradient, HOG (Histogram Oriented Gradient) sur six canaux, et canaux couleur LUV. Dans cette étude [29], ACF est entraîné sur le jeu de données INRIA. Un exemple de détection est illustré dans la figure (2.22) .

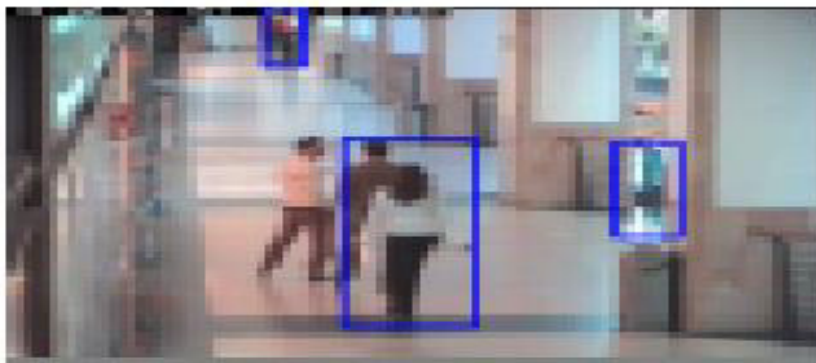


Figure 2. 21 Exemples de détection.

2.3.1 Détecteur ACF ajusté en tant que région

Le détecteur ACF proposé par Dollar et al. pour la détection des piétons, des canaux à résolution inférieure à 10 sont extraits de l'image d'entrée en tant que caractéristiques. Ceux-ci sont constitués de trois canaux de couleur LUV, d'une amplitude de gradient normalisée de l'image d'entrée et des six derniers canaux correspondant à l'histogramme de gradient qui sont calculés en utilisant des caractéristiques HOG pour six groupes d'orientation différents. Ces canaux sont ensuite vectorisés et transmis à un algorithme d'amplification appris pour obtenir une sortie de détection. Afin de réduire la corrélation locale entre les canaux ci-dessus, on a calculé les vecteurs propres correspondant à la matrice de covariance des 10 canaux et les quatre premiers vecteurs propres sont utilisés pour filtrer les canaux comme proposé par Nam et al. [30]. Ces canaux localement décorrélés ont encore amélioré les performances. En utilisant les canaux filtrés ci-dessus comme caractéristiques, on a effectué une analyse détaillée en faisant varier quatre paramètres significatifs du détecteur pour optimiser les performances dans l'ensemble de données Caltech.

1. Taille de la fenêtre coulissante: En augmentant la taille de la fenêtre coulissante, le nombre des fonctions disponibles pour l'apprentissage seront augmentées et ainsi le pouvoir discriminant du détecteur sera amélioré. Bien que, en augmentant la taille de la fenêtre au-delà d'une limite, il en résultait un grand nombre de caractéristiques qui augmentaient les calculs et entraînaient également la disparition de petites instances. D'où un optimum taille de la fenêtre de 120×60 a été utilisé comme un compromis entre ces deux facteurs.

2. Taille du dataset de formation: plus de nombre d'instances de haute et basse résolution peut être détecté en augmentant la taille de l'ensemble de données d'apprentissage mais en augmentant une limite entraînait un ajustement excessif du classificateur. Dans le jeu de données Caltech des images disponibles pour la formation peut être modifié en changeant le nombre de images extraites des vidéos (six séries de vidéos sont disponibles pour la formation l'ensemble de données pédestre Caltech). L'ensemble de données d'entraînement obtenu en prenant chaque troisième image (42782 images) a fourni les meilleures performances du détecteur.

3. Glissement de la fenêtre coulissante: la valeur de la glissière de la fenêtre coulissante définit le nombre de pixels ignorés entre les balayages de fenêtre glissante adjacents. Même si augmenter la valeur de la foulée (la taille du pas se déplace le filtre convolution) permettra de réduire les calculs, le détecteur manquerait de petites instances de piétons. On a donc réduit la valeur de foulée de quatre dans le détecteur de base à deux afin de détecter plus de petites instances des piétons.

2.3.2 Synoptique général

Le synoptique d'un système usuel de suivi par détection est illustré par la figure(2.23) ci-dessous :

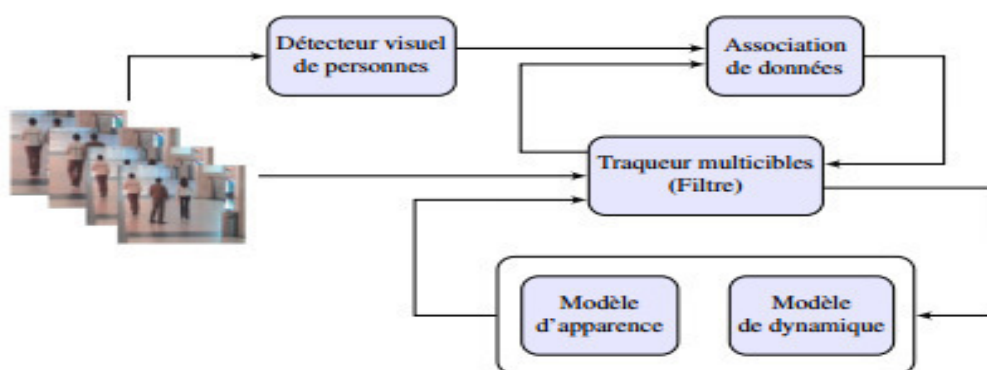


Figure 2. 22 Synoptique d'un système usuel de suivi par détection [30].

2.3.3 Notre implémentation de la méthode ACF et Résultats

2.3.3.1 Détection indoor avec mono caméra (webcam)

Expérience 1 : Image d'une personne visage frontale en lumière modérée



Figure 2. 23 Détection par méthode ACF (lumière sombre)

Expérience 2 : Exemple sur image de personne frontales indoor (variation de lumière et distance caméra)



Figure 2. 24 Détection par méthode ACF (Variante lumière et distance)

Expérience 3 : Exemple sur image personne frontales indoor



Figure 2. 25 Détection par méthode ACF (variante distance et lumière)

Expérience 4 : avec cam mobile indoor et outdoor (vue frontale)

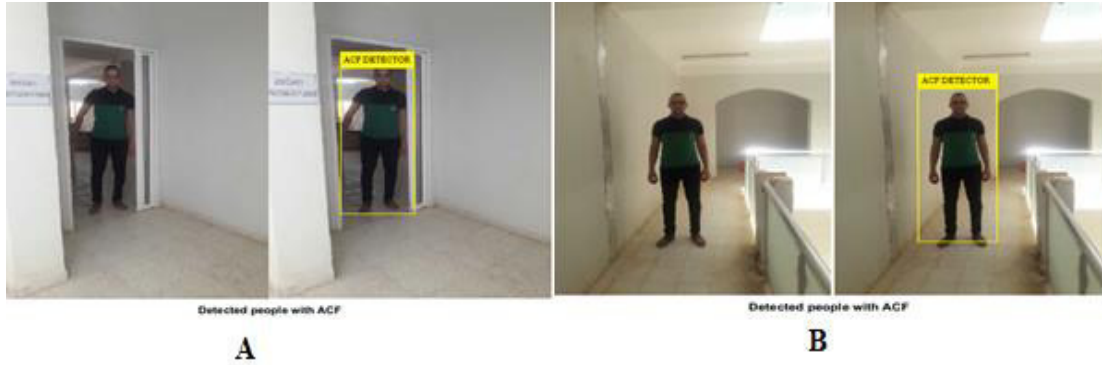


Figure 2. 26 Détection par la méthode ACF utilisation cam mobile

Notons que de l'expérience 1 à l'expérience 4, où nous avons utilisé une seule caméra pour la détection des personnes dans l'environnement interne et externe, le taux d'erreur atteint est inexistant, ce qui signifie que la méthode de détection ACF a été efficace.

2.3.3.2 Détection d'image stéréo (avec caméra binoculaire)

Expérience 1 : Vue frontale et vue de profil



Figure 2. 27 Détection d'image stéréo dans différentes poses

Expérience 2 : Orientation de la tête



Figure 2. 28 Détection d'image stéréo avec orientation de la tête

Discussion : Malgré que le profil et la rotation de la tête reste un défi pour la détection et la reconnaissance, le détecteur ACF reste efficace dans toutes les expériences menées.

2.4 Méthode de détection par HOG (Histogramme de gradient orienté)

2.4.1 Description générale de la méthode

L'idée importante derrière le descripteur HOG [31] est que l'apparence et la forme locale d'un objet dans une image peuvent être décrites par la distribution de l'intensité du gradient ou la direction des contours. Ceci peut être fait en divisant l'image en des régions adjacentes de petite taille, appelées cellules, et en calculant pour chaque cellule l'histogramme des directions du gradient ou des orientations des contours pour les pixels à l'intérieur de cette cellule. La combinaison des histogrammes forme alors le descripteur HOG. Pour de meilleurs résultats, les histogrammes locaux sont normalisés en contraste, en calculant une mesure de l'intensité sur des zones plus larges que les cellules, appelées des blocs, et en utilisant cette valeur pour normaliser toutes les cellules du bloc. Cette normalisation permet une meilleure résistance aux changements d'illuminations et aux ombres. Dalal et Triggs ont testé de nombreuses combinaisons de pré et post-traitements ou de valeurs de paramètres. On ne décrit ici que les valeurs et les méthodes retenues.

2.4.2 Construction du descripteur

2.4.2.1 Calcul du gradient

Une étape de pré-traitement peut être effectuée avant le calcul du gradient, afin que les couleurs de l'image soient normalisées et une correction gamma correcte. Cette étape ne s'est finalement pas avérée nécessaire, la normalisation du descripteur lui-même s'avérant suffisante. Plusieurs espaces de couleurs ont été testés, le meilleur étant RGB. La première étape de la méthode est le calcul du gradient, la méthode la plus courante pour cela consistant à appliquer un filtre dérivatif 1D centré, dans les directions horizontales et verticales. Les masques suivants sont utilisés pour cela:

$$[-1, 0, 1] \text{ et } [-1, 0, 1]^T$$

Dans le cas des images couleurs, le gradient est calculé séparément pour chaque composante, et on retient pour chaque pixel le gradient de plus grande norme.

D'autres types de masques plus complexes ont été testés, comme des filtres de Sobel 3x3, ou des masques diagonaux, ou non centrés. Des essais d'appliquer un filtrage gaussien avant l'application du masque a été essayé, mais toutes ces opérations font baisser sensiblement les performances par rapport à l'application du masque dérivatif le plus simple, et sans filtrage préalable.

2.4.2.2 Construction de l'histogramme

La seconde étape est la création des histogrammes de l'orientation des gradients. Ceci est fait dans des cellules carrées de petite taille (de 4x4 à 12x12 pixels). Chaque pixel de la cellule vote alors pour une classe de l'histogramme, en fonction de l'orientation du gradient à ce point. Le vote du pixel est pondéré par l'intensité du gradient en ce point. Les histogrammes sont uniformes de 0 à 180° (cas non signé) ou de 0 à 360° (cas signé). Une quantification fine de l'histogramme est nécessaire, et on obtient les meilleurs résultats avec un histogramme à 9 classes. Prendre en compte le signe du gradient n'améliore pas les performances pour la détection de personnes, mais peut être significatif pour d'autres types d'objets.

2.4.2.3 Formation des blocs

Une étape importante est la normalisation des descripteurs afin d'éviter les disparités dues aux variations d'illumination. Cette étape introduit également de la redondance dans le descripteur. Pour cela, les auteurs regroupent plusieurs cellules dans un bloc, qui est l'unité sur laquelle est effectuée la normalisation. Les blocs se recouvrent, donc une même cellule participe plusieurs fois au descripteur final, comme membre de blocs différents. Deux types de géométrie de blocs sont proposées: rectangulaire (R-HOG) ou circulaire (C-HOG). Les expériences faites ont montré que la meilleure performance était obtenue pour des blocs rectangulaires contenant 3x3 cellules de 6x6 pixels chacune. Une amélioration mineure de performance est obtenue en pondérant les pixels par une fenêtre gaussienne sur le bloc, diminuant la contribution des pixels sur les bords. Les auteurs remarquent que la formation du descripteur R-HOG est assez semblable au descripteur SIFT, à la différence que les R-HOG sont calculés sur une grille dense et à une seule échelle, et non sur des points d'intérêts à plusieurs échelles comme les SIFT.

2.4.2.4 Normalisation des blocs

Quatre types de normalisation sont explorées. Le vecteur non normalisé contenant tous les histogrammes d'un seul bloc est désigné par \mathbf{v} , sa k-norme par $\|\mathbf{v}\|_k$ et ϵ est une constante de faible valeur. Le facteur de normalisation est alors défini par:

$$\begin{aligned}
 \text{L2-norme: } f &= \frac{\mathbf{v}}{\sqrt{\|\mathbf{v}\|_2^2 + \epsilon^2}} && \text{(Equ 2.11)} \\
 \text{L1-norme: } f &= \frac{\mathbf{v}}{(\|\mathbf{v}\|_1 + \epsilon)} \\
 \text{L1-racine: } f &= \sqrt{\frac{\mathbf{v}}{(\|\mathbf{v}\|_1 + \epsilon)}}
 \end{aligned}$$

Une quatrième norme L2-hys, consistant à calculer v tout d'abord par la L2-norme, puis limiter les valeurs maximales de v à 0.2, et ensuite à renormaliser. Les normes L2-Hys, L2-norme, et L1-racine obtiennent des performances similaires, tandis que L1-norme obtient de moins bons résultats, mais toutefois bien meilleurs que l'absence de normalisation.

2.4.3 Classification

L'étape finale dans le processus de détection d'objet est l'utilisation des descripteurs HOG pour entraîner un classifieur supervisé. Cette étape ne fait pas partie de la définition du descripteur HOG à proprement parler et différents types de classifieurs peuvent être utilisés. Dans certains travaux un classifieur simple, un SVM à noyau linéaire, est choisi afin de mesurer essentiellement l'apport des HOG. Il serait aussi précisent en particulier qu'il serait intéressant de développer une méthode à base de cascade comme la méthode de Viola et Jones, en utilisant les HOG.

L'état de l'art précise que l'utilisation d'un noyau gaussien améliore les performances de 3 %, pour un taux de faux positifs par fenêtre de 10^{-4} , mais un coût de calcul bien plus élevé.

2.4.4 Extensions

Lors du workshop PASCAL sur la détection de classe d'objet en 2006, Dalal et Triggs ont présenté des résultats de l'application des HOG sur d'autres types d'objets, en particulier des voitures, des bus, des vélos, ainsi que des animaux comme des chiens, chats, vaches. Pour obtenir de bons résultats, les paramètres de taille de blocs et la méthode normalisation doivent être adaptés pour chaque classe d'objet. Toujours en 2006, à ECCV, Dalal et Triggs, avec Cordelia Schmid ont proposé une extension à la vidéo, pour détecter des personnes dans des films. Cela consiste à combiner les HOG calculés sur chaque image de la vidéo, avec un histogramme de mouvement (Internal Motion Histograms IMH) entre deux images successives. Les IMH utilisent l'amplitude du gradient provenant du Flux optique calculé sur deux images consécutives, et ces amplitudes sont utilisées de la même manière que les HOG calculés sur les images statiques. Lors de tests sur des corpus extraits de plusieurs DVD de films, la méthode qui combine HOG et IMH obtient un taux de ratés de 0.1, pour 10^{-4} de taux de fausse alarme. Les HOG ont rapidement été repris par d'autres chercheurs pour des systèmes de détection de personnes. En 2006, des chercheurs de l'INSA Rouen et de l'université de Parme proposent un système basé sur les HOG et deux caméras infrarouges, et la position des personnes est estimée par stéréovision.

Lors de la conférence CVPR de 2006, des chercheurs de Mitsubishi Electric proposent d'utiliser les HOG en combinaison avec une cascade de classifieurs, similaire à celle présente dans la méthode de Viola et Jones. Ils introduisent également des blocs qui varient en taille et en position, et les meilleurs sont sélectionnés par AdaBoost. Leurs expériences donnent des résultats similaires à ceux de Dalal et Triggs, mais avec des temps de détection jusqu'à 70 fois plus rapide. Un brevet a été déposé par Mitsubishi sur cette méthode.

2.4.5 Expériences et résultats de la détection par la méthode HOG

2.4.5.1 Détection avec mono caméra



Figure 2. 29 Détection d'une personne indoor avec différentes variantes

Discussion : A travers toutes ces expériences nous remarquons que le détecteur de HOG reste efficace même avec un éclairage modéré, une distance assez importante et sur différentes poses.

2.4.5.2 Détection avec caméra binoculaire

1^{re} Expérience :

Détection en présence des variantes : lumière, rotation de la tête , profil , distance.

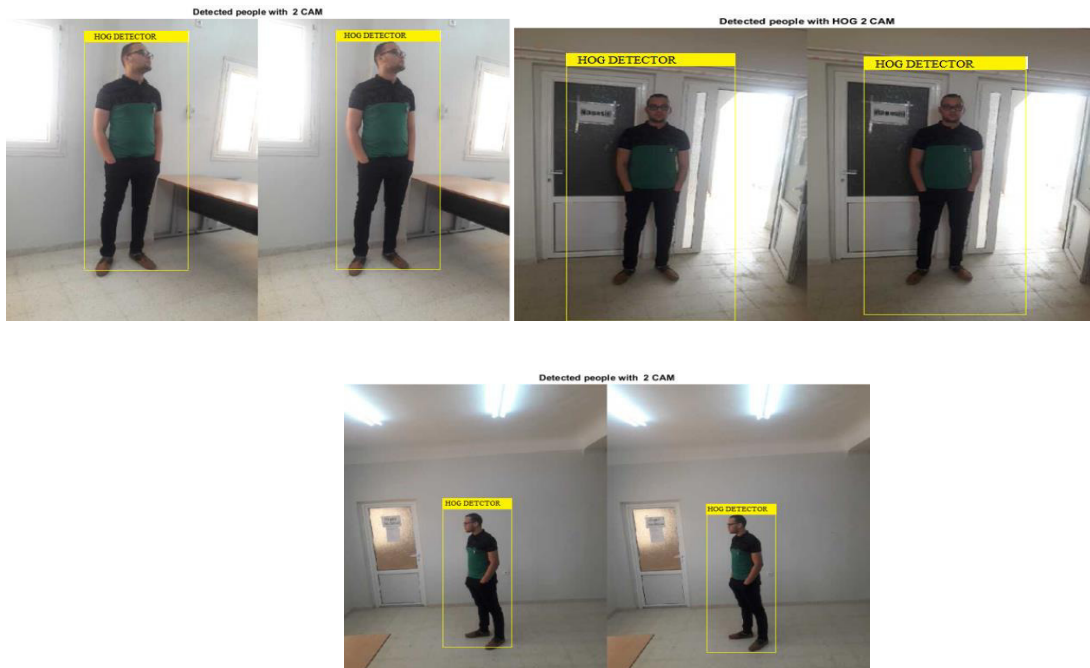


Figure 2. 30 Détection en présence des variantes lumière, rotation de la tête, profil, distance.

2^{ème} Expérience : en mouvement



Figure 2. 31 Détection d'une personne en mouvement indoor avec différentes variantes (lumière, profil, distance)

Discussion : D'après les résultats obtenus, nous pouvons dire que le détecteur de HOG reste un bon candidat même dans le cas des images en présence de mouvement. Nous verrons par la suite, comment utiliser ce détecteur pour la localisation indoor.

Conclusion

Dans ce chapitre nous avons présenté et étudié quatre méthodes de détection jugées par les chercheurs des plus populaires et efficaces. Nous avons ensuite conçu et testé ces quatre différentes méthodes, puis nous les avons validées sur des images indoor et outdoor présentant diverses variantes. D'après notre étude et les expériences menées sur les quatre détecteurs : GMM, Viola Jones, ACF et HOG ; nous pouvons dire que HOG et ACF sont les meilleurs détecteurs aptes à réaliser notre objectif qui est la localisation indoor. A travers toutes les expériences réalisées nous choisissons donc HOG pour le reste de notre travail. Dans le chapitre suivant, nous nous intéressons à l'étude et la conception de la calibration et localisation.

Chapitre 3

Etude et Conception de Calibration et Localisation

Introduction

Avec les puissances de calcul actuelles, il est possible de construire un système de suivi visuel à une fréquence vidéo qui peut suivre de multiples cibles en utilisant simplement des processus de détection au niveau pixel. Les processus de détection au niveau pixel peuvent être définis en utilisant des algorithmes tels que la soustraction de fond adaptative, la différence d'image, et des tableaux d'indexation de couleurs. Le suivi permet de restreindre la détection aux régions où les cibles sont susceptibles d'apparaître, ce qui réduit considérablement les besoins en puissance de calcul. Le calcul robuste en temps réel peut être assuré par une régulation dynamique des paramètres de détection, incluant la résolution des cibles ainsi que le nombre de cibles suivies.

Le suivi visuel est un processus cyclique d'estimation récursif. De même que pour d'autres technologies d'estimation de localisation, le filtre de Kalman fournit un cadre classique pour la conception du suivi visuel. Le suivi conserve temporellement les informations sur les cibles, permettant une cohérence des objets. Cette cohérence des objets assure qu'un label affecté à une cible à un instant T1 peut être utilisé à un instant T2. Le suivi permet également de focaliser l'attention du système en appliquant le processus de détection approprié uniquement dans les régions de l'image où une cible sera potentiellement détectée. La position ainsi que la vitesse fournie par le suivi peuvent être très importantes afin de décrire des situations.

Les systèmes de suivi visuel sont généralement composés de quatre phases: la prédiction, l'observation, la détection et la mise à jour. La phase de prédiction actualise les attributs précédemment estimés pour un ensemble d'entités à une valeur estimée pour un instant donné. La phase d'observation applique la prédiction à la donnée courante afin d'actualiser l'état de chaque cible. La phase de détection détecte de nouvelles cibles. Enfin, la phase de mise à jour actualise la liste des cibles afin de tenir compte des nouvelles cibles ainsi que de celles qui ont disparu. Les expériences décrites ci-dessous utilisent un processus robuste de suivi visuel contenant plusieurs caractéristiques innovantes. Notre système de suivi est contrôlé par un processus superviseur intégrant des phases pour la reconnaissance, une auto-régulation et la communication dans un cycle d'exécution de base. Dans la phase de reconnaissance, le suiveur interprète des méthodes de reconnaissances qui peuvent générer des événements ou des flux en sortie. La phase d'auto-régulation détermine la qualité des mesures du service tel que la durée d'un cycle entier et adapte la liste des cibles ainsi que leurs paramètres afin de maintenir la qualité désirée. Durant la phase de communication, le

superviseur répond aux demandes provenant d'autres processus. Ces demandes peuvent concerner des descriptions de l'état ou des capacités des processus, ou peuvent fournir des spécifications de nouvelles méthodes de reconnaissance. Le composant de supervision du système de vision fournit des interprétations de commandes, un ordonnancement des exécutions, le traitement des événements, la régulation des paramètres, et une description réflexive. Le superviseur agit comme un interpréteur programmable recevant des morceaux de code de scriptes qui détermine la composition et la nature du cycle du processus d'exécution et la manière dont le processus réagit aux événements. Le superviseur agit comme un planificateur, sollicitant l'exécution de modules de manière asynchrone. Le superviseur régule les paramètres des modules à partir des résultats de l'exécution. Des rapports auto-critiques provenant des modules permettent au superviseur d'adapter dynamiquement l'exécution. Enfin, le superviseur répond à des requêtes externes avec une description de l'état courant et des capacités.

3.1 Etude de la calibration des image

L'étalonnage de la caméra géométrique, également appelé résection de la caméra, permet d'estimer les paramètres d'une lentille et d'un capteur d'image d'une caméra vidéo ou d'une image. Vous pouvez utiliser ces paramètres pour corriger la distorsion de l'objectif, mesurer la taille d'un objet dans les unités du monde ou déterminer l'emplacement de la caméra dans la scène. Ces tâches sont utilisées dans des applications telles que la vision industrielle pour détecter et mesurer des objets. Ils sont également utilisés en robotique, pour les systèmes de navigation, et la reconstruction de scènes 3D.

3.2 Etude de la localisation indoor

La vidéo et les dispositifs recevant des images d'une scène permettent d'effectuer d'une part une détection de présence d'un élément dans une scène, mais aussi de localiser cet élément dans la scène. La localisation est effectuée grâce à des transformations entre l'image de la scène et les angles de vues de la caméra [32, 33]. Une utilisation possible de cette technique est de détecter les intrusions dans une zone. Grâce aux techniques de reconnaissance de contours, un objet est repérable sur une image. Il est possible de suivre le déplacement de ce contour tant qu'il reste dans le champ de vision de la caméra. [34] présente une technique de poursuite de cible grâce à la vidéo. Ce système est aussi utilisé en robotique. Les nouveaux robots arrivant sur le marché commencent à gagner en autonomie grâce aux

systèmes de vision. Ces robots peuvent se repérer dans l'espace et donc se déplacer. Cette technique possède comme faiblesse la portée limitée du système. Dans les environnements indoor, la portée se trouve restreinte à une seule pièce (emplacement de la caméra). Des problèmes d'identification se posent. Ce problème n'est pas négligeable car les applications requièrent, en plus de la position d'un mobile, un identifiant permettant de distinguer un mobile par rapport aux autres. Or avec cette technologie, différencier deux objets mobiles n'est pas simple. Lorsque deux objets se croisent et sont assez proches, l'un des objets masque l'autre pendant un bref instant. Ce masquage est suffisant pour que le système de détection par vidéo conclut qu'il n'y a qu'un seul objet dans la scène. Si un instant suivant, ces deux objets se séparent, ces deux cibles sont vues comme de nouvelles cibles pour le dispositif par vidéo. Le problème du système est de déterminer quel était le nom affecté à chacune des cibles précédentes et de redonner à chacune le bon nom suite à cet événement de fusion

3.3 Acquisition des images

Dans nos expériences l'acquisition est réalisée de plusieurs manières, au départ nous avons juste la webcam de notre pc. Mais, nous nous sommes rendus compte qu'avec ce type de capteur ; on ne peut réaliser des expériences performantes. Car ce type de matériel est dédié à des acquisitions en face du pc et à une distance limitée avec la contrainte de mauvaise résolution.

Dans notre deuxième série d'expériences nous utilisons une caméra binoculaire pour une acquisition stéréo qui serait plus adaptée à la localisation.

Là aussi, l'acquisition présente la limite de mauvaise résolution en plus un réglage du logiciel développé nécessite un réglage pour chaque série d'acquisition. Notre objectif est de réaliser une localisation automatique basée sur les algorithmes de traitement d'image, pour cette raison nous avons pensé à réaliser une acquisition avec deux smartphones. Les trois types d'acquisition sont présentés dans la section suivante :

3.3.1 Acquisition par webcam

Une **webcam** est une **caméra vidéo** (figure 3.1) qui alimente ou diffuse son image en temps réel vers ou via un ordinateur vers un réseau informatique. Lorsqu'il est "capturé" par l'ordinateur, le flux vidéo peut être sauvegardé, visualisé ou envoyé à d'autres réseaux via des systèmes tels que l'Internet, et envoyé par e-mail en pièce jointe. Lorsqu'il est envoyé à un emplacement distant, le flux vidéo peut être enregistré, visualisé ou envoyé. Contrairement à une caméra IP (qui se connecte via Ethernet ou Wi-Fi), une webcam est généralement

connectée par un câble USB, ou un câble similaire, ou intégré dans du matériel informatique, tel qu'un ordinateur portable.

Le terme "webcam" (un composé écrêté) peut également être utilisé dans son sens original d'une caméra vidéo connectée au Web en continu pendant une durée indéterminée, plutôt que pour une session particulière, fournissant généralement une vue à quiconque visite sa page Web. Certains d'entre eux, par exemple, ceux utilisés comme caméras de circulation en ligne, sont des caméras vidéo professionnelles robustes et coûteuses.



Figure 3. 1 Webcam utilisée et ses caractéristiques

3.3.2 Acquisition par caméra IP

Une **caméra IP** ou **caméra réseau** est une caméra de surveillance utilisant le Protocole Internet pour transmettre des images et des signaux de commande via une liaison FastEthernet. Certaines caméras IP [35] sont reliées à un enregistreur vidéo numérique (DVR) ou un enregistreur vidéo en réseau (NVR) pour former un système de surveillance vidéo.

L'avantage des caméras IP est qu'elles permettent aux propriétaires et aux entreprises de consulter leurs caméras depuis n'importe quelle connexion internet via un ordinateur portable ou un smartphone.

Une caméra IP peut être câblée avec du RJ45 vers un routeur ou « box ADSL », ce qui lui permet à la fois d'être alimentée et les images visionnées sur le réseau, ou alors par Wi-Fi (une alimentation en courant électrique devient alors nécessaire). Contrairement aux webcams USB, la compatibilité avec les logiciels de visioconférence n'est pas toujours garantie.

Les caméras de surveillance IP tendent à se démocratiser, plusieurs modèles sur le marché sont à la disposition des particuliers. Les modèles cloud sont basés sur des services payants, les caméras autonomes fonctionnent avec une interface web ainsi que des modèles fixes ou motorisés.

Lorsque l'adresse IP est dynamique, la caméra est dotée d'un client type DynDns. Il est nécessaire de configurer son routeur pour pouvoir y accéder depuis l'extérieur, de manière à rediriger un ou plusieurs ports déterminés vers l'adresse IP locale de la caméra.

3.3.2.1 Application utilisée de la caméra IP



Figure 3. 2 Logo d'application

IP Webcam est une application qui nous permet de convertir notre appareil Android en une caméra Internet avec de multiples options de vue que l'on peut voir sur n'importe quelle plateforme en utilisant le VLC Player ou un navigateur Internet.

IP Webcam permet d'enregistrer des vidéos en Webm, MOV et MPEG4 (seulement offert pour les appareils avec le système d'exploitation Android 4.1 ou une version ultérieure). La transmission audio, de l'autre côté, peut être en tant que fichier Wav, Opus ou ACC (le dernier de ceux-ci nécessite Android 4.1 ou une version ultérieure). Les options de détection de mouvement permettent de laisser le téléphone pointer quelque part et lorsque quelque chose bouge devant, il commencera à enregistrer. Ce mode, bien sûr, consomme beaucoup de la batterie, donc il s'agit seulement d'une bonne idée si l'appareil est branché. IP Webcam est un outil qui permet de convertir l'appareil Android dans un outil de surveillance vidéo. On a qu'à laisser l'appareil connecté au chargeur (ou l'ordinateur) et suivre ce qui se déroule n'importe où.

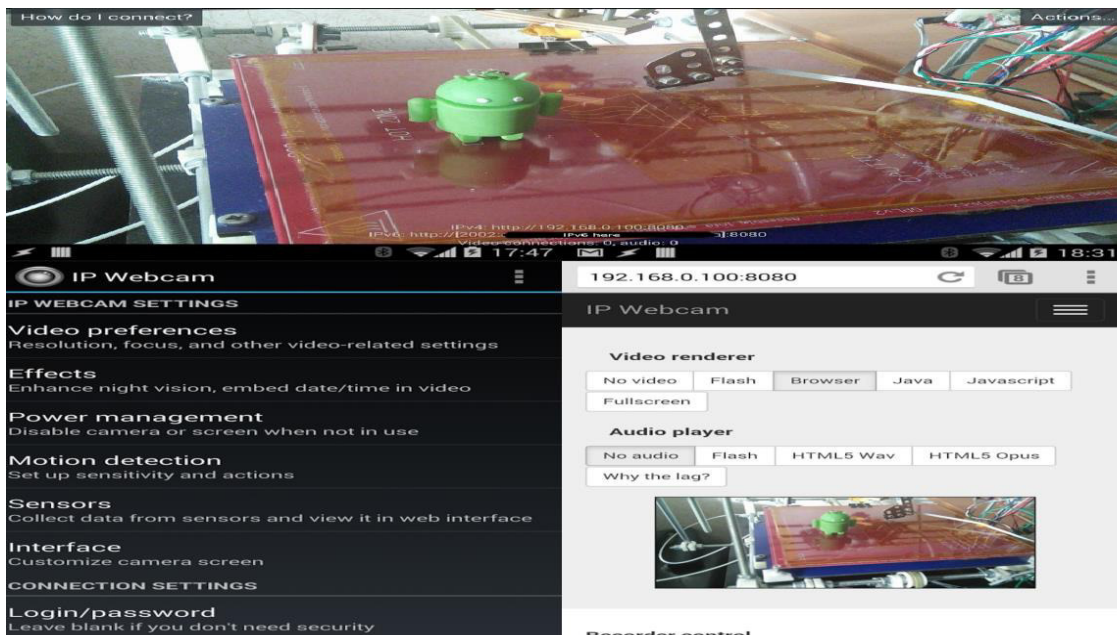


Figure 3. 3 Exemple d'utilisation de l'application IP webcam [36]

3.3.2.2 Installation du toolbox caméras IP dans matlab

Le lien de téléchargement du toolbox est : <https://fr.mathworks.com/hardware-support/ip-camera.html>

3.3.2.3 Photo du PC et ses caractéristiques



Figure 3. 4 Photo du PC et ses caractéristiques

Le langage de programmation estle matlab version 2016.

3.3.3 Acquisition stéréo binoculaire

3.3.3.1 Equation de disparité de base

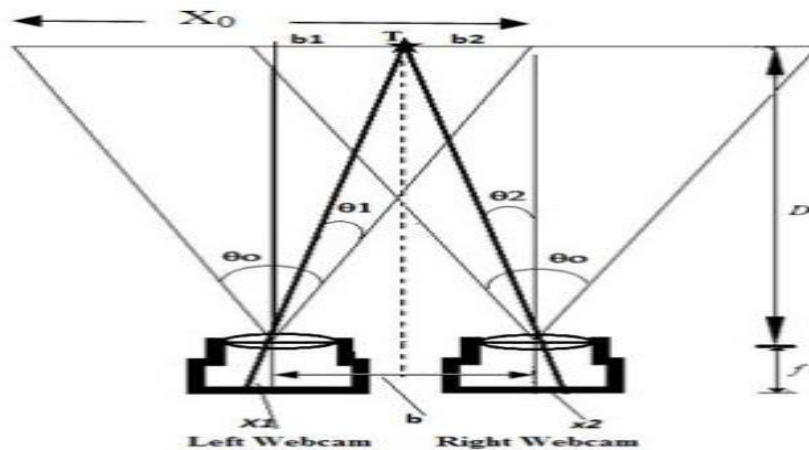


Figure 3. 5 Image montrant le calcul de dimension pour la calibration

Le diagramme schématique de la fig. (3.5) montre deux webcams, à gauche webcam et webcam droite alignés de sorte que leurs axes optiques parallèles et à une distance b l'un de l'autre. Les deux webcams ont les mêmes paramètres, c'est-à-dire la même distance

focale f et même angle de vue θ de la cible T sera à distance x_1 dans la caméra de gauche et à une distance x_2 dans la caméra de droite. L'application de la triangulation de base entraînera:

$$\frac{b_1}{D} = \frac{-x_1}{f} \quad , \quad \frac{b_2}{D} = \frac{x_2}{f} \tag{Equ 3.1}$$

Depuis $b=b_1+b_2$, puis, $b = \frac{bD}{f}(x_2 - x_1)$ $D = \frac{bf}{x_2-x_1}$ (Equ 3.2)

$$\tan\left(\frac{\theta}{2}\right) = \frac{x_0}{D} = \frac{x_1}{f} \quad f = \frac{x_0}{2 \tan\left(\frac{\theta}{2}\right)} \tag{Equ 3.3}$$



Figure 3. 6 Installation et mise en œuvre des caméras binoculaires sur PC

3.3.4 Acquisition par Smartphones

3.3.4.1 Photo et caractéristiques des smartphones utilisés



Figure 3. 7 Photos des deux smartphones utilisés

Caractéristiques du smartphone 1 :

CAMERA	Primary	13 MP (f/1.9, 28mm), autofocus, LED flash, check quality
	Features	Geo-tagging, touch focus, face detection, panorama, HDR
	Video	1080p@30fps, check quality
	Secondary	5 MP, f/1.9, LED flash

Caractéristiques du smartphone 2 :

CAMERA	Primary	13 MP (f/1.9, 28mm), autofocus, LED flash
	Features	Geo-tagging, touch focus, face detection
	Video	1080p@30fps
	Secondary	5 MP (f/2.2, 23mm), LED flash

3.3.4.2 Installation et mise en œuvre de l'acquisition par smartphones

Figure 3. 8 Installation et mise en oeuvre de l'acquisition par smartphone

Nous avons rencontré des difficultés à mettre au point l'acquisition par smartphones particulièrement dans la stabilité de ceux-ci. Mais, nous sommes enfin arrivés à faire quelques acquisitions suffisantes pour nos expériences.

3.4 Application de calibreur de caméra stéréo

3.4.1 Présentation du calibreur de caméra stéréo

Nous avons utilisé l'application Stéréo CameraCalibrator pour calibrer une caméra stéréo, que l'on a ensuite utilisé pour récupérer la profondeur des images [37]. Un système stéréo se compose de deux caméras: caméra 1 et caméra 2. L'application peut soit estimer ou importer les paramètres de caméras individuelles. L'application calcule également la position et l'orientation de la caméra 2, par rapport à la caméra 1.

L'application Stereo Camera Calibrator produit un objet contenant les paramètres de la caméra stéréo. Cet objet peut être utilisé pour :

- Rectifier les images stéréo ;
- Reconstruire la scène 3D ;

- Calculer des emplacements 3D correspondant à des paires de points d'image en utilisant la fonction de triangulation.

La suite de fonctions d'étalonnage utilisée par l'application Stereo Camera Calibrator fournit le flux de travail pour l'étalonnage du système stéréo. Ces fonctions sont utilisées directement dans l'espace de travail MATLAB.

3.4.1.1 Calibration de la Stéréo Caméra

Nous suivons ce flux de travail pour calibrer notre caméra stéréo en utilisant l'application:

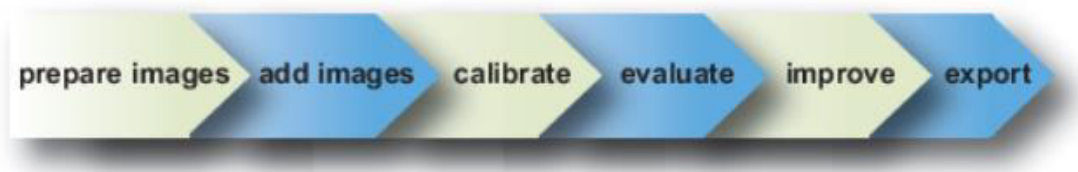


Figure 3. 9 Image montrant les étapes de calibration

1. Préparer des images, une caméra et un motif d'étalonnage ;
2. Ajouter des paires d'images ;
3. Calibrer la caméra stéréo ;
4. Évaluer la précision d'étalonnage ;
5. Ajuster les paramètres pour améliorer la précision (si nécessaire) ;
6. Exporter l'objet de paramètres ;
7. Dans certains cas, les valeurs par défaut fonctionnent bien et on n'a pas besoin d'apporter d'améliorations avant d'exporter les paramètres.

3.4.1.2 Ouverture du calibrateur de caméra stéréo

- Barre d'outils MATLAB: Dans l'onglet Applications, dans la section Traitement de l'image et Vision par ordinateur, on clique sur l'icône Stereo Camera Calibrator.
- Commande MATLAB: Entrez stereoCameraCalibrator

3.4.1.3 Préparation d'un motif, une caméra et des images

Pour améliorer les résultats, on utilise entre 10 et 20 images du modèle d'étalonnage. Le calibrateur nécessite au moins trois images. Des images non compressées ou des formats de compression sans perte tels que PNG sont utilisées. Le modèle d'étalonnage et la configuration de la caméra doivent satisfaire à un ensemble d'exigences pour fonctionner avec le calibrateur. Pour une meilleure précision d'étalonnage, suivre ces instructions pour préparer le motif, configurer l'appareil photo et capturer les images.

3.4.1.4 Choix de la forme du motif en damier (carré)

L'application Camera Calibrator utilise un motif en damier. Un motif en damier est une cible d'étalonnage pratique. Si l'on souhaite utiliser un motif différent pour extraire les points clés, on peut utiliser les fonctions MATLAB d'étalonnage de la caméra directement. Voir *Étalonnage de la caméra unique* pour la liste des fonctions. On peut imprimer (à partir de MATLAB) et utiliser le motif en damier fourni. Le motif en damier que nous utilisons ne doit pas être carré. Un côté doit contenir un nombre pair de carrés et l'autre côté doit contenir un nombre impair de carrés. Par conséquent, le motif contient deux coins noirs le long d'un côté et deux coins blancs du côté opposé. Ce critère permet à l'application de déterminer l'orientation du motif. Le calibrateur affecte le côté le plus long à la direction x.



Figure 3. 10 Choix de la forme du motif en damier (carré)

3.4.1.5 Choix des primitives du motif en damier

1. Fixer l'impression en damier sur une surface plane. Les imperfections sur la surface peuvent affecter la précision de l'étalonnage.
2. Mesurer un côté du carré en damier. On a besoin de cette mesure pour l'étalonnage. La taille des carrés peut varier en fonction des paramètres de l'imprimante.

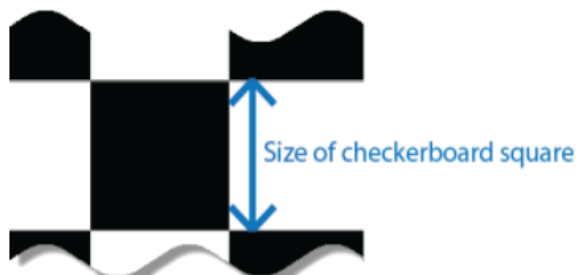


Figure 3. 11 Choix des primitives du motif en damier

3. Pour améliorer la vitesse de détection, configuration du motif avec le plus petit fouillis d'arrière-plan possible.

3.4.1.6 Configuration de la caméra

Pour étalonner la caméra, il faut suivre ces règles:

- Garder le motif au point, mais n'utiliser pas l'autofocus
- Si on modifie les paramètres de zoom entre les images, la distance focale change.

3.4.1.7 Capture des images

Pour de meilleurs résultats, on utilise au moins 10 à 20 images du motif d'étalonnage. Le calibrateur nécessite au moins trois images. Utilisation des images non compressées dans des formats de compression sans perte tels que le format PNG. Pour une plus grande précision d'étalonnage:

- Capturer les images du motif à une distance à peu près égale à la distance entre la caméra et les objets d'intérêt. Par exemple, si on prévoit de mesurer des objets à partir de 2 mètres, garder le motif à environ 2 mètres de la caméra.
- Placer le damier à un angle inférieur à 45 degrés par rapport au plan de la caméra.

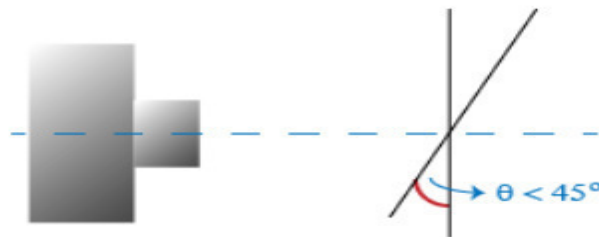


Figure 3. 12 Damier à un angle inférieur à 45 degrés par rapport au plan de la caméra.

- Ne pas modifier les images (par exemple, ne les recadrer pas).
- N'utiliser pas la mise au point automatique ou ne modifier pas les paramètres de zoom entre les images.
- Capturer les images d'un motif en damier à différentes orientations par rapport à la caméra.
- Capturer une variété d'images du motif afin que l'on ait pris en compte le plus possible de l'image. La
- La distorsion de l'objectif augmente radialement depuis le centre de l'image et n'est parfois pas uniforme à travers le cadre de l'image. Pour capturer cette distorsion de l'objectif, le motif doit apparaître près des bords des images capturées.

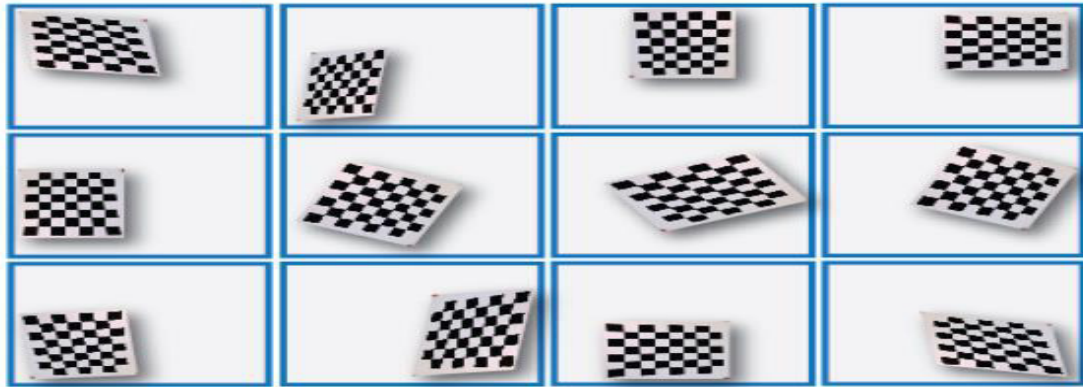


Figure 3.13 Statut des images prises pour calibration

- Assurons-nous que le motif en damier est entièrement visible dans les deux images de chaque paire stéréo.

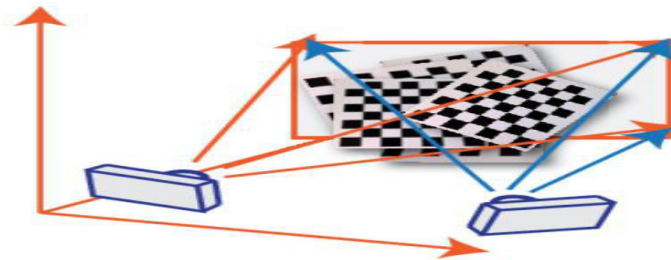


Figure 3.14 Réglage des positions de la paire de caméras

- Garder le motif stationnaire pour chaque paire d'images. Tout mouvement du motif entre la prise de l'image 1 et l'image 2 de la paire affecte négativement l'étalonnage.
- Créer un affichage stéréo ou anaglyphe en positionnant les deux caméras à environ 55 mm l'une de l'autre. Cette distance représente la distance moyenne entre les yeux humains.
- Pour une plus grande précision de reconstruction sur de longues distances, place les caméras plus loin.

3.4.2 Ajouter des paires d'images

Pour commencer l'étalonnage, on doit cliquer sur ajouter des images, en particulier deux jeux d'images stéréo du damier, un jeu pour chaque caméra.

3.4.2.1 Chargement des images

On peut ajouter des images de plusieurs dossiers en cliquant sur Ajouter des images dans la section Fichier de l'onglet Etalonnage. Puis, Sélectionner l'emplacement des images

correspondant à la caméra 1 à l'aide du bouton Parcourir, puis faites de même pour la caméra 2. Ensuite nous spécifions la taille du carré en damier en saisissant la longueur d'un côté d'un carré à partir du motif en damier.

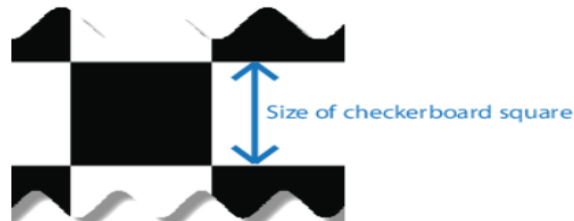


Figure 3. 15 Définition de la taille de la boîte d'échecs

La boîte d'échec est définie comme étant la primitive du damier de calibrage.

3.4.2.2 Analyse des images

Dix images sont prises à partir de la caméra de gauche pour une utilisation dans calibration



Figure 3. 16 Série d'images acquises pour la calibration (caméra gauche)

Dix images sont prises à partir de la caméra de droite pour une utilisation dans calibration



Figure 3. 17 Série d'images acquises pour la calibration (caméra droite)

Le calibrateur tente de détecter un damier dans chacune des images ajoutées, affichant une fenêtre de barre de progression Analyse des images, indiquant la progression de la détection. Si l'une des images est rejetée, la boîte de dialogue Résultats de la détection apparaît, qui contient des informations de diagnostic. Les résultats indiquent combien d'images totales ont été traitées et parmi celles traitées, combien ont été acceptées, rejetées ou ignorées. Le calibrateur ignore les images en double.

Les résultats de calibration sont présentés sur la figure ci-dessous :

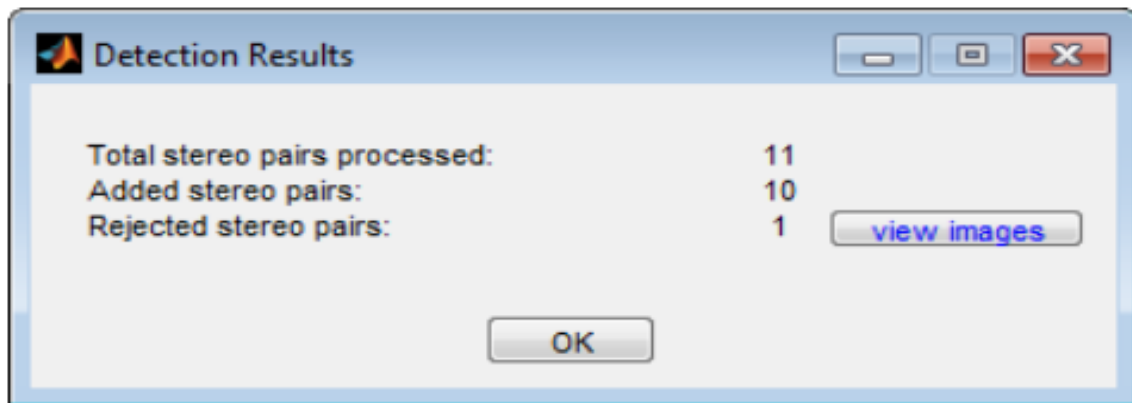


Figure 3. 18 Exemple de test sur la calibration

Le nombre totale d'images correctement calibrées comme le montre la figure est égale à 10 ; alors que celle que la calibration n'a pu prendre en charge est au nombre de un. Donc 10 images/ 11 ont été correctement calibrées en mode stéréo.

Pour afficher les images rejetées, cliquer sur Afficher les images. Le calibrateur rejette les images en double. Il rejette également les images où l'ensemble du damier n'a pas pu être détecté. Les raisons possibles de ne pas détecter sont une image floue ou un angle extrême du motif. La détection prend plus de temps avec des images plus grandes et avec des motifs contenant un grand nombre de carrés.

3.4.2.3 Visualisation des images et points détectés

Le volet Navigateur de données affiche une liste de paires d'images avec des ID. Ces paires d'images contiennent un motif détecté. Pour afficher une image, il faut la sélectionner dans le volet Navigateur de données.

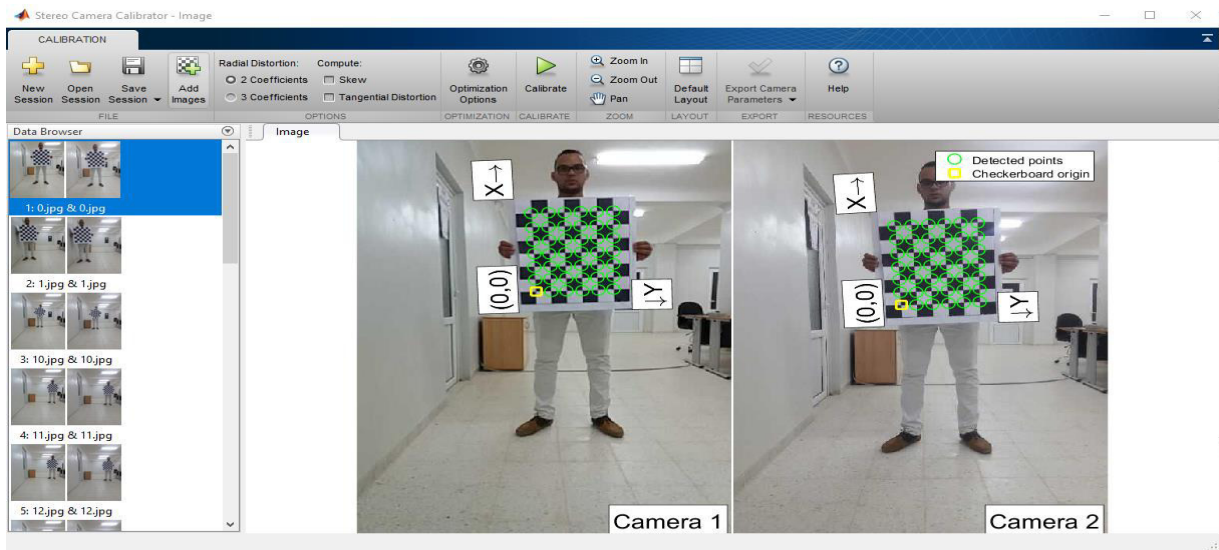


Figure 3. 19 Affichage des images avec des points détectés

Le volet Image affiche la paire d'images en damier sélectionnées avec des cercles verts pour indiquer les points détectés. On peut vérifier que les coins ont été détectés correctement à l'aide des commandes de zoom. Le carré jaune indique l'origine (0,0). Les flèches X et Y indiquent l'orientation des axes en damier.

3.4.2.4 Visualisation de la carte intrinsèques (MATLAB)

On peut choisir pour l'application de calculer les intrinsèques de la caméra ou on peut charger des intrinsèques fixes précalculés. Pour charger des intrinsèques dans l'application, il faut sélectionner Utiliser les intrinsèques fixes dans la section Intrinsèque de l'onglet Etalonnage. Les options Radial Distortion et Compute de la section Options sont désactivées lorsque l'on charge des intrinsèques. Pour charger des intrinsèques en tant que variables à partir de l'espace de travail, il faut cliquer sur LoadIntrinsics.

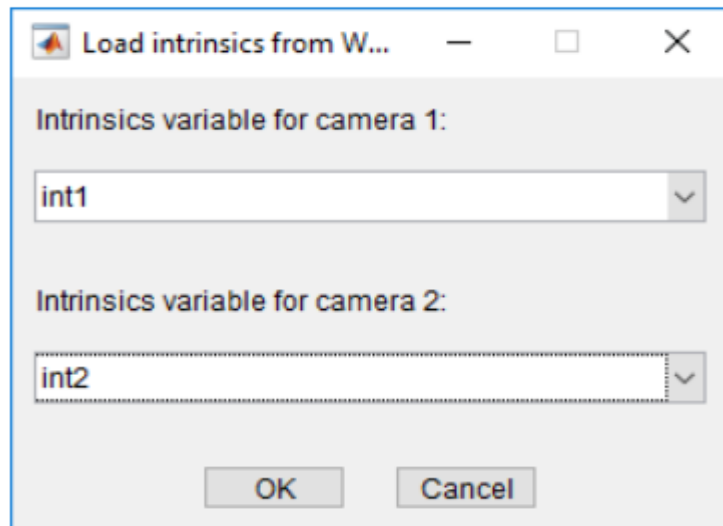


Figure 3. 20 Sélection des images prises à partir des caméras 1 et 2

3.4.3 Mise en oeuvre de la calibration

Une fois que les paires d'images sont acceptées (correctement calibrées), il faut cliquer sur le bouton Calibrer dans l'onglet Calibration. Les paramètres de calibrage par défaut supposent l'ensemble minimum de paramètres de la caméra. Commençons par exécuter le calibrage avec les paramètres par défaut. Après avoir évalué les résultats, on peut essayer d'améliorer la précision de l'étalonnage en ajustant les paramètres et en ajoutant ou en supprimant des images, puis procéder à un nouveau calibrage.

Optimisation : Lorsque la distorsion de l'objectif est grave, l'application peut ne pas pouvoir calculer les valeurs initiales des intrinsèques de la caméra. Si on possède les spécifications du fabricant pour l'appareil photo et que l'on connaît la taille de pixel, la distance focale ou les caractéristiques de l'objectif, on peut définir manuellement des suppositions initiales pour l'intrinsèque de la caméra et la distorsion radiale. Pour définir les suppositions initiales, cliquer sur Options > Options d'optimisation.

- Cocher la case du haut, puis entrer une matrice 3 par 3 pour spécifier les intrinsèques initiaux. Si on ne spécifie pas de supposition initiale, la fonction calcule la matrice intrinsèque initiale en utilisant les moindres carrés linéaires.
- Sélectionner la case à cocher inférieure, puis entrer un vecteur à 2 ou 3 éléments pour spécifier la distorsion radiale initiale. Si on ne fournit pas de valeur, la fonction utilise 0 comme valeur initiale pour tous les coefficients.

3.4.4 Évaluation des résultats de calibration

On peut évaluer la précision de l'étalonnage en examinant les erreurs de reprojection, en examinant les extrinsèques de la caméra ou en visualisant l'image non déformée. Pour obtenir les meilleurs résultats d'étalonnage, utilisez les trois méthodes d'évaluation.

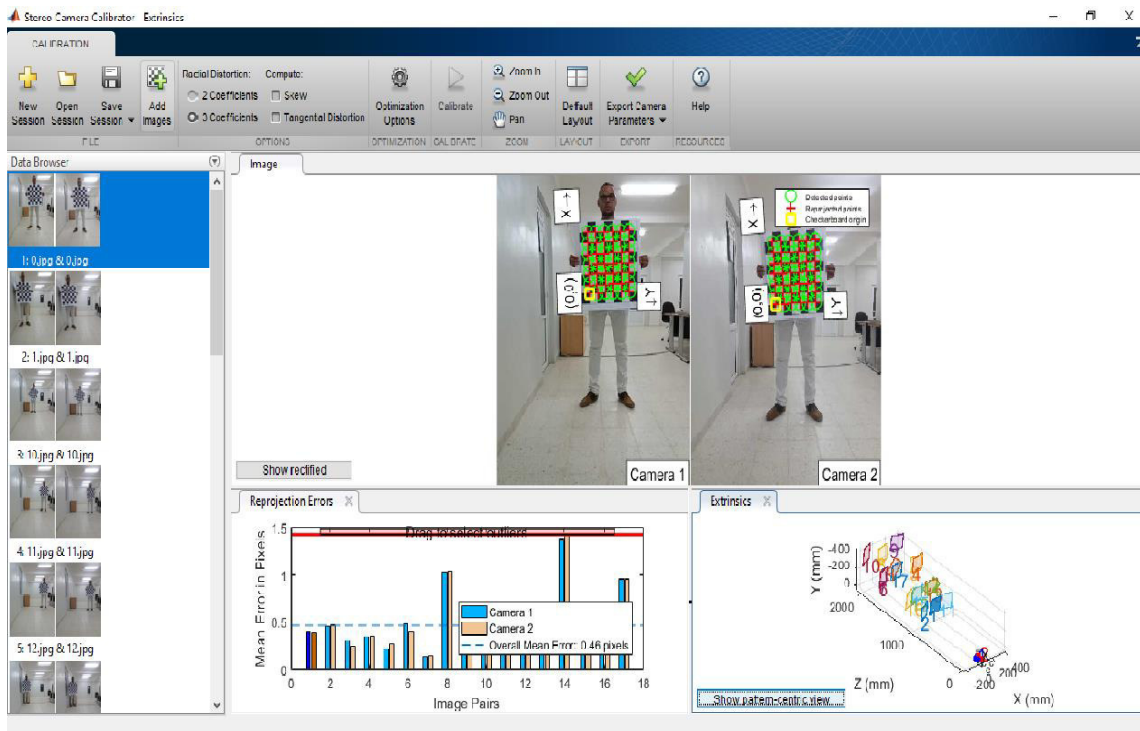


Figure 3. 21 Évaluation des résultats de calibration

3.4.4.1 Examen des erreurs de reprojection

Les erreurs de reprojection sont les distances, en pixels, entre les points détectés et les points reprojétés. L'application Stereo Camera Calibrator calcule les erreurs de reprojection en projetant les points du damier à partir des coordonnées du monde, définies par le damier, en coordonnées de l'image. L'application compare ensuite les points reprojétés aux points détectés correspondants. En règle générale, les erreurs moyennes de reprojection de moins d'un pixel sont acceptables.

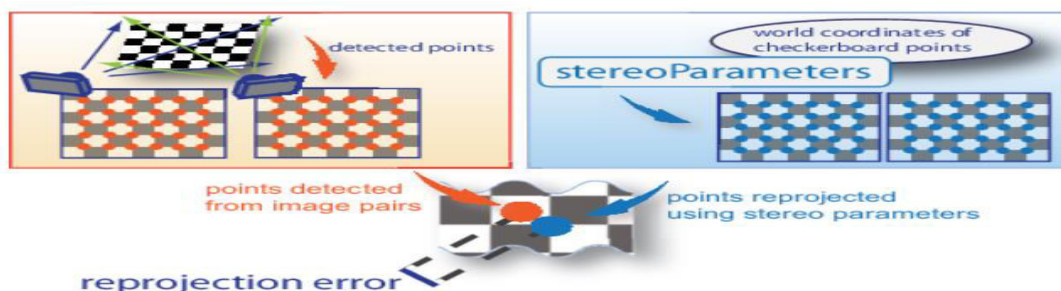


Figure 3. 22 Vérification des erreurs de photo

L'application Stereo Calibration App affiche, en pixels, les erreurs de reprojection sous forme de graphique à barres. Le graphique vous aide à identifier les images qui contribuent de manière négative à l'étalonnage. Il faut sélectionner l'entrée du graphique à barres et supprimer l'image de la liste des images dans le volet Explorateur de données.

3.4.4.2 Erreurs de projection (graphique à barres)

Le graphique à barres affiche l'erreur moyenne de reprojection par image, ainsi que l'erreur moyenne globale. Les étiquettes de barre correspondent aux ID d'image. Les barres en surbrillance correspondent à la paire d'images sélectionnée.

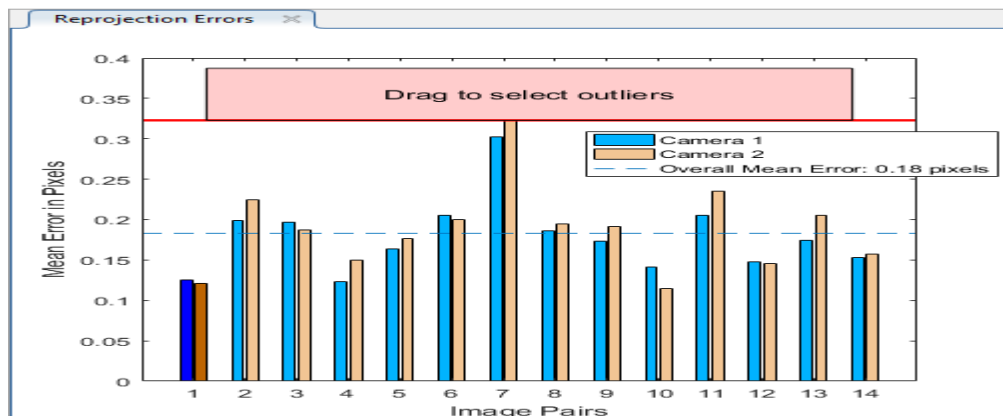


Figure 3. 23 Graphique à barres des erreurs de projection

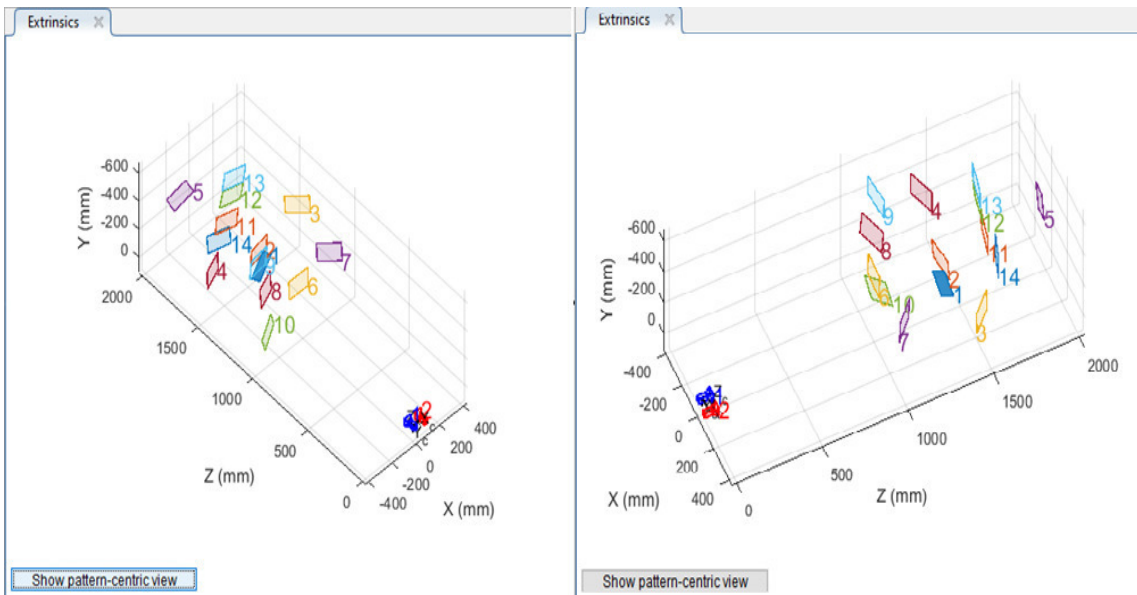
Sélection d'une paire d'images de l'une des manières suivantes:

- En cliquant sur la barre correspondante dans le graphique.
- Sélection de la paire d'images dans la liste du volet Navigateur de données.
- Ajustement de l'erreur moyenne globale. (Cliquez et faites glisser la ligne rouge vers le haut ou vers le bas pour sélectionner les images aberrantes).

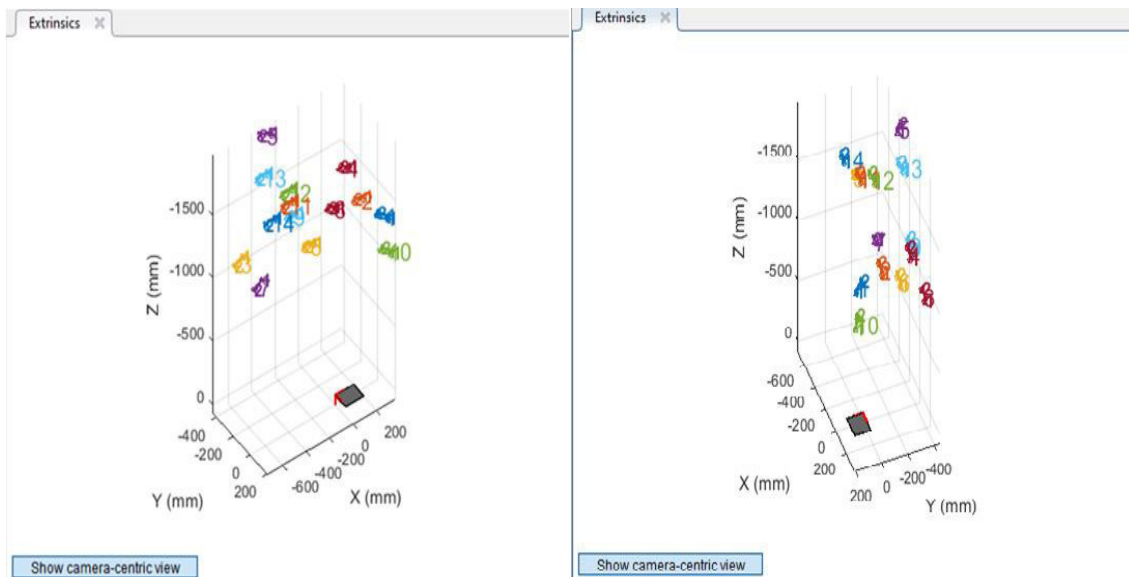
3.4.4.3 Visualisation des paramètres extrinsèques

Le tracé des paramètres extrinsèques 3D fournit une vue centrée sur la caméra des motifs et une vue centrée sur les motifs de la caméra. La vue centrée sur la caméra est utile si l'appareil photo était à l'arrêt lorsque les images ont été capturées. La vue centrée sur les motifs est utile si le motif est stationnaire. On peut cliquer sur le curseur et maintenir le bouton de la souris enfoncé avec l'icône de rotation pour faire pivoter la figure. Cliquer sur un damier (ou une caméra) pour le sélectionner. Les données en surbrillance dans les visualisations correspondent à l'image sélectionnée dans la liste. Examiner les positions

relatives du motif et de la caméra pour déterminer si elles correspondent à ce que l'on attend. Par exemple, un motif qui apparaît derrière la caméra indique une erreur de calibrage.



(a)



(b)

Figure 3. 24 Représentation en 3D de la position des caméras et des images acquises en représentation (a) horizontale et (b) verticale

3.4.4.4 Affichage des images rectifiées

Pour afficher les effets de la correction stéréo, il faut cliquer sur Afficher rectifié dans la section Vue de l'onglet Calibration. Si l'étalonnage est précis, les images ne sont pas déformées et sont alignées. Ensuite, vérifier les images rectifiées, c'est important même si les erreurs de reprojection sont faibles.

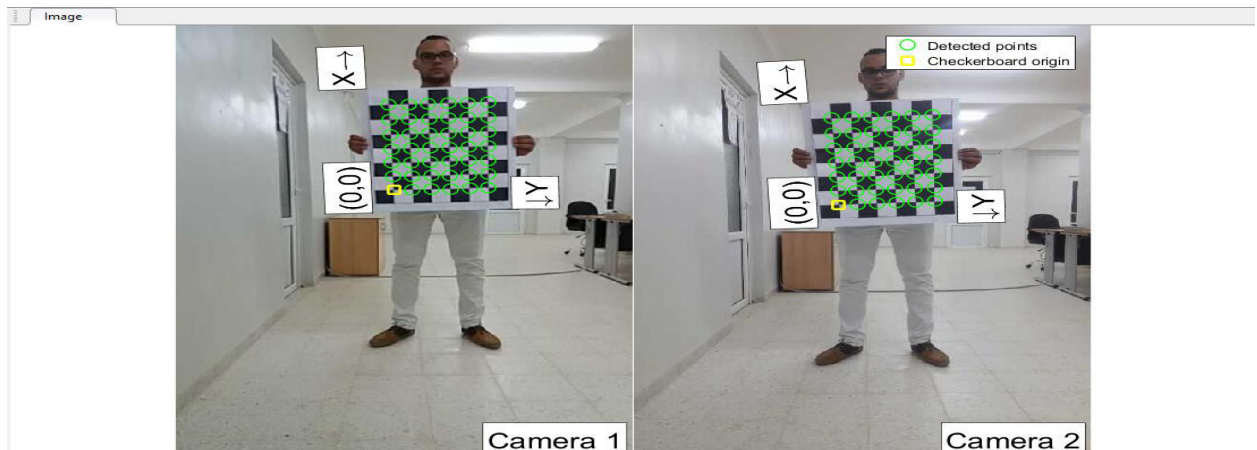


Figure 3. 25 Affichage des images rectifiées

Par exemple, si le motif ne couvre qu'un faible pourcentage de l'image, l'estimation de distorsion peut être incorrecte, même si le calibrage a entraîné quelques erreurs de reprojection. L'image suivante montre un exemple de ce type d'estimation incorrecte pour un calibrage unique.

3.4.5 Amélioration de calibration

Pour améliorer le calibrage, on peut supprimer les paires d'images à haute erreur, ajouter plus de paires d'images ou modifier les paramètres du calibrateur.

Ajout ou suppression des images

Envisager d'ajouter plus d'images si:

- On a moins de 10 images.
- Les motifs ne couvrent pas assez le cadre de l'image.
- Les motifs n'ont pas une variation d'orientation suffisante par rapport à la caméra.

Envisagez de supprimer des images :

- Les images ont une erreur moyenne de reprojection élevée.
- Les images sont floues.

- Les images contiennent un damier à un angle supérieur à 45 degrés par rapport au plan de la caméra voir figure 3. 26 .

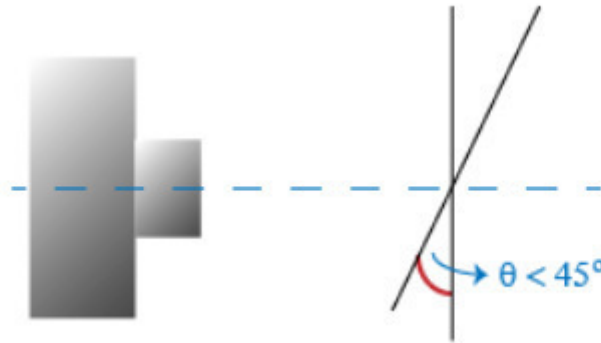


Figure 3. 26 Illustration du réglage du damier pour la calibration

- Les images contiennent des points en damier détectés de manière incorrecte.

3.4.6 Exporter les paramètres de la caméra

Lorsque on est satisfait de la précision de l'étalonnage, il faut cliquer sur Exporter les paramètres de la caméra. On peut enregistrer et exporter les paramètres de la caméra vers un objet en sélectionnant Exporter les paramètres de la caméra ou générer les paramètres de la caméra en tant que script MATLAB.

3.4.6.1 Exportation des paramètres de la caméra

Sélectionner Exporter les paramètres de caméra> Exporter les paramètres dans Workspace pour créer un objet stereoParameters dans votre espace de travail. L'objet contient les paramètres intrinsèques et extrinsèques de la caméra et les coefficients de distorsion. On peut utiliser cet objet pour diverses tâches de vision par ordinateur, telles que la distorsion d'image, la mesure d'objets planaires et la reconstruction 3D. Voir Mesure d'objets planaires avec une caméra calibrée. On peut éventuellement exporter l'objet stereoCalibrationErrors, qui contient les erreurs standard des paramètres de caméra stéréo estimés, en cochant la case Exporter les erreurs d'estimation.

3.4.6.2 Génération d'un script MATLAB

Sélectionner Exporter les paramètres de caméra> Générer un script MATLAB pour enregistrer les paramètres de votre caméra dans un script MATLAB, ce qui vous permet de reproduire les étapes de votre session d'étalonnage.

3.5 Localisation

Après exécution du programme les résultats de calibration et localisation obtenus sont présentés sur les figures se trouvant dans la section suivante.

3.5.1 Chargement des paramètres de la caméra stéréo

Charger l'objet stereoParameters, qui résulte de calibration de la caméra à l'aide de l'application stereoCameraCalibrator ou de la fonction estimateCameraParameters.

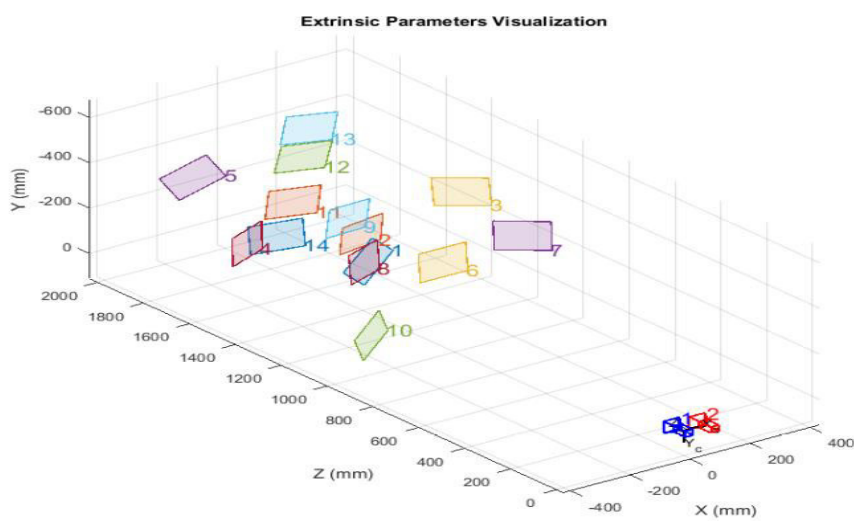


Figure 3. 27 Résultat de calibration de la caméra

3.5.2 Création des lecteurs de fichiers vidéo et le lecteur vidéo

Créer des objets système pour lire et afficher la vidéo.

3.5.3 Lecture et rectification des images vidéo

Les images des caméras gauche et droite doivent être rectifiées afin de calculer la disparité et reconstruire la scène 3D. Les images rectifiées ont des lignes épipolaires horizontales et sont alignées en ligne. Cela simplifie le calcul de la disparité en réduisant l'espace de recherche pour les points correspondants à une dimension. Les images rectifiées peuvent également être combinées en un anaglyphe, qui peut être vu en utilisant les lunettes stéréo rouge-cyan pour voir l'effet 3D.



Figure 3. 28 Lecture et rectification les images vidéo et reconstruire la scène 3D

3.5.4 Calcul de la disparité

Dans les images stéréo rectifiées, toute paire de points correspondants est située sur la même rangée de pixels. Pour chaque pixel de l'image de gauche, on doit calculer la distance au pixel correspondant dans l'image de droite. Cette distance est appelée la disparité, et elle est proportionnelle à la distance du point du monde correspondant de la caméra.

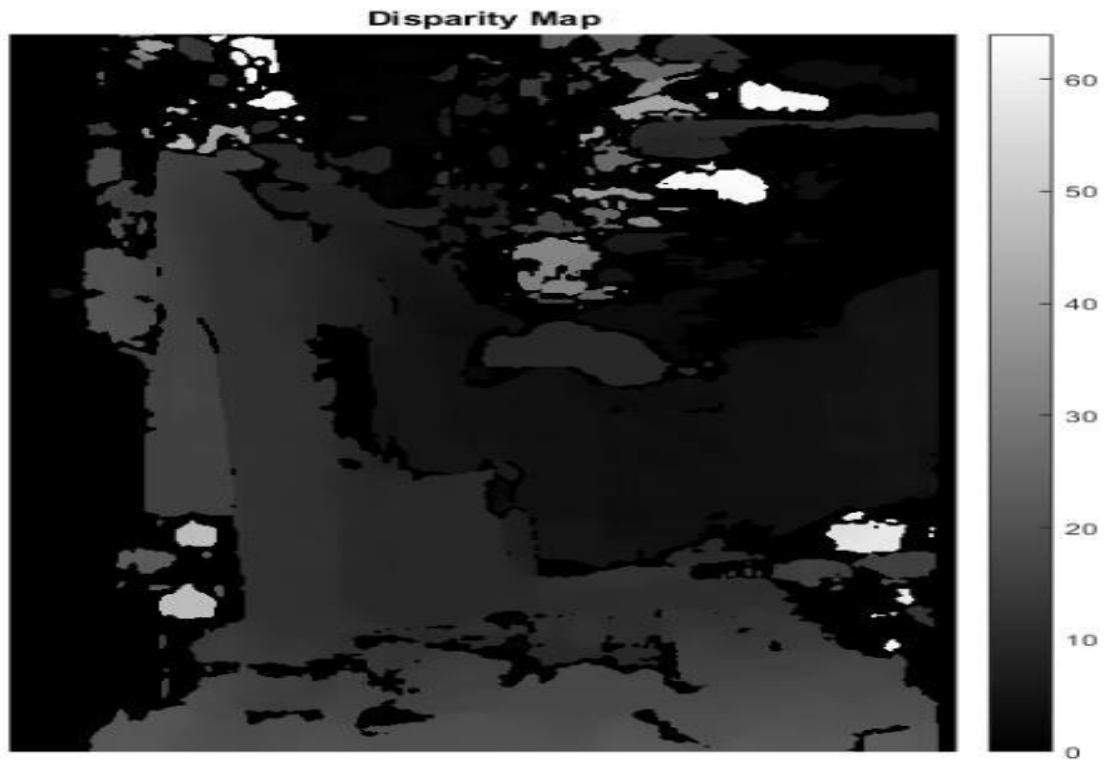


Figure 3. 29 Carte de disparité

3.5.5 Reconstruction de la scène 3D

Reconstruire les coordonnées du monde 3D des points correspondant à chaque pixel de la carte de disparité.

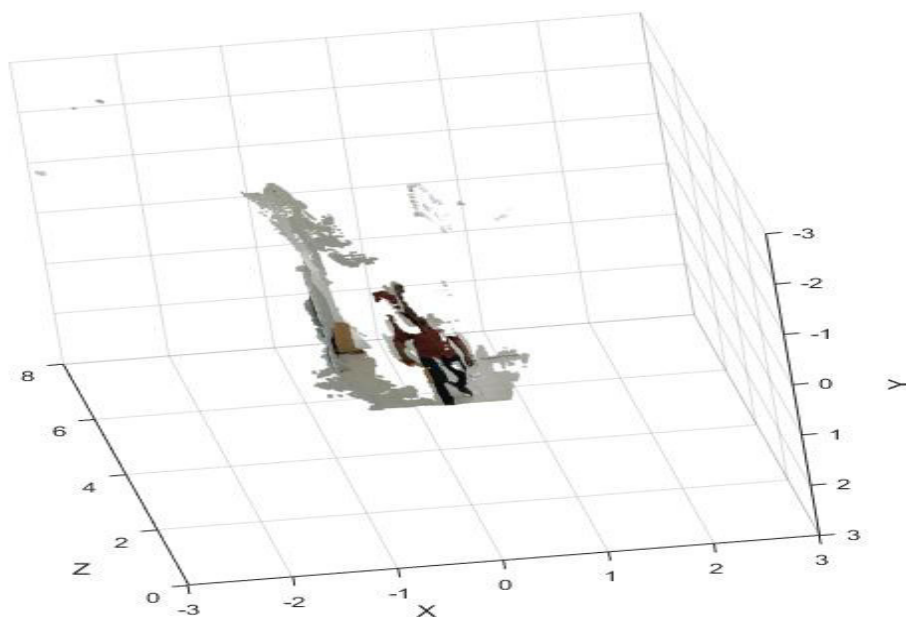


Figure 3. 30 Reconstruction de la scène 3D



Figure 3. 31 Localisation de la 1^{ère} image de la séquence vidéo

3.5.6 Détection et avertissement

C'est une étape d'avertissement où une personne est détectée avec une trace et un avertissement est émis si elle s'approche de la zone d'accès restreint.



Figure 3. 32 Exemple de détection, localisation et avertissement

3.6 Résultats finaux sur séquence vidéo

Les images successives suivantes représentent les résultats obtenus sur une sequence video:













Discussion :

D'après toutes les expériences menées, nous confirmons que la localisation indoor est opérationnelle et donne de bons résultats. Le détecteur HOG est retenu car il est insensible à toutes les variantes du milieu indoor incontrôlé, notamment l'illumination, la distance, la mobilité de la personne à détecter (localiser) et les arrières plan. Le mouvement de la personne n'affecte pas notre approche de détection et de localisation indoor. Nous rajoutons que la localisation de personne dans l'image stéréo ou séquences vidéo est meilleure dans le cas de l'acquisition par smartphone, ceci est dû à la bonne résolution des caméras.

Conclusion

Nous avons proposé une architecture et implémenté un système qui rend possible l'intégration de technologies de localisation à différents niveaux d'abstraction. Le cas particulier de l'intégration multi-technologies présenté ici peut paraître spécifique aux technologies proposées, mais il peut en fait être généralisé aux technologies fournissant des données avec des propriétés similaires, correspondant à des modèles homologues de l'espace. Finalement, les technologies de localisation devraient être vues comme des services logiciels composables, en cohérence avec le paradigme des architectures orientées services.

Les applications basées sur la localisation pourraient donc utiliser, dans un périmètre donné, une combinaison de solutions centralisées à large échelle comme les systèmes satellitaires ou cellulaires terrestres, avec, potentiellement, d'autres services de localisation indépendants qui pourrait leur être composés dynamiquement. Une infrastructure logicielle locale de localisation telle que décrite ci-dessus est une solution possible pour consolider cette intégration au niveau d'abstraction le plus élevé, leur combinaison aux niveaux inférieurs étant réalisable suivant le modèle.

Conclusion générale

Ce travail s'inscrit dans le contexte de la localisation des terminaux mobiles pour les environnements indoor. L'utilisation des réseaux WiFi (utilisés normalement pour communiquer entre différents équipements) s'est faite naturellement car ces réseaux deviennent de plus en plus pervasifs et des informations utilisables à des fins de localisation sont disponibles dans ces réseaux. L'exploitation de l'information de puissance du signal s'est faite à l'aide de la technique dite du fingerprinting qui nécessite de disposer d'une base de données représentative de la couverture radio pour les différents points d'accès présents. La disponibilité du plan du bâtiment est une autre information sur laquelle on peut s'appuyer puisqu'elle permet d'effectuer un "map matching" utile pour délivrer à l'application finale une trajectoire conforme à celle que l'utilisateur peut avoir suivi. L'exploitation d'estimateurs comme le filtrage particulaire permet de déterminer un optimum en fonction des mesures WiFi, du plan du bâtiment et du modèle de trajectoire choisi pour le mouvement. L'exploitation de ces différents éléments permet d'aboutir à des précisions de localisation comprises entre 1.5 et 2 m d'erreur moyenne. Cette précision est suffisante pour les applications visant à déterminer la position d'une personne. Connaissant cette information, une seconde personne peut facilement retrouver son correspondant. Cette erreur est intéressante en terme de précision, mais en terme de justesse, elle est limitative. En effet, lorsque la personne rase les murs, une incertitude subsiste sur le fait qu'elle soit dans le couloir ou plutôt dans la pièce, notamment dans les situations statiques. Afin de satisfaire le critère de justesse, il est nécessaire d'avoir une technologie de localisation dont la précision est de l'ordre de la dimension des parois de l'environnement. Ceci n'est pas réaliste pour WiFi. Aujourd'hui, les activités de recherche doivent se tourner vers l'allègement de la procédure de mise en place de la base de données nécessaire à la technique du fingerprinting. À l'usage, la base de données, même si les opérations sont simples, reste le frein au déploiement à grande échelle de cette technique de localisation. Des outils de modélisation de la propagation des ondes peuvent être utilisés, mais nécessitent tout de même une étape de calibration souvent non négligeable. L'étude portant sur la localisation en ULB a montré la capacité de cette technologie à fournir une localisation dont la précision est de l'ordre de 50 cm. Cette précision est obtenue grâce à un système détectant l'instant d'arrivée du premier trajet sur la réponse impulsionnelle du canal radio (obtenue à l'aide d'un code d'étalement – sondage de canal). La nécessité de rechercher le premier trajet vient du fait que le système est déployé à l'intérieur des bâtiments dont la structure génère de nombreux multi-trajets. La

limitation du système proposé ici provient du faible niveau de puissance (limité par la densité spectrale de puissance et la répétition dans le temps des impulsions) autorisé en émission. Cette forte contrainte imposée limite la portée radio des équipements. Aussi, la zone couverte par le système de localisation par ULB est limitée à 20 m, ce qui est peu pour un système de localisation. Pour couvrir un bâtiment plus important, il est nécessaire de densifier le réseau de capteurs et d'avoir un appareil pour synchroniser l'ensemble des voies d'acquisition. La mise en place de ce système devient vite fastidieuse et coûteuse. L'exploitation des réseaux de capteurs sera certainement une approche à privilégier, chacun de ces capteurs étant doté de capacités à estimer la distance qui le sépare des autres capteurs à portée radio. Une nouvelle stratégie de détection de ce premier trajet sera à mettre en place, car l'échantillonnage, tel qu'il est présenté dans ces travaux, est une contrainte pour des équipements de type capteurs à ressources d'énergie et mémoire de stockage limitées. Néanmoins, on montre que de très bonnes performances en termes de localisation (statique et dynamique) sont atteintes, ce qui reste très encourageant pour la mise en place des capteurs ULB. Chaque méthode de localisation possède des avantages et des limitations qui ne sont pas les mêmes. La limitation pour WiFi est la constitution de la base de données et la méthode employée ne permet d'atteindre que des précisions de l'ordre de 1.5 à 2 m sur une zone relativement vaste (la portée d'un point d'accès étant d'une cinquantaine de mètres en indoor), alors que la technologie ULB permet d'obtenir une localisation très précise, mais sur une zone restreinte (portée de 20 m au maximum).

Aujourd'hui, de plus en plus de pouvoirs locaux ont recours aux caméras (les vidéosurveillances) pour surveiller leur territoire et le cas échéant, réagir de manière appropriée. La multitude de caméras génère toutefois un grand nombre d'images. Le suivi 'en direct' de toutes ces images requiert dès lors la présence d'une grande équipe d'opérateurs mais il n'est pas toujours évident de libérer l'effectif nécessaire. En réponse à ce problème, apparaissent de plus en plus de fonctions de caméras intelligentes qui facilitent le travail des opérateurs en montrant des situations, des personnes ou des véhicules suspects. L'installation d'un tel système nécessite un investissement supplémentaire conséquent dans le chef des villes et communes.

Dans ce travail, nous nous intéressons à deux des problèmes majeurs d'analyse d'image et de séquence vidéo ainsi que la localisation par les algorithmes du traitement d'images sur images fixes et dans une séquence d'images par des caméras fixes. Pour aboutir à notre objectif qui

est la localisation indoor , la détection et le suivi d'objets en mouvement dans des scènes simples ont été réalisés. Cette première partie est fondamentale et essentielle dans notre travail.

Dans la première partie de ce mémoire nous présentons quelques concepts et définitions de la vidéo surveillance, et dans une deuxième partie nous proposons les méthodes de détection d'objets, la troisième partie présente l'ensemble des résultats expérimentaux obtenus pour la détection et le suivi d'objets en mouvement avec un détail sur la phase de calibration fondamentale à la localisation. Dans le cadre de ce mémoire et après une étude bibliographique sur certaines méthodes se rapportant à la vidéosurveillance et à la détection. Nous avons étudié et conçu quatre méthodes de détection des plus populaires et performantes de l'état de l'art dans notre cas d'application. Nous avons choisi la meilleure pour notre type d'application indoor expérimenté sur une série d'images et vidéo en mono et stéréo en changeant le type de capteurs (webcam, caméras binoculaires et caméras de deux smartphones). La détection HOG (Histogram Oriented Gradient) s'est avérée la meilleure et ses résultats sont excellents. Pour cela, nous l'avons choisie pour la détection des images de personnes fixes ou en mouvement. Elle s'affranchit de la soustraction de l'arrière-plan d'une image dans le processus de détection du mouvement. Elle trouverait donc, son application pour des scènes où l'arrière-plan change d'une image à l'autre à cause du mouvement de la caméra. Notre approche se base essentiellement sur une bonne détection et calibration de caméra(s) fixe(s) qui nous permet de détecter les personnes indoor. Nous avons pu valider la méthode de détection de mouvement à l'aide d'un algorithme pour détecter les objets abandonnés en mouvement. Cette méthode a fourni de bons résultats et des objets en mouvement ont pu être détectés à travers la séquence d'images. L'application de cette méthode sur des images vidéos présentant différents problèmes habituels tels que la forte compression, le flou provoqué par des mouvements rapides, la présence de reflets lumineux sur certaines surfaces..., constituerait un défi pour la robustesse de cette méthode, qu'il n'est pas impossible de relever. Nous avons fait des expériences sur des images fixes avec différentes variantes (pose, illumination, éloignement caméra, profil, expression, occlusions...) et sur vidéos et les résultats sont prometteurs. Il s'agirait de déterminer adéquatement certains paramètres comme la bonne vitesse d'obturation pour avoir une meilleure qualité d'images et donc une meilleure extraction des points d'intérêts, comme nous avons pu le démontrer à travers les images. Nous avons amélioré nos résultats par l'application de la 3D en utilisant 2 caméras et coordonnant les données par des images stéréos contribuant ainsi à l'avancement du projet. L'expérience était effectuée et les résultats

d'étalonnage des caméras pour calculer la structure 3D a été obtenue. Après plusieurs expériences nous avons constaté que les augmentations d'erreur de position est due à la caractéristique de la caméra : posture relative des caméras, résolution d'image, et la distorsion optique peut entraîner du bruit.

Nous pouvons conclure que pour appliquer la reconstruction 3D des images stéréo des objets (personnes) : la technologie des caméras, les algorithmes de stabilisation sur la reconnaissance automatique d'objets, le processus de réduction du développement de bruit optique et l'amélioration des performances matérielles sont nécessaires.

En perspectives, nous proposons de :

1. Chercher une méthode unique pour la détection et le suivi d'objets mobiles pour les deux types de mouvement (lent et rapide).
2. Appliquer des méthodes de détection et de suivi sur les séquences à fonds dy

Références Bibliographique

- [1] Boudjellal. S. Détection et identification de personne par méthode biomé-trique. Mémoire de Magister : Electronique-Téledétection. Tizi-ouzou : Université Mouloud Mammeri, 95p.
- [2] Ming-Hsuan Yang, David J. Kriegman et Narendra Ahuja. Detecting faces images : A survey. Dans IEEE Transaction on Pattern Analysis and Machine Intelligence, 2002, vol.24(1), 34-58.
- [3] Commercial Security Devices[en ligne] (page consultée le 01/10/2013) .
<http://www.commercialSecuritydevises.com/fr/détection-de-visage.html>
- [4] Cheng-Chin, Wen-Kai Tai, Mau-Tsuen Yang, Yi-Ting Huang, and chi-Janng Huang. A novel method for detecting lips, eyes and faces in real time. Real-Time Imaging, 9(4) : 277-287, 2003.
- [5] Wenlong Zheng and Suchendra M. Bhandarkar. Face detection and haking using a boosted adaptative particle filter. Journal of visual communication and Imog Representation, 20(1) : 9-27, 2009.
- [6] H. A. Rowley, S. Baluja, and T. Kanade. Neural Network-based face detection. IEEE Transachen on Pattern Analysis and Machine Intelligence, 20(1) : 23-38, 1998.
- [7] H. Schneidermand and T. Kanade. Probabilistic modeling of local appearance and spacial relation ships for object recognition. Computer Vision and Pattern Recognition, IEEE Computer Society Conférence on 0 :45, 1998.
- [8] Paul Viola and Michael Jones. Robust real-time object detection. In Second international work shop on statistical and computation atheories of vision, Vancouver, Canada, July 13 2001.
- [9] M.Van Wambeke, Reconnaissance et suivi de visages et implémentation en robotique temps-réel. Master Ingénieur Civil en Génie Biomédical. Belgique. Université !catholique de Lovain. 2010, 93p

[10] <https://www.timcod.fr/solutions/geolocalisation-indoor-outdoor/>

[11] Technologies de géolocalisation indoor Orange Labs Frédéric EVENNOU, Recherche & Développement 14/10/2010

[12] <https://blog.clever-age.com/fr/2015/05/21/les-technologies-de-geolocalisation-indoor/>

[13] <http://slideplayer.fr/slide/5411863/>

[14] Fishkin, K., Jiang, B; Philipose M. and Roy, S. "Unobtrusive detection of interactions with RFID-tagged object", UbiComp, Nottingham, September 2004

[15] Schwerdt K. Crowley, J.L., "Robust Face Tracking using color", 4th IEEE International Conference on Automatic Face and Gesture Recognition", Grenoble, France, March 2000

[16] Indrabayu 1, Rizki Yusliana Bakti , Intan Sari Areni, A. Ais Prayogi Informatics Study Program Electrical Engineering Study Program Hasanuddin University Makassar, Indonesia

[17] D. Djamaluddin, T. Indrabulan, Andani, Indrabayu, and S. W.Sidehabi, "The simulation of vehicle counting system for trafficsurveillance using Viola Jones method." Makassar Int. Conf. Electr.Eng. Inform. MICEEI, pp. 130 – 135, 2014.

[18] Indrabayu, Basri, A. Achmad, I. Nurtanio, and F. Mayasari, "BlobModification in Counting Vehicles using Gaussian Mixture Models Under Heavy Traffic." Asian Res. Publ. Netw. ARPN, vol. 10, 2015.

[19] W. L. Khong, W. Y. Kow, H. T. Tan, H. P. Yoong, and K. T. K. Teo, "Kalman Filtering Based Object Tracking in Surveillance VideoSystem," CUTSE Int. Conf., 2011.

[20] C. S. Rao and P. Darwin, "Frame Difference and Kalman Filter Techniques for Detection of Moving Vehicles in VideoSurveillance," *Int. J. Eng. Res. Appl. IJERA*, vol. 2, no. 6, pp. 1168–1170, 2012.

- [21] S. Han, and N. Vasconcelos, “Object-Based Regions of Interest for Image Compression,” University of California, San Diego.
- [22] A. Nurhadiyatna, B. Hardjono, A. Wibisono, I. Sina, W. Jatmiko, M.A. Ma’sum, and P. Mursanto, “Improved Vehicle Speed Estimation Using Gaussian Mixture Model and Hole Filling Algorithm,” *ICACSYS*, 2013.
- [23] I. A. Siradjuddin, M. R. Widyanto, T. Basaruddin, “Particle Filter with Gaussian Weighting for Human Tracking.” *TELKOMNIKA*, vol. 10, 2012.
- [24] H. S. Parekh, D. G. Thakore, and U. K. Jaliya, “A Survey on Object Detection and Tracking Methods,” *Int. J. Innov. Res. Comput. Commun. Eng. IJIRCCE*, vol. 2, no. 2, 2014.
- [25] Lim R., Sutjiadi R. and Setyati E. 2011. “Adaptive Background Extraction-Gaussian Mixture Models Method for Vehicle Counting Application in video base”. *National Conference of Informatics Engineering and Information System (SeTISI)*. Pp 19. in Bahasa.
- [26] https://fr.mathworks.com/examples/computer-vision/mw/vision-ex35748399-detecting-cars-using-gaussian-mixture-models?s_tid=srchtitle
- [27] Novelty detection and segmentation based on Gaussian mixture models: A case study in 3D robotic laser mapping Drews Jr. P., Nunez P., Rocha R.P., Campos M., Dias J. (2013) *Robotics and Autonomous Systems*, 61 (12) , pp. 1696-1709.
- [28] https://fr.wikipedia.org/wiki/Méthode_de_Viola_et_Jones
- [29] Nam, W., Dollár, P., Han, J.H.: Local decorrelation for improved pedestrian detection. In: *Advances in Neural Information Processing Systems*, pp. 424–432 (2014)
- [30] P. Dollár, R. Appel, S. Belongie, P. Perona, Fast feature pyramids for object detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 36 (8) (2014) 1532–1545.

[31] https://fr.wikipedia.org/wiki/Histogramme_de_gradient_orienté

[32] Dirk Schulz, Wolfram Burgard, Dieter Fox, and Armin B. Cremers. People tracking with a mobile robot using sample-based joint probabilistic data association filters. February 2003.

[33] Blue Eye Video. Blue Eye Video, video intelligence solutions. Website, <http://www.blueeyevideo.com/main.php>, 2003.

[34] Dirk Schulz and Dieter Fox. Bayesian color estimation for adaptive vision-based robot localization. Proceedings of IROS, 2004.

[35] https://fr.wikipedia.org/wiki/Caméra_IP

[36] <https://ip-webcam.fr.uptodown.com/android>

[37] <https://fr.mathworks.com/help/vision/ug/stereo-camera-calibrator-app.html>