Université Mohamed Khider de Biskra
Faculté des Sciences et de la Technologie
Département de Génie Electrique

MÉMOIRE DE MASTER

Sciences et Technologies
Filière : Télécommunication
Spécialité : Réseaux et Telecommunication

Réf. : .....

Présenté et soutenu par :
GHORMA Yahia

Le : Samedi 06 juillet 2019

Localisation Indoor basé sur la détection Deep Learning

Jury :

Mr.	SBAA Salim	Pr	Université de Biskra	Président
Mme.	BELAHCENE Mebarka	Pr	Université de Biskra	Encadreur
Mme.	ATAMENA Noura	MAA	Université de Biskra	Examinateur

Année universitaire : 2018 - 2019

DÉDICACES

Je rends grâce à Dieu de m'avoir donné le courage et la volonté. Ainsi que la conscience d'avoir pu terminer mes études.

Pour mes parents. A celle qui est toujours à coté de mon cœur, à celle qui
m'apprit le vrai Sens de la vie.

A mes frères et sœurs pour leur soutien moral. A toute ma famille grande et petite.

. A touts mes amis et mes collègues

A tous les enseignants, et étudiants de département Gène Electrique

Remerciements

Avant tout, nous remercions ***Dieu*** *de nous avoir donné la force pour réaliser ce présent travail.*

Nous tenons à exprimer notre très profonde gratitude à Madame ***BELAHCENE Mebarka****, qui n'a ménagé aucun effort pour nous prendre en charge pour la réalisation de ce travail. Sa clairvoyance, sa générosité, sa gentillesse, ses connaissances, le temps qui nous a dispensé, et sa grande disponibilité dont il a fait preuve; nous ont énormément facilité notre tâche.*

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, mes très chers parents, mes frères et tous mes ami(e)s que j'aime tant

Afin de n'oublier personne, mes vifs remerciements s'adressent à tous ceux qui m'ont aidée à la réalisation de ce modeste mémoire

Liste de figure

Liste d'abréviation

GPS	Global Positioning System
GSM	Global System for Mobile
GPRS	General Packet Radio Service
UMTS	Universal Mobile Telecommunications System
LTE	Long Term Evolution
DL	Deep Learning
LOS	Line Of Sight
NLOS	Non Line Of Sight
RFID	Radio Fréquence Identification
AP	Access Point
SNR	Signa-to-Noise Ratio
TDOA	Time Difference Of Arrival
RTS	Request To Send
CTS	Clear To Send
RTT	Round Trip Time
CNN	Convolution Neural Network
RCNN	Region Convolution Neural Network
UWB	ultra-wide Band
RL	Reinforcement Learning
IoU	Intersection over Union
DQL	Deep Q-Learning
MDP	Markov Decision Processes
YOLO	You Only Look Once
FC	Fully Connected
NASNet	Neural Architecture Search Network
RNN	Recurrent Neural Network

R-CNN	Region-based Convolutional Network
R-FCN	Region-based Fully Convolutional Network
RoI	Region of Interest
RPN	Region Proposal Network
ResNet	Residual Network
RMSE	Root-Mean-Square Error

Sommaire

Résumé

La localisation par satellites et la démocratisation du GPS ont permis à la navigation de se développer et de devenir aujourd'hui un élément incontournable du quotidien. Les smartphones, qui accompagnent constamment leur détenteur, sont un élément clé de cette réussite. Par ailleurs, les technologies de l'information et de la communication (TIC) prennent aujourd'hui une place croissante dans les services proposés aux utilisateurs. Ces technologies, comme l'utilisateur lui-même, ne peuvent plus se contenter de la localisation en extérieur. La localisation à l'intérieur des bâtiments est donc un défi qui pourrait élargir la palette d'applications connectées utilisant cette information. Les solutions existantes proposent d'utiliser les signaux terrestres pour identifier la position de l'utilisateur du smartphone. Elles nécessitent cependant une connaissance du bâtiment dans lequel il se trouve. Cetravail s'intéresse au problème de localisation indoor, dans le but de fournir une solution de navigation équivalente à celle du GPS en extérieur. En particulier, nous étudions la faisabilité d'utiliser les algorithmes de traitement d'image pour localiser et détecter des individus. Nous proposons pour cela une application fondée exclusivement sur la localisation basée sur les algorithmes de vision et la détection en profondeur basée sur les régions Faster RCNN. . Pour cela, nous contournons l'imprécision des caméras et smartphones, en étudiant la locomotion humaine et son effet sur les signaux des capteurs. Une campagne de tests nous a permis de montrer que nous obtenons une précision de l'ordre du mètre dans la plupart deux bâtiments différents.

ملخص: أتاح تحديد المواقع عبر الأقمار الصناعية وإضفاء الطابع الديمقراطي على نظام تحديد المواقع العالمي (GPS) إمكانية تطوير الملاحة وأصبحت اليوم عنصرًا أساسيًا في الحياة اليومية الهواتف الذكية ، التي تصاحب أصحابها باستمرار ، هي عنصر رئيسي في هذا النجاح. بالإضافة إلى ذلك، أصبحت تكنولوجيا المعلومات والاتصالات (TIC) مهمة بشكل متزايد في الخدمات المقدمة للمستخدمين. هذه التقنيات، مثل المستخدم نفسه، لم تعد راضية عن الموقع الخارجي. و بالتالي فإن الموقع داخل المباني يمثل تحديًا يمكن أن يوسع نطاق التطبيقات المتصلة باستخدام هذه المعلومات. تقترح الحلول الحالية استخدام الإشارات الأرضية لتحديد موقع مستخدم الهاتف الذكي. أنها تتطلب مع ذلك ، معرفة المبنى الذي يقع فيه. يركز هذا العمل على مشكلة الموقع الداخلي، من أجل توفير حل ملاحة يعادل حل GPS في الهواء الطلق. على وجه الخصوص، نحن ندرس امكانية استخدام خوارزميات معالجة الصور لتحديد مكان الأفراد واكتشافهم. نقترح تطبيقًا يعتمد بشكل حصري على التعريب استنادًا إلى خوارزميات الرؤية والكشف العميق استنادًا إلى مناطق Faster RCNN. . لذلك نتفادى عدم دقة الكاميرات والهواتف الذكية، من خلال دراسة الحركة البشرية وتأثيرها على إشارات المستشعرات. سمحت لنا حملة اختبار أن نظهر أننا نحقق دقة تبلغ غالبا حوالي متر واحد في مبنيين مختلفين.

Mots clés : Localisation ; Calibrage ; RCNN ; Détection.

Introduction générale

Contexte

Se localiser, localiser ses pairs ou trouver son chemin sont des problèmes qui ont toujours existé. A chaque époque, les solutions de localisation ont utilisé les derniers ` moyens technologiques existants : la connaissance de la position des étoiles, les phares (maritimes depuis l'antiquité jusqu'à aujourd'hui, et aéronautiques depuis les années 1920), les satellites. Si, longtemps, la navigation s'est limitée à un positionnement par rapport à des points de référence connus (les villes, les étoiles, les phares), elle a ensuite cherché à atteindre une meilleure précision. Deux techniques se sont démarquées : le positionnement par triangulation, où l'utilisateur mesure les angles par rapport aux points de référence pour en déduire sa position (notamment grâce à l'invention du sextant(instrument de navigation à réflexion servant à mesurer la distance angulaire entre deux points aussi bien verticalement que horizontalement), au XVIIe siècle) ; et la navigation à l'estime, où l'utilisateur connaît sa position initiale, puis il suit ses mouvements grâce à des instruments de mesure comme les compas - méthode utilisée depuis le XVe siècle, et popularisée par Christophe Colomb lors de sa découverte des Amériques. Ce n'est qu'à la fin du XXe siècle que les systèmes de trilatération (d'abord terrestres puis par satellites avec le GPS) ont pris le dessus, en raison de leur précision et de leur couverture. A l'ère de l'information, les technologies de l'information et de la communication (TIC) prennent une place croissante dans notre vie quotidienne.

Problématique

De nouvelles applications nécessitant la localisation voient le jour avec cette nouvelle ère informatique, tels que le stationnement intelligent, la gestion de files d'attente, la surveillance de l'environnement, la gestion de la mobilité, les réseaux sociaux ou encore les jeux géolocalisés, récupérant l'ensemble des informations captées par le smartphone de l'utilisateur. Cela a particulièrement favorisé la démocratisation du GPS et sa présence dans tous les smartphones. Cependant, alors que les systèmes satellitaires répondent à une grande partie des besoins, l'intérieur des bâtiments reste une zone blanche pour la localisation des équipements mobiles, car le signal GPS n'y est pas

disponible (perte de signal). Ce manque est particulièrement important pour les smartphones, éléments déterminants dans la croissance des services proposés aux utilisateurs. Il est le seul outil technologique que la grande majorité des utilisateurs transportent sur eux quasiment en permanence. Muni de plus en plus de capteurs, de mémoire et de capacités de calcul, il permet d'offrir des services aux utilisateurs, grâce aux informations géolocalisées remontées par le mobile. La croissance de ces services a forcé la mise en œuvre de solutions de localisation complémentaires, notamment pour l'indoor, grâce aux autres capteurs du smartphone. Pour des raisons d'imprécision et/ou de coût d'installation, ces solutions ont connu des difficultés d'implantation. De nombreuses propositions ont émergé pour tenter de résoudre ces problèmes, mais aucune n'a suffisamment convaincu pour être déployée à grande échelle.

Après des décennies de recherche, il n'existe toujours pas de produits présents pour la localisation intérieure alors que la demande de services basés sur la localisation intérieure augmente rapidement dans les villes intelligentes. Les dernières années ont vu beaucoup de travail sur la localisation intérieure. La plupart d'entre eux essaient de fournir un système de capteurs largement utilisé pour la localisation en intérieur et d'obtenir des performances satisfaisantes comme le GPS dans les environnements extérieurs. Généralement, certaines méthodes incluent le Wi-Fi et les empreintes magnétiques. Les deux sont basés sur l'hypothèse que chaque emplacement a une caractéristique de signal unique.

Motivation

L'imprécision des capteurs de mouvement (centrale inertielle) a long temps retardé l'avènement de solutions de navigation. Pourtant, on ne désire pas toujours connaître la position d'un utilisateur à un instant précis, mais plutôt le trajet qu'il a effectué. De la même façon que le GPS, qui est dans la plupart des cas utilisé comme un instrument de navigation et non de positionnement, une solution de localisation serait particulièrement efficace si elle était en fait une application de navigation. Les solutions existantes sont également limitées par les contraintes qu'elles font peser sur l'utilisateur, les besoins en infrastructure ainsi que les coûts de déploiement. Pour cette raison, l'objectif est de proposer une solution dans le traitement d'image qui réponde aux besoins des applications : limitant l'erreur de localisation.

Aujourd'hui, de nombreux nouveaux développements technologiques ont eu lieu. À la suite de ces progrès technologiques, les gens peuvent être confrontés à plusieurs problèmes. Certaines des négativités de tels problèmes à expérimenter avec la localisation et la détection peuvent être minimisées par diverses approches. Par conséquent, des études ont été effectuées et réalisées sur la détection et jour après jour de nouvelles solutions et les algorithmes sont développés avec de nouvelles études. Deep Learning et les technologies pertinentes sont un autre sujet principalement élaboré au cours des périodes récentes. Si nous définissons des méthodes d'apprentissage en profondeur comme les méthodes qui comprennent les réseaux de neurones artificiels, comprenant au moins une couche cachée. Ils capable de réaliser un processus de détection de caractéristiques automatiquement à partir de grandes quantités de données d'entraînement étiqueté. Les méthodes d'apprentissage en profondeur utilisent des algorithmes connus sous le nom de réseaux de neurones, qui sont inspirés par les méthodes de traitement de l'information de biologique des systèmes nerveux tels que le cerveau et ces méthodes permettent aux ordinateurs d'apprendre ce que chaque donnée représenteet ce que chaque modèle correspondant signifie réellement. Dans cette étude, les approches de détection de personne et d'apprentissage en profondeur sont combinées.

Les méthodes de récupération d'objet comprennent généralement deux étapes, c'est-à-dire la recherche d'images contenant l'objet requête et la localisation de l'objet dans l'image avec un cadre de sélection.

La première étape est essentielle pour la détection des images et de nombreux résultats de haute performance sont obtenus en utilisant le réseau de neurones à convolution (CNN). Ce type de technique permet d'apprendre des fonctionnalités à partir de jeux de données d'image bruts dans un processus orienté tâche. Contrairement aux approches conventionnelles, les descripteurs à usage général de CNN peuvent être utilisés dans différentes tâches. Cependant, le CNN général ne peut être utilisé que dans la détection d'image et ne peut pas réaliser la localisation d'objet. Le réseau de neurones de convolution basé sur une région (R-CNN) peut identifier et localiser de manière synchrone un objet dans une image. Premièrement, R-CNN extrait environ 2 000 propositions de région à l'aide de la recherche sélective (SS) ou de bord Box.

Deuxièmement, il élabore les fonctionnalités de toutes les propositions utilisant un CNN, telles que ConvNet.

Plan du mémoire

Les travaux de ce mémoire sont présentés en cinq chapitres. Le chapitre 1 est une généralité sur la localisation indoor.

Dans le chapitre 2, nous nous intéressons aux approches technologiques de la localisation indoor en établissant un état de l'art.

Le chapitre 3 est consacré à la méthodologie pour la localisation indoor basée sur la Deep détection.

Le chapitre 4 présente la conception de la localisation indoor basée sur la Deep détection.

Et le chapitre 5 est destiné à l'implémentation et aux résultats.

Enfin, une conclusion générale termine notre mémoire et quelques perspectives sont proposées.

Chapitre 1 :

Généralités sur la localisation Indoor

Introduction

Les hommes ont toujours eu besoin de localiser les objets et de se situer dans l'environnement. Pour répondre à cette nécessité, plusieurs techniques ont été utilisées. Au début de l'Humanité, l'homme utilisait les pierres (ou montagnes) pour se repérer. Les particularités du relief lui servaient de repère pour retrouver son chemin à travers la jungle et les déserts. Les précurseurs de la navigation ont laissé des traces sur leur passage comme des marques sur des pierres ou des arbres. Le concept de base à toute localisation est donc la "référence". C'est sur cette notion que reposent tous les systèmes de localisation qui se succèdent pour fournir la position d'un objet ou d'une personne [1].

L'importance de la localisation des personnes à l'intérieur a considérablement augmenté au cours de la dernière décennie. Les gens passant plus de 85% de leur temps dans des environnements intérieurs, une localisation précise peut avoir un impact considérable non seulement en simplifiant la vie des personnes, mais également en aidant les pompiers, la police, les soldats et le personnel médical à sauver des vies et à effectuer des tâches spécifiques.

1.1 Définition et objectif de la localisation

D'après la définition Wikpédia : « la **géolocalisation** est un procédé permettant de positionner un objet, un véhicule, ou une personne sur un plan ou une carte à l'aide de ses coordonnées géographiques. Certains systèmes permettent également de connaître l'altitude (géolocalisation - dans l'espace - en 3D).Cette opération est réalisée à l'aide d'un terminal capable d'être localisé grâce à un système de positionnement par satellites et un récepteur GPS par exemple, ou par d'autres techniques ; de plus, le terminal est en mesure de publier, en temps réel ou de façon différée, ses coordonnées géographiques latitude/longitude. Les positions enregistrées peuvent être stockées au sein du terminal et être extraites ultérieurement, ou être transmises en temps réel vers une plateforme logicielle de géolocalisation. La transmission en temps réel nécessite un terminal équipé d'un moyen de télécommunication de type GSM / GPRS, UMTS, LTE, radio ou satellite lui permettant d'envoyer les positions à des intervalles plus ou moins réguliers. Cela

permet à la plateforme de visualiser la position du terminal au sein d'une carte. La plateforme est le plus souvent accessible depuis Internet ». La **figure 1.1** définit en image la **géolocalisation** [2].

Figure 1. 1: Schéma de principe de la géolocalisation par GPS [2]

D'une manière générale, la localisation d'un objet consiste en la détermination de sa position géographique. Ce mot peut aussi faire référence à l'adaptation d'un objet pour une région particulière. On définit par l'action de localiser, de situer par le fait d'être localisé ou d'être situé dans l'espace ou le temps. C'est aussi la localisation d'un bruit ou la localisation d'un engin spatial par rapport à la terre.

La géolocalisation à l'intérieur présente plusieurs axes de recherche. Plusieurs technologies et techniques sont proposées pour localiser l'utilisateur dans un lieu fermé. Une des solutions proposées est la localisation par signaux WiFi des réseaux locaux déjà déployés dans la plupart des bâtiments. Dans notre projet, l'axe principal est la mise en œuvre d'une architecture géolocalisation indoor associée à une détection basée sur le Deep Learning. Notre projet consiste à concevoir une application qui estime la position d'un client Android au sein d'un lieu fermé équipé de réseau WiFi. Ce travail consiste en la conception d'un système de géolocalisation professionnel complet en ajoutant la cartographie spatiale et la navigation en intérieur. L'objectif du projet est d'étudier dans un premier temps les techniques, les technologies et les algorithmes qui peuvent être utilisées dans la conception de notre projet, ensuite de proposer une solution à base d'algorithmes choisis et des techniques appropriées. La mise au point d'un système de positionnement, facile à maintenir et performant avec une précision améliorée, et qui pourra faire partie d'autres systèmes de géolocalisation dans les futurs travaux, est notre objectif principal en tenant compte du délai imparti et des moyens matériels et logiciels présents.

1.2 Les moyens de localisation actuels

1.2.1 Les systèmes satellitaires

Il existe actuellement trois services mondiaux de positionnement par satellite:

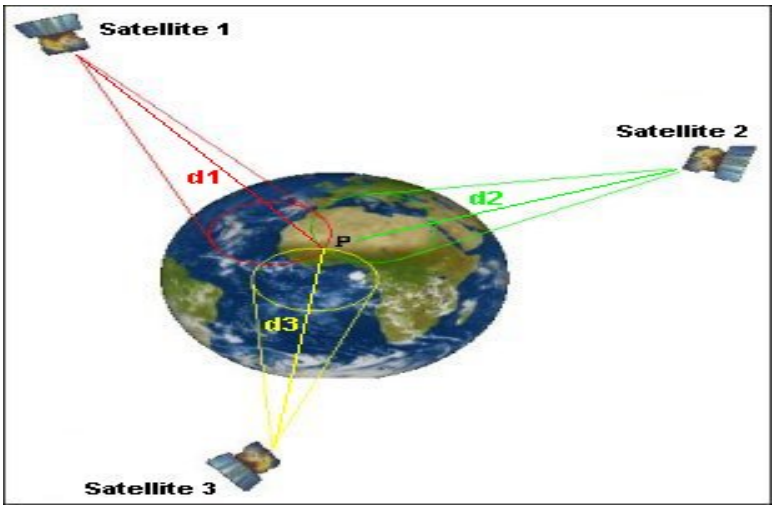


Figure 1. 2: Systèmes satellitaires pour localisation [3]

- Le GPS (Global Positioning System) ou système de positionnement par satellite, dispositif américain mis en service depuis 1978 et graduellement amélioré. Jusqu'en 2007, seul GNSS était opérationnel. Le système de référence associé au GPS est le WGS-84 (World Geodetic System).
- GLONASS, dispositif militaire russe mis en service en 1982.
- GALILEO, dispositif civil européen mis en service en 2011, et devant être 100% opérationnel en 2019 [4].

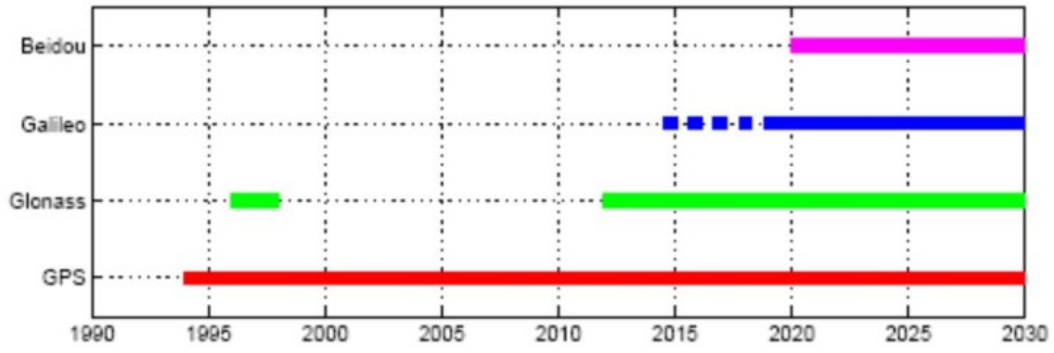


Figure 1. 3: Phase d'opérationabilité des différents GNSS existants ou en projet [4]

1.2.2 Les systèmes de localisation par réseaux terrestres

Aujourd'hui, de nombreux réseaux cellulaires ou sans fil existent. Ces réseaux communiquent avec les équipements mobiles par radio.

Le premier système de ce type est le système LORAN C dans les années 1960.

Ensuite les réseaux comme le GSM et l'UMTS ou le réseau de TNT (Télédiffusion Numérique Terrestre) sont apparus pour relayer le système LORAN C.

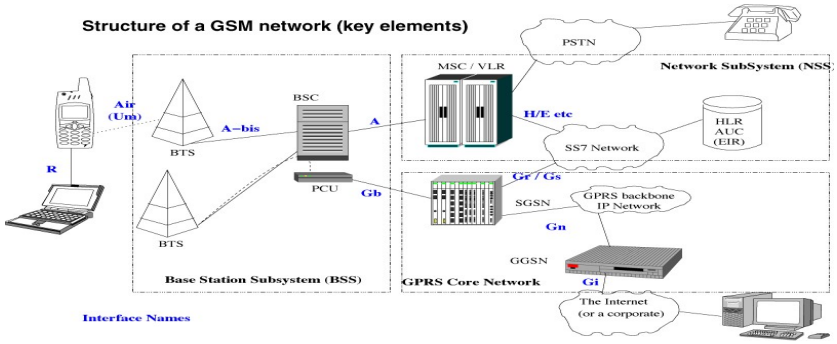


Figure 1. 4: Schéma d'un réseau GSM [5]

L'exploitation de ces différents réseaux terrestres est une nouvelle source potentielle de revenus pour les propriétaires de ces réseaux (opérateurs de télécommunication par exemple). Un désavantage de ces systèmes de localisation est que la portée de chacune de ces stations terrestres est limitée. Pour les réseaux de télécommunication, le réseau est construit de manière à ce que les cellules se juxtaposent les unes à côté des autres en évitant autant que possible les recouvrements entre cellules (optimisation des réseaux cellulaires) afin de couvrir le plus grand territoire avec le minimum de stations terrestres. Comme pour le système GPS, un certain nombre d'informations provenant de plusieurs stations terrestres est nécessaire pour localiser un équipement mobile. La portée de chacune des stations de base est limitée à une zone bien précise, plus ou moins grande suivant l'environnement, et la disponibilité de plusieurs stations de base pour une position n'est pas garantie surtout dans des situations de visibilité directe (très grand éloignement entre les stations de base). Des systèmes de localisation exploitent ces réseaux terrestres et sont abordés par la suite avec les techniques de localisation qui leurs sont associées [1].

1.2.3 La navigation par mesures inertielles

Ce système de navigation utilise un ou plusieurs capteurs délivrant des informations quant au comportement de l'utilisateur. Ces capteurs sont embarqués sur le mobile lui-même. L'exploitation des équations de la mécanique, comme l'équation du mouvement, permet de déterminer la position du mobile à partir des informations délivrées par les différents capteurs. Généralement, les capteurs utilisés sont des accéléromètres mesurant l'accélération, des gyroscopes mesurant des vitesses angulaires, des compas mesurant une direction par rapport au nord magnétique, des sondes barométriques,...ce type de navigation est souvent utilisé dans des applications

militaires [6]. Les données issues des capteurs sont disponibles en permanence (pas de problème de couverture radio comme pour les technologies précédentes). Les traitements des données provenant de ces capteurs se font localement, c'est à dire sur l'objet mobile, ce qui garantit un élément de sécurité. L'autonomie de ce système de navigation est très importante. Cette navigation est utilisée pour le guidage de missiles balistiques par exemple. Des applications civiles utilisent ce type d'éléments pour obtenir une localisation. Des domaines comme l'aéronautique civile ou l'automobile utilisent ces capteurs afin d'affiner la localisation GPS, ou alors en substitution de la navigation par GPS si celle-ci est indisponible.

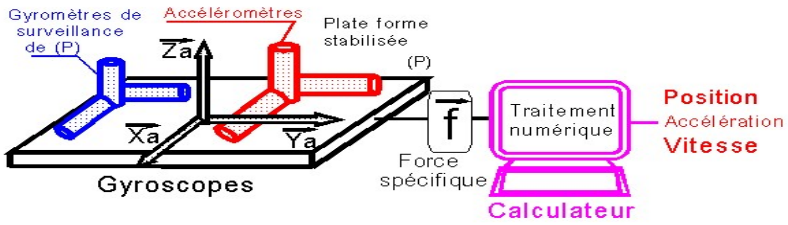


Figure 1. 5: Digramme fonctionnel du navigateur inertiel [7]

1.3 Localisation indoor

La localisation indoor 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.

Figure 1. 6: Géolocalisation indoor [7]

Les systèmes de géolocalisation indoor permettent de connaitre 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.

1.4 Les moyens de localisation indoor

Afin d'obtenir des précisions de l'ordre du mètre ou voire meilleures, plusieurs technologies de proximité ont été explorées. La proximité de tous les éléments du système (mobile, stations de bases du réseau) permettent d'atteindre une précision métrique. L'émergence des nouveaux réseaux sans fil est une des solutions à envisager pour se localiser à l'intérieur des bâtiments. D'autres technologies comme celles par tags actifs/passifs ou de vision sont autant de moyens pour se localiser. Cependant, ces technologies présentent des points faibles qui peuvent devenir des freins quant à leur déploiement, notamment le coût et la complexité d'installation (synchronisation des éléments entre eux, conditions particulières d'installation= (angle de vue)). Le prix des éléments d'un système gêne parfois le déploiement d'une technologie, et particulièrement lorsque de très nombreux éléments relativement onéreux doivent être installés.

1.4.1 La localisation par ultrason

Les systèmes à ultrason sont utilisés pour déterminer la position d'un mobile. La plupart des systèmes de localisation par ultrason sont combinés avec une autre

technologie afin d'obtenir une estimation de la distance émetteur/récepteur. Dans le système Cricket [8], les informations provenant d'une interface ultrason sont combinées avec celles provenant d'une interface RF. Cette combinaison permet d'estimer la distance émetteur/récepteur puis la position occupée par le mobile. Des émetteurs sont placés au plafond du bâtiment et émettent des signaux RF contenant des informations de localisation. En même temps que ces signaux RF sont émis, une onde ultrasonore est émise à partir de ce même émetteur. Le récepteur reçoit successivement l'onde RF et l'onde ultrasonore. Il effectue une corrélation de ces deux signaux reçus pour extraire la différence des temps d'arrivée entre chacune de ces ondes. Ceci permet d'estimer la distance le séparant de l'émetteur qui a émis ces deux signaux. En réitérant cette même mesure avec plusieurs émetteurs, on détermine précisément la position du mobile (ici le récepteur) dans l'environnement. Ceci est basé sur le fait que l'onde sonore, et l'onde radio possèdent des vitesses de propagation différentes.

Le récepteur mesure la différence de temps existante entre les instants d'arrivée de ces deux ondes au niveau du récepteur.

Cette technique mesure la distance séparant un émetteur d'un récepteur. Un certain nombre de traitements est nécessaire car les objets se trouvent dans des environnements générant beaucoup de multi-trajets et la situation de visibilité directe (LOS) n'est pas toujours garantie. Dans une situation de non visibilité directe (NLOS), le trajet le plus fort au niveau de la réponse impulsionnelle n'est pas le trajet le plus court. Il est nécessaire de mettre en place des algorithmes pour estimer cet instant d'arrivée du premier trajet afin de minimiser les erreurs sur l'estimation de la distance émetteur/récepteur. Un mécanisme d'accession au medium est en place dans le système Cricket pour minimiser les risques d'interférence. Ces interférences rendent la détection de cet instant d'arrivée pour chacun des émetteurs encore plus difficile [9].

Le système de localisation Active Bat repose sur les mêmes principes que ceux présentés ci-dessus. La principale différence entre ces deux systèmes est que l'un est basé sur une architecture centralisée (Active Bat), tandis que pour le système Cricket, les traitements sont effectués par l'équipement mobile. Dans le système Active Bat, dès qu'un émetteur est détecté, le contrôleur général (Master qui se trouve sur le réseau) envoie un signal RF à l'équipement mobile (Bat). Le Bat transmet alors une série d'impulsions ultrasonores. Toutes les 200 ms, un message radio contenant

uniquement les 16 bits d'adressage d'un des tags est émis par un contrôleur relié à un PC. Le PC décide de l'adresse qui doit émettre. Les équipements mobile reçoivent ce message et le décodent. Le mobile qui reconnaît son adresse dans le message reçu passe dans un mode d'émission et émet un message par ultrason durant 50 µs. Une fois la série de pulses émise, le tag se remet en mode économie d'énergie, et scrute le canal 195 µs plus tard. Le PC servant de centre nerveux au système de récepteurs situés au plafond émet vers chacun des récepteurs via la liaison série un signal de reset lorsqu'il émet en même temps que le signal RF à destination des tags à localiser. Durant 20 ms à partir de ce moment, le dispositif électronique associé à chacun des récepteurs cherche à détecter un signal ultrason. De multiples détections se produisent à cause des multi-trajets. Ensuite, le PC central interroge chacun des récepteurs composant le réseau en récupérant les intervalles de temps entre le signal de reset et celui de la détection du pic de signal ultrason.

1.4.2 La localisation par infrarouge

Le système Active Bat est l'un des premiers systèmes de localisation. AT&T l'a élaboré entre 1989 et 1992. Ce système exploite la technologie infrarouge. Le mobile à localiser est équipé d'un tag infrarouge émettant un signal infrarouge toutes les 10 secondes. Les récepteurs sont installés au plafond dans chaque pièce de l'environnement. Ces récepteurs sont reliés entre eux pour former un réseau permettant de détecter le tag actif. Comme dans le système Active Bat/Cricket, le système infrarouge émet une série de pulses. Cette technologie a été retenue à cette période car elle est peu coûteuse. De plus, la portée des capteurs utilisés est de 6 m. Pour des utilisations dans de petites pièces, de nombreuses réflexions sont présentes et facilitent la détection. Le désavantage par rapport aux technologies radio, c'est que les signaux ne traversent pas les murs, ce qui réduit la portée du système.

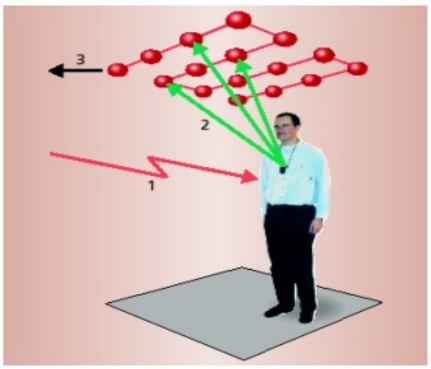


Figure 1. 7: Principe de fonctionnement du système Active Bat [1]

Les émissions infrarouge se font toutes les 15 s (durant 0.1 s) afin d'économiser l'énergie, mais aussi pour permettre à plusieurs tags d'être localisés et éviter ainsi les problèmes d'interférence. Une période de répétition de 15 s est rédhibitoire, car en 15 s une personne peut effectuer un déplacement important. Pour des environnements indoor, ceci n'est pas totalement justifié, et lors de l'exploitation du système, une bonne précision a été obtenue. La présence de la lumière du jour est un frein au développement de cette technologie car cette lumière perturbe la transmission infrarouge entre l'émetteur et le récepteur. Ainsi cette technique de localisation est orientée vers une détection de présence du mobile dans l'environnement (ou une de ses parties). On parle de localisation par zone. On retrouvera ce même type d'information binaire lors de l'exploitation des données remontant de capteurs RFID (Radio Frequency Identifier) [10].

1.4.3 La localisation par vidéo

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 [11]. 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. 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, différentier 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 conclue 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/séparation [1].

1.4.4 La localisation par onde radio (Wi-Fi, Bluetooth, RFID)

À l'inverse du GPS, les réseaux de communication à courte portée permettent de se localiser à l'intérieur des bâtiments. Ces réseaux sont déployés dans les bâtiments. Des précisions de l'ordre du mètre sont atteignables grâce aux réseaux locaux. On distingue plusieurs catégories d'interfaces de localisation. On trouve des réseaux d'étiquettes actives ou passives, dites RFID qui permettent de détecter si un objet se trouve dans un certain périmètre autour du lecteur d'étiquettes. Le projet SpotOn est basé sur ces étiquettes RFID. Il a été réalisé à l'Université de Washington [12]. Ce type de localisation donne bien souvent une bonne précision, car elle est de l'ordre de quelques centimètres, voire dizaines de centimètres. Cependant la portée de ces étiquettes et de ces lecteurs d'étiquettes n'est pas très importante (de l'ordre de quelques mètres). Le procédé repose sur une détection de la présence ou non de l'élément étiquette dans le giron de la borne. Une connaissance de la position occupée par la borne permet de remonter à la position de l'étiquette. Cette information binaire (présence / non présence), n'est pas toujours facilement exploitable notamment lorsque les zones de couvertures des différents lecteurs sont disjointes. Pour obtenir une couverture quasi continue du service de localisation, il faut équiper l'environnement d'une grande quantité de capteurs. Malgré le faible coût de ces capteurs (RFID passif ou RFID actif) cela devient rapidement contraignant et coûteux et se transforme en un frein au déploiement de cette technologie.

Dans certains environnements comme dans des entrepôts, de nombreuses pièces entreposées peuvent disposer de leur propre étiquette, et une localisation permanente de ces pièces n'est pas nécessaire. Une simple détection de la sortie de l'équipement est nécessaire. Cette technologie est adaptée à ce type d'application où il faut connaître l'état de transition d'une zone à une autre par l'information binaire de présence ou non. La faible portée du système et la contrainte de passer l'étiquette sur un lecteur font qu'un autre système radio à plus grande portée doit être utilisé afin de rendre une certaine liberté à l'utilisateur et que le service de localisation soit disponible sur l'ensemble du bâtiment avec un minimum d'infrastructure. Les réseaux locaux sans fil de type 802.11 sont une bonne alternative. Le réseau est composé d'un

ensemble de points d'accès (AP) servant de relais de communication entre les objets mobiles et le réseau. Aujourd'hui ces réseaux 802.11a/b/g sont présents dans de plus en plus de lieux publics, ainsi que chez les particuliers. Le premier but de ces réseaux est d'effectuer des communications d'informations. L'idée est de détourner l'usage de ces réseaux à des fins de localisation. Un certain nombre de signaux de contrôles sont émis pour gérer le roaming lors du déplacement de l'équipement mobile. Ces signaux lui permettent de rester connecté avec le point d'accès avec lequel le rapport signal à bruit (RSB) est le meilleur. RADAR de Microsoft est basé sur ce principe. Une technique de localisation associée à ces réseaux est l'empreinte par puissance de signal reçu. Les algorithmes de recherche du plus proche voisin (Pattern Matching) et de recherche probabiliste sont présentés dans la section. La portée des points d'accès peut aller jusqu'à 300 m pour des environnements outdoor, mais en indoor, cette portée est généralement comprise entre 50 à 70 mètres, ce qui est suffisant pour couvrir bon nombre de bâtiments. D'un autre côté leur prix étant aujourd'hui relativement bas, ce type de technologie devient accessible et la couverture radio intégrale d'un grand bâtiment et d'une ville est peu coûteuse.

La technologie Bluetooth est aussi disponible dans certains environnements indoor. Cette technologie est similaire à la précédente du fait qu'elle se base sur un réseau composé de bornes servant de points d'accès entre un réseau sans fil et un réseau filaire. À l'intérieur des bâtiments, la portée de ces bornes est plus restreinte que celle des points d'accès 802.11, car elle n'est généralement que d'une dizaine de mètres. AeroScout [13] (Cisco) a exploré cette technologie avant de passer à un système actuel employant des RFID et la technique de localisation TDOA avec un réseau Wi-Fi. L'exploitation de ce système n'est pas simple, car cela nécessite l'achat de points d'accès propriétaires. Les informations temporelles utiles pour le TDOA sont indisponibles dans les produits commerciaux actuels. Les performances annoncées pour ce système sont de l'ordre du mètre, tout comme pour le système de localisation par mesure de puissance. Hitachi étudie aussi un système de localisation similaire. Un moyen d'obtenir ces informations temporelles est d'effectuer des mesures de temps entre l'instant de départ du dernier bit du paquet RTS et l'instant d'arrivée du premier bit du paquet CTS. En tenant compte des latences de chacun des équipements, on remonte au temps mis par un paquet RTS ou CTS pour effectuer la liaison émetteur/récepteur. Cette méthode est appelée la méthode d'aller/retour ou Round

Time Trip (RTT). Des travaux sont menés sur ce sujet, notamment à l'Université de Catalogne en Espagne et à l'université de Berlin [14].

1.5 Localisation indoor par le traitement d'image

Pour permet à un système 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 système rentre dans cette zone. Et déterminer la distance par rapport un repère.

Dans un contexte de navigation visuelle, il est indispensable d'avoir une vue d'ensemble sur l'environnement qui nous entoure pour pouvoir y évoluer et circuler. Toute information est bonne à prendre : la structure de l'environnement, les repères à suivre, les obstacles à éviter, etc. Ces informations vont servir à se localiser dans l'environnement et accomplir les tâches souhaitées.

A fin de se localiser, il est nécessaire de se baser sur des repères visuels sémantiques stables. La détection de ces repères est une problématique qui a suscité beaucoup d'intérêt, et a fait l'objet de nombreux travaux à la fois dans le domaine de la navigation et la reconstruction de l'environnement [15].

Différents types de repère peuvent être utilisés. De façon non exhaustive on peut citer:

- **Point de fuite:** un repère des plus basiques permettant de déterminer l'orientation de la perspective.
- **Points saillants:** de type Haar, SIFT etc. ces points possèdent des caractéristiques bien spécifiques permettant de les identifier et de les suivre lors de la navigation.
- **Amers visuels:** ces repères sont créés spécialement pour la localisation (typiquement un cône posé par terre). Leurs caractéristiques sont bien modélisées et leurs positions permettent une localisation immédiate.
- **Objets de l'environnement:** ce dernier type de repère peut être représenté par n'importe quel objet faisant partie intégrante de l'environnement (porte, fenêtre, escalier, etc.).

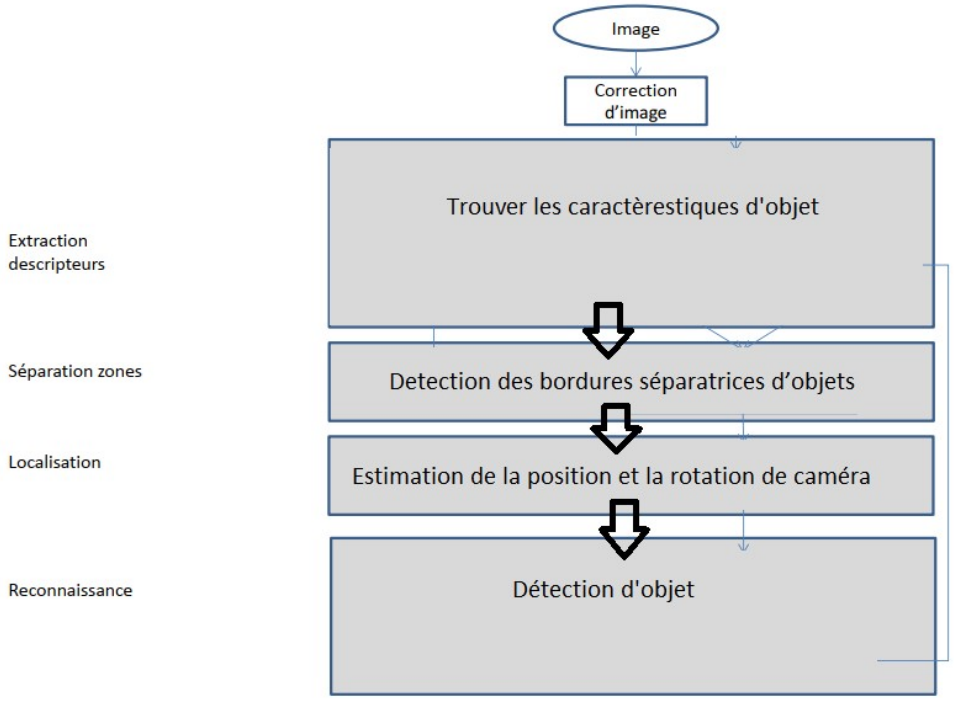


Figure 1. 8: Schéma principale de la détection et la localisation [15]

1.6 Localisation indoor par le traitement d'image par Deep Learning

Le terme Deep Learning (en français : apprentissage profond) est très en vogue ces derniers temps. C'est bien simple, lorsque l'on parle d'intelligence artificielle, on parle presque systématiquement de Deep Learning. A tel point que dans l'esprit de beaucoup, ces deux termes sont synonymes. C'est pourtant inexact.

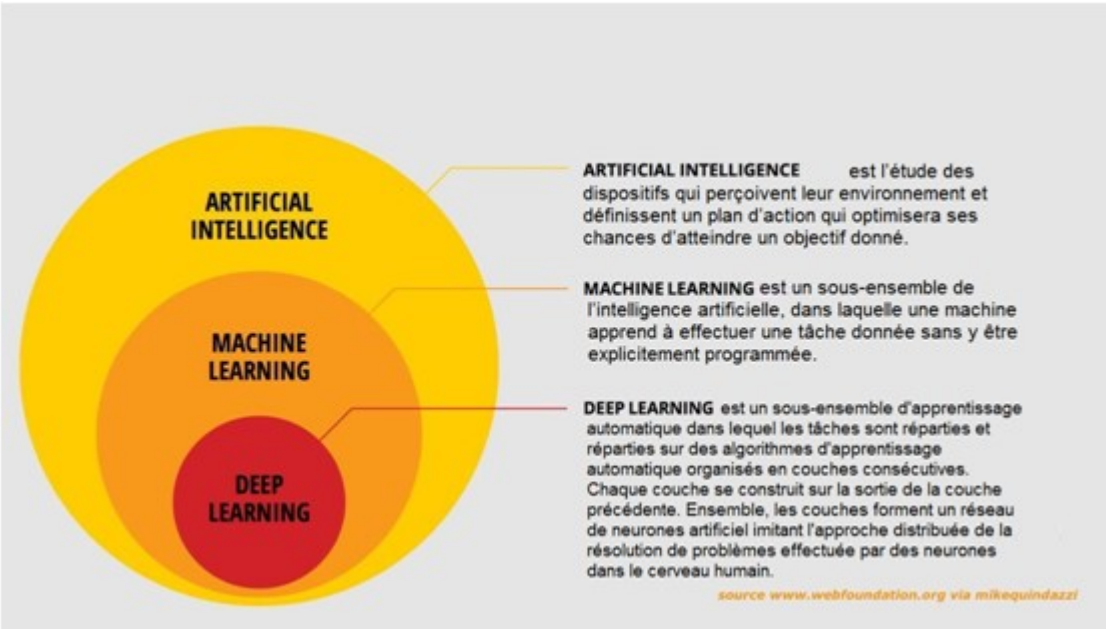


Figure 1. 9: Artificial intelligence(AI), Machine learning et Deep learning [16]

1.7 Localisation et détection par Deep Learning

La localisation et la détection d'objets sont deux des tâches principales de Computer Vision, car elles sont appliquées dans de nombreuses applications du monde réel telles que les véhicules autonomes et la robotique. Donc, si vous voulez travailler dans ces industries en tant que spécialiste en vision par ordinateur ou si vous voulez créer un produit relatif, vous feriez mieux de le maîtriser [17].

Premières choses d'abord. Faisons un bref récapitulatif des termes les plus utilisés et de leur signification pour éviter les idées fausses:

- Classification / Reconnaissance: Étant donné une image avec un objet, découvrez ce qu'est cet objet. En d'autres termes, classifiez-le dans une classe à partir d'un ensemble de catégories prédéfinies.
- Localisation: Trouvez où se trouve l'objet et tracez un cadre de sélection autour de celui-ci.
- Détection d'objet: classifiez et détectez tous les objets de l'image. Attribuez une classe à chaque objet et tracez un cadre de sélection autour de celui-ci.
- Segmentation sémantique: classifiez chaque pixel de l'image dans une classe en fonction de son contexte, de sorte que chaque pixel soit affecté à un objet
- Segmentation d'instance: classifiez chaque pixel de l'image dans une classe de sorte que chaque pixel soit affecté à une instance différente d'un objet

Rappelez-vous cependant que ces termes ne sont pas clairement définis dans la communauté scientifique, vous pouvez donc en rencontrer un dans un sens différent. À mon sens, ce sont les interprétations correctes.

Au moment de bien comprendre les termes de base, il est temps de faire de la localisation et de détecter des objets. Comment faisons-nous ça? De nombreuses approches ont été utilisées au fil des ans, mais depuis l'arrivée de Deep Learning, les réseaux de neurones convolutionnels sont devenus la norme de l'industrie. N'oubliez pas que notre objectif est de classer l'objet et de le localiser. Mais sommes-nous sûrs qu'il n'y a qu'un seul objet? Est-il possible qu'il y ait deux, trois ou quinze objets? En fait, la plupart du temps, c'est le cas.

C'est pourquoi nous pouvons diviser notre problème en deux problèmes différents. Dans le premier cas, nous connaissons le nombre d'objets (nous parlerons du

problème comme classification + localisation) et dans le second cas, nous ne le ferons pas (détection d'objet). Je vais commencer par le premier car c'est le plus simple.

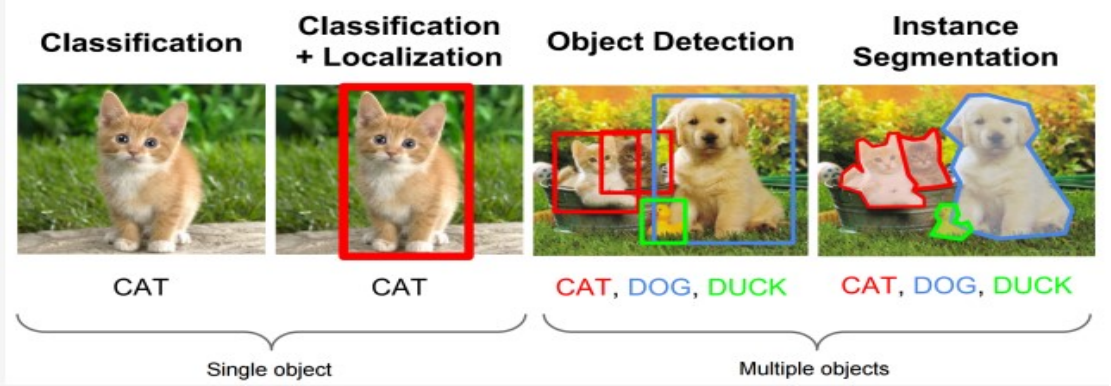


Figure 1. 10: Etapes de DL pour localiser et détecter [17]

1.7.1 Classification et localisation

Si nous n'avons qu'un seul objet ou si nous connaissons le nombre d'objets, c'est en réalité trivial. Nous pouvons utiliser un réseau de neurones de convolution et l'entraîner non seulement pour classer l'image, mais également pour produire 4 coordonnées pour le cadre de sélection. De cette manière, nous traitons la localisation comme un simple problème de régression.

Par exemple, nous pouvons emprunter un modèle bien étudié, tel que ResNet ou Alexnet, qui consiste en un groupe de couches de convolution, de regroupement et autres, et de réutiliser la couche entièrement connectée pour produire le cadre de sélection en dehors de la catégorie. C'est si simple que nous nous demandons si cela donnera des résultats. Et cela fonctionne plutôt bien dans la pratique. Bien sûr, vous pouvez avoir l'imagination et modifier l'architecture pour traiter des problèmes spécifiques ou améliorer sa précision, mais l'idée principale demeure.

Assurez-vous de noter que pour utiliser ce modèle, nous devrions avoir un ensemble de formation avec des images annotées pour la classe et le cadre de sélection. Et ce n'est pas très amusant de faire de telles annotations.

Mais que se passe-t-il si nous ne connaissons pas le nombre d'objets a priori? Ensuite, nous devons entrer dans le trou du lapin et parler de choses hardcore. Es-tu prêt? Voulez-vous prendre une pause avant? Bien sûr, je comprends mais je vous préviens de ne pas partir. C'est là que commence le plaisir.

1.7.2 Détection d'objet

On rigole. Il n'y a rien de hardcore dans les architectures qui seront discutées. Il n'ya que quelques idées intelligentes pour rendre le système intolérant au nombre de sorties et pour réduire ses coûts de calcul. Nous ne connaissons donc pas le nombre exact d'objets dans notre image et nous souhaitons tous les classer et dessiner un bœuf qui les entoure. Cela signifie que le nombre de coordonnées que le modèle doit générer n'est pas constant. Si l'image a 2 objets, nous avons besoin de 8 coordonnées. Si on a 4 objets, on veut 16.

Une idée clé de la vision par ordinateur traditionnelle est la proposition des régions. Nous générons un ensemble de fenêtres susceptibles de contenir un objet à l'aide d'algorithmes CV classiques, telles que la détection de bord et de forme, et nous n'appliquons que ces fenêtres (ou régions d'intérêt) au CNN. Pour en savoir plus sur la façon dont les régions sont proposées.

C'est la base d'un document fondamental qui a introduit une nouvelle architecture appelée RCNN.

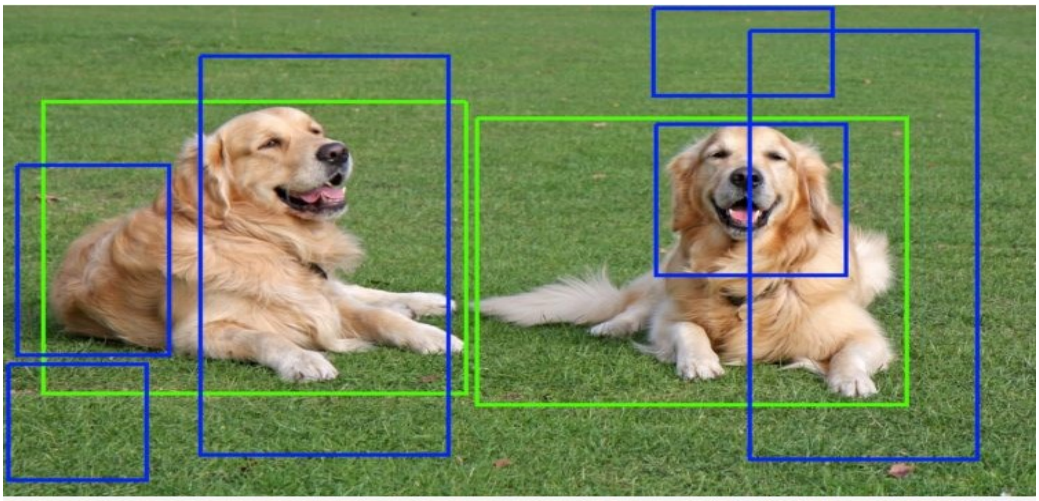


Figure 1. 11: Détection d'objet avec DL [17]

Conclusion

Pour permet à un système 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 système rentre dans cette zone. Et déterminer la distance par rapport un repère.

Chapitre 02 :

La localisation Indoor : Etat de l'Art technologique

Introduction

La localisation indoor (littéralement « en intérieur ») connaît un succès sans précédent. Là où les signaux GPS peinent à se propager, de nouvelles technologies viennent prendre le relais. Atteignant parfois des précisions centimétriques en environnement complexe, ces systèmes de positionnement révèlent un potentiel d'utilisation certain.

2.1 La localisation indoor

Que ce soit pour déterminer l'emplacement d'un service et/ou d'un équipement ou bien se repérer soi-même dans l'environnement, les besoins en termes de géolocalisation sont de plus en plus marqués. Toujours plus spécifiques, les exigences et les technologies associées doivent évoluer régulièrement afin de répondre à de nouvelles contraintes de précision et de disponibilité.

Si la technologie GPS s'impose aujourd'hui en outdoor, offrant à grande échelle une précision de l'ordre de quelques mètres, le standard peine à fournir des résultats concluants en environnement fermé. En cause, des signaux satellites bridés en précision ou une fiabilité jugée trop faible au regard des utilisations souhaitées.

Ainsi, avec le déploiement omniprésent des réseaux sans fil, de nouveaux systèmes de positionnement indoor voient le jour. Les services fournis couvrent aussi bien des besoins de navigation que de suivi ou encore de surveillance. Et si les acteurs du marché tendent à se multiplier, il en est de même pour les applications.

Certains systèmes opèrent en milieu hospitalier, musées ou parkings souterrains dans une optique de guidage ou suivi temps réel. De même en logistique, l'optimisation des déplacements constitue un avantage stratégique indéniable afin de réduire le coût d'acheminement des marchandises. Dans les commerces, cette localisation fournit des analyses comportementales sur le parcours des clients. Identifier la position d'un utilisateur permet de créer de l'interactivité en fonction du lieu où il se trouve.

Les systèmes s'adaptent également à des contextes de surveillance. Les technologies peuvent être mises à profit afin de suivre des personnes sensibles telles que des enfants en bas âge ou des personnes âgées (par exemple).

Les systèmes de positionnement indoor permettent de localiser toute entité munie d'un tag ou encore de repérer des objets les uns par rapport aux autres.

Bien que certains de ces enjeux soient similaires à ceux de la localisation outdoor, les technologies utilisées n'en demeurent pas moins radicalement différentes.

Au sein d'espaces cloisonnés, de nouvelles contraintes apparaissent. La propagation des signaux en ligne de vue directe n'est plus garantie, les technologies doivent également s'accommoder des potentiels phénomènes de multi-trajets.

La localisation indoor soulève donc de nouveaux défis de conception et d'installation.

Les systèmes de positionnement, désormais déployés en environnement dynamique, doivent allier robustesse et fiabilité tout en répondant à des exigences de précision, de temps critique, d'efficacité énergétique et de coûts de déploiement.

2.2 Quelles technologies mettre en œuvre?

Les technologies à l'œuvre au sein de tels systèmes peuvent se répartir en deux catégories :

- D'une part, celles fonctionnant de manière autonome sans interaction quelconque entre l'usager et les équipements externes.
- D'autre part celles nécessitant le déploiement d'une infrastructure en amont de la localisation (capteurs ou antennes, répartis dans la pièce à des positions connues).

2.2.1Les technologies autonomes (sans infrastructure)

2.2.1.1 Capteurs inertiels et dead reckoning

Ces premiers systèmes de positionnement s'appuient sur une méthode dite de « dead reckoning », celle-ci consistant à déduire la position actuelle à partir de la dernière position connue. L'utilisateur a ainsi recours à des capteurs inertiels (accéléromètres, gyroscopes, etc.) parfois utilisés conjointement avec une boussole ou un podomètre. Ces différents dispositifs permettent de quantifier les déplacements dans l'espace. En pratique l'utilisateur dispose d'un Smartphone ou d'une carte embarquée intégrant la technologie inertielle. Le plan du bâtiment étant disponible, il est nécessaire de renseigner une position d'origine dans l'application de positionnement (connaissance de points de repères visuels).

Les données d'accélération récoltées pendant la phase de mouvement permettent ensuite de déterminer de nouvelles positions. Le positionnement est relatif, la position à l'instant T-1 permettant de déterminer celle à l'instant T.

La carte ou le téléphone support peuvent par ailleurs communiquer avec un serveur. L'opération permet ainsi de centraliser les données tout en réduisant les coûts calculatoires au niveau du terminal embarqué.

Atouts et faiblesses : De tels systèmes présentent l'avantage d'être totalement autonomes une fois le plan du bâtiment acquis. Toutefois la précision atteinte reste limitée de par les erreurs progressivement cumulées.

En pratique les capteurs inertiels sont usuellement utilisés en support d'autres technologies, selon un principe de fusion des données.

2.2.1.2Étude des champs magnétiques

Que ce soit au niveau d'un couloir, d'une porte ou encore d'une cage d'ascenseur, les structures métalliques présentes dans un bâtiment engendrent une perturbation dans le champ magnétique terrestre qui leur est propre. L'utilisation d'un capteur à effet Hall afin d'enregistrer ces variations permet alors de déterminer une position.

Cette méthode de positionnement, bien qu'indépendante de tout périphérique externe nécessite une première phase de calibration. Aussi l'utilisateur doit dans un premier temps disposer d'informations spécifiques au milieu étudié.

Deux approches sont alors possibles :

- L'utilisation des valeurs du champ pour chaque position, mesurées sur différents axes
- La reconnaissance de patterns représentatifs d'un élément

La première démarche s'appuie sur un principe dit de « Fingerprinting ». Dans une première phase de mapping l'usager utilise le capteur à effet Hall intégré à son téléphone afin de réaliser un ensemble de mesures. A différentes positions de l'espace il enregistre donc les valeurs du champ magnétique selon chacun des 3 axes du repère (O, x, y, z).

Figure 2. 1: Repère d'étude lié à l'apparail embarqué[Image Google]

Par la suite, lors de la phase de libre déplacement, le téléphone est en mesure de comparer les valeurs actuelles avec celles de la base de données précédemment constituée. Il repère alors l'usager à la position correspondante [18].

Néanmoins, certaines conditions doivent être réunies afin que les mesures acquises restent exploitables. Notamment, malgré les potentiels changements d'orientation du téléphone, il est

primordial que le repère d'étude reste fixe. Ce procédé est réalisable par exploitation des données gyroscopiques selon les axes de roulis, tangage et lacet. Une fois ces variations connues il est possible d'appliquer une matrice de rotation dans l'espace afin de conserver les différents axes selon une orientation stable.

Toutefois, le processus d'empreinte seul permet rarement d'assurer une bonne fiabilité de positionnement. En cause, les valeurs de champ pouvant être proches ou identiques, et ce, pour plusieurs positions de l'espace. Afin de contourner ce défaut, certains systèmes décident d'ajouter une dimension à l'équation : celle du temps, et donc de la « vitesse ». Ainsi ce ne sont plus trois valeurs qui sont accessibles mais trois valeurs et leurs dérivées.

Le principe consiste alors à enregistrer les perturbations magnétiques rencontrées lors du déplacement. Par la suite, un algorithme de reconnaissance de patterns permet d'indiquer la position de l'usager. Afin de réaliser cette correspondance, il est néanmoins nécessaire que celui-ci soit en mouvement.

Atouts et faiblesses : Bien que moins connu du grand public, le positionnement par champ magnétique apparaît comme un procédé prometteur et disponible à faible coût. Les systèmes offrent une bonne précision (couramment inférieure au mètre), le tout, sans aucune exigence d'infrastructure. Néanmoins l'usage de tels principes se limite à des configurations bien particulières. En effet, l'introduction d'une composante de mouvement suggère que l'ensemble des déplacements de l'usager soient prévisibles, et donc que l'environnement présente des parcours déjà définis (exemple d'un magasin avec différents rayons ou de couloirs étroits dans un bâtiment).De fait, la liberté de mouvement doit être limitée, sous peine d'introduire une complexité excessive. Par ailleurs le milieu doit exhiber une densité de mobilier suffisante, de manière à provoquer des perturbations locales, essentielles à l'algorithme de positionnement.

2.2.2Les technologies s'appuyant sur une infrastructure

2.2.2.1L'exploitation des ondes radio

Les ondes radiofréquences disposent de caractéristiques attrayantes d'un point de vue localisation. De fait, leur omniprésence au quotidien ainsi que leur capacité de pénétration des cloisons sont autant d'atouts justifiant leur utilisation. ***Aujourd'hui elles sont à la base de la majorité des systèmes de positionnement.*** Entre Ultrasons, Wi-Fi, Bluetooth LowEnergy, RFID ou encore UWB, les standards sont multiples et s'adaptent à de nombreux cas d'utilisation. L'architecture usuelle se compose d'un tag mobile, rattaché à l'usager ou à

l'équipement à localiser, ainsi que d'un réseau d'antennes, déployées dans l'environnement (voir **figure 2.2**)

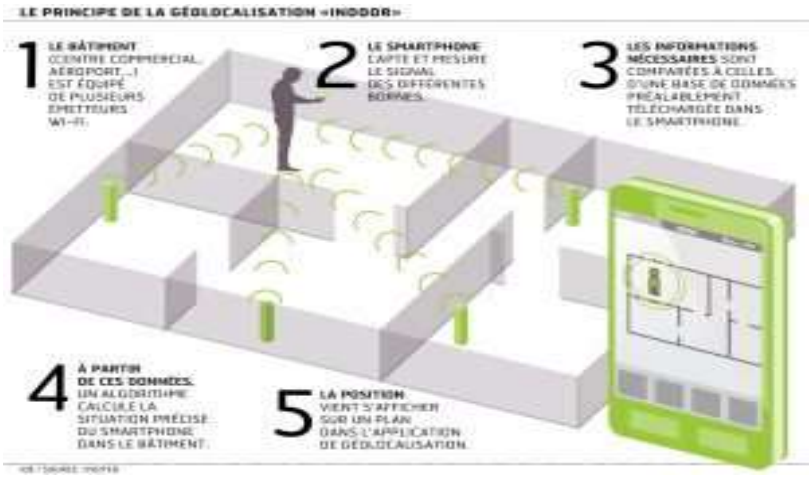


Figure 2. 2: Exemple d'utilisation des signaux Wi-Fi pour localisation indoor (Principe de Fingerprinting appliqué aux ondes radio) [19]

Deux cas de figure peuvent alors se présenter :

- Soit le tag émet et les antennes sont à l'écoute
- Soit chaque antenne émet à destination du tag, chargé de traiter l'information

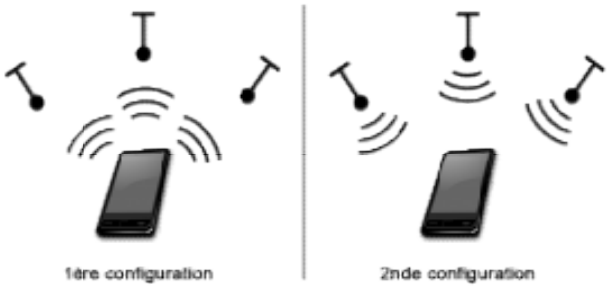


Figure 2. 3: Configuration possibles dans les systèmes de positionnement par ondes radio

Chacune de ces configurations propose divers avantages :

La première permet la réduction d'éléments « actifs » au sein du système. En effet, seul le tag est chargé d'émettre à destination des antennes environnantes, ce qui facilite notamment les aspects de synchronisation. Après réception du signal celles-ci communiquent avec un serveur en charge d'effectuer le traitement algorithmique. Grâce à cette approche, les ressources de calcul requises au niveau du tag sont réduites, la position de l'utilisateur est par ailleurs rendue accessible au plus grand nombre par l'accès au serveur. La seconde configuration concentre l'ensemble du traitement au niveau de l'équipement à localiser, cela permet une

mise à jour immédiate de la position sans avoir recours à un service externe. La disponibilité d'une connexion réseau n'est plus nécessaire.

2.2.2.2 Les techniques d'imagerie

Cette première méthode consiste à équiper le lieu dans lequel la localisation est souhaitée avec un ensemble de capteurs (caméras, LIDAR) qui restent à une position fixe et connue. Ces capteurs repéreront l'objet à surveiller et en déduiront la position.

De tels procédés reposent sur un même concept : analyser l'environnement afin d'en extraire des anomalies. Les supports sont divers, allant de la caméra simple ou dotée de fonctionnalités infrarouges, aux systèmes de traitement laser comme les lidars.

En ce qui concerne les caméras, leur usage est fréquent dans des cadres de surveillance ou de détection. Grâce aux techniques de reconnaissance de contours, un objet est ainsi repérable sur une image tant qu'il reste dans le champ de vision.

Si ces caméras sont souvent fixes, elles peuvent également être directement embarquées sur un équipement mobile, solidaire de la personne à localiser (procédé notamment utilisé en robotique). Similairement, le positionnement s'appuie sur un traitement de l'image, l'opération permet alors une estimation des distances et un repérage fidèle dans l'espace.

Bien que plus coûteuse, l'emploi de lidars est également une alternative efficace permettant de reconstituer l'environnement. Ici le balayage est réalisé par un faisceau laser, le système est alors en mesure de notifier des mouvements dans la pièce tout en réalisant une cartographie 3D. C'est ce principe qui est notamment utilisé dans le contexte automobile pour la fonctionnalité de régulateur de vitesse adaptatif (ACC – Adaptative Cruise Control).

Atouts et faiblesses : Les systèmes d'imagerie atteignent souvent une très bonne précision en milieu fermé, de surcroît ils ne nécessitent aucun dispositif rattaché à l'utilisateur, ce qui constitue un avantage certain (le système est dit « device free »).

Cependant leurs limites apparaissent en ligne de vue non directe (NLOS). Enfin, lorsque plusieurs personnes sont présentes, des difficultés d'identification surviennent. Des solutions existent, mais cela requiert des capacités de calculs élevées, venant alourdir la facture des équipements à déployer.

La solution *pourrait résider dans le développement d'algorithmes d'apprentissage en profondeur en début du système d'authentification des personnes. Ce qui veut dire dans la phase détection et aussi la classification. Ceci nous motive à pousser nos investigations sur l'application du Deep Learning pour la détection après une localisation basé sur les algorithmes de traitement d'image.*

En pratique le choix de la solution est à la charge du concepteur, selon l'utilisation souhaitée et la technologie support. En somme, la figure suivante synthétise la classification du chapitre:

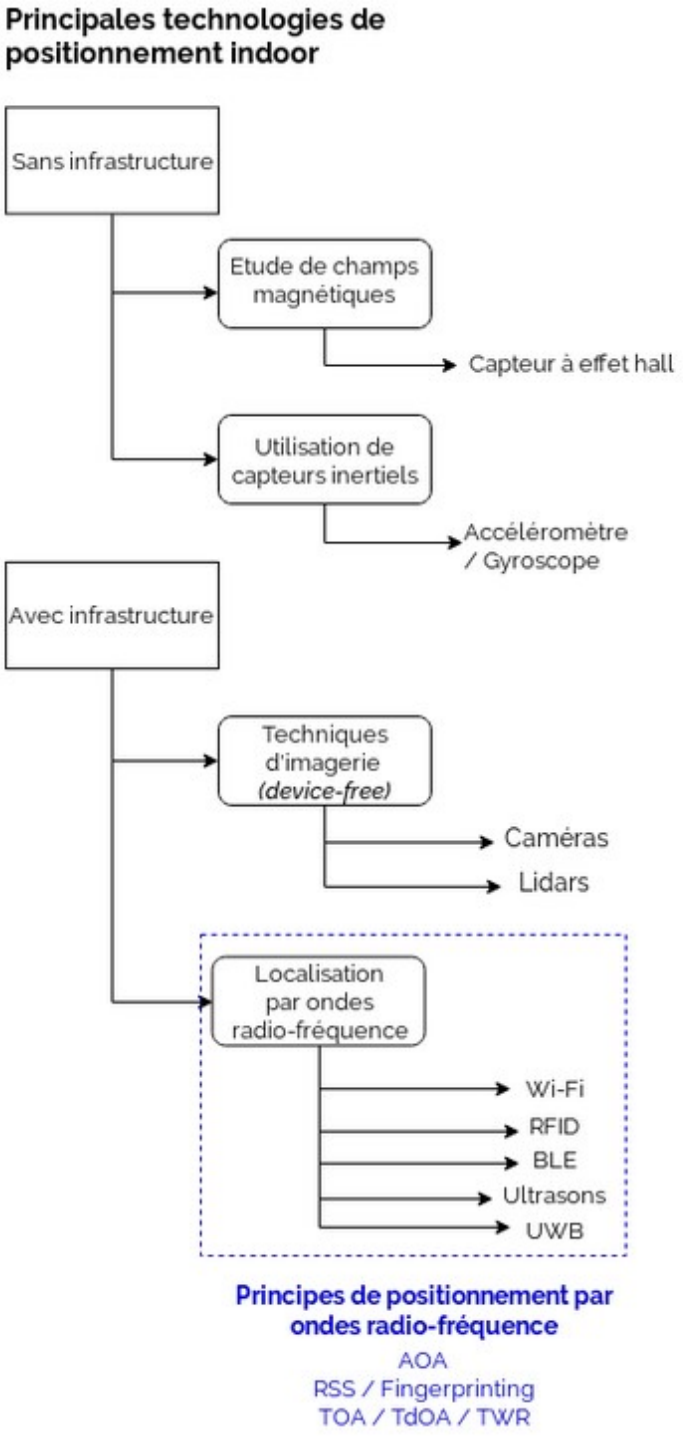


Figure 2. 4: Classification des technologies de positionnement indoor selon l'architecture à déployer [Image Google]

Cette première partie amorce la présentation des divers procédés de localisation en milieu fermé. Si les ondes radiofréquences dirigent aujourd'hui le marché, certaines architectures

apparaissent plus propices à d'autres contextes d'utilisation (systèmes device-free dans un contexte de surveillance, infrastructure-basé lorsque le milieu est connu). Les principes de positionnement par ondes radio restent encore énigmatiques et fascinants.

2.3 Techniques de localisation par les techniques d'imagerie basées sur le DL:

La détection d'objets est l'un des domaines de la vision par ordinateur qui mûrit très rapidement. Merci à l'apprentissage en profondeur! Chaque année, de nouveaux algorithmes et modèles continuent à surperformer les précédents. En fait, l'un des tout derniers logiciels de détection d'objet a été publié la semaine dernière par l'équipe d'intelligence artificielle de Facebook. Le logiciel s'appelle Detectron. Il intègre de nombreux projets de recherche pour la détection d'objets et est alimenté par le framework d'apprentissage en profondeur Caffe2.

Dans cette partie, l'accent est mis principalement sur les méthodes de localisation d'objet utilisant RL. Par conséquent, avant d'expliquer ces méthodes, le contexte RL sera brièvement présenté. Il convient de souligner que contrairement aux méthodes expliquées précédemment, ces algorithmes se concentrent simplement sur la localisation d'objets et non en fait les détecter.

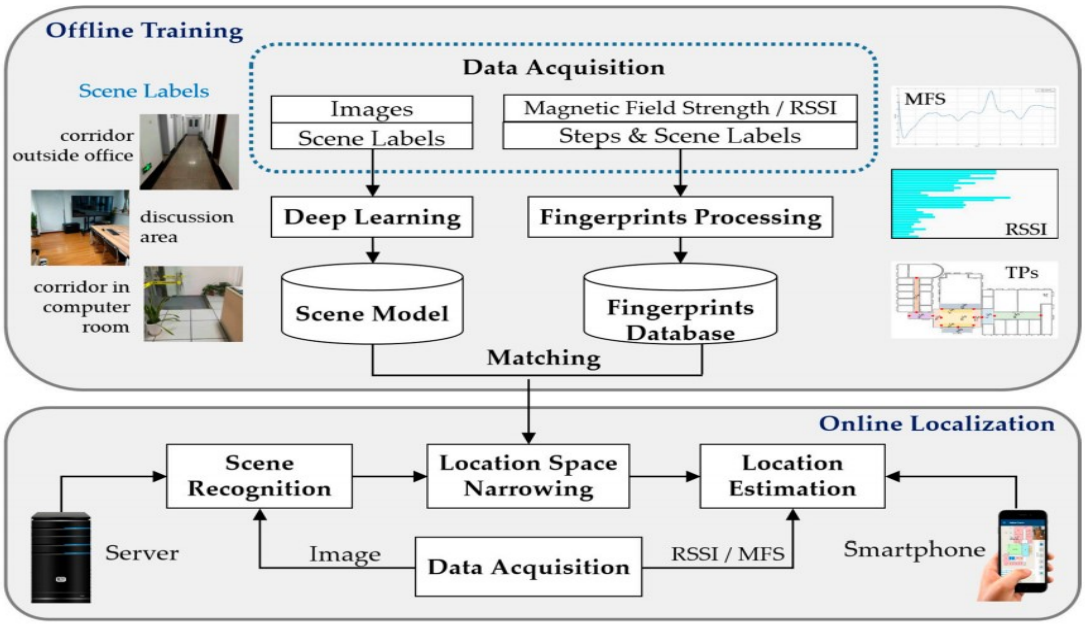


Figure 2. 5: Architecture d'un système de localisation et détection pour la reconnaissance [Google Image]

2.3.1 Apprentissage par renforcement

L'apprentissage par renforcement (AR ou RL) est considéré comme un cadre mathématique pour un apprentissage autonome basé sur l'expérience. En utilisant RL, un agent est capable

d'apprendre la structure d'un environnement par essais et erreurs dans le temps pour atteindre son (ses) objectif (s) [20]. RL conteste l'idée d'approches d'apprentissage automatique qui supposent toujours devrait être un enseignant (données annotées) pour apprendre une tâche. Au lieu d'avoir des données annotées, RL fournit une structure qui permet à un agent d'apprendre une tâche en maximisant les avantages à long terme en interagissant avec un environnement incomplètement connu. Basé sur la conséquence des interactions avec l'environnement, l'agent apprend à modifier son comportement en réponse aux récompenses reçues [21].

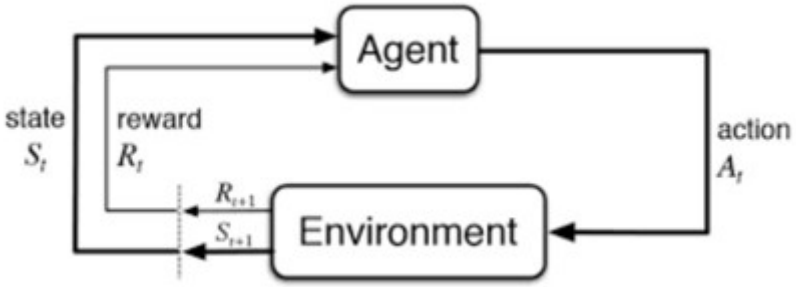


Figure 2. 6: la source de boucle d'apprentissage action de perception [22]

2.3.2 Apprentissage arborescent-structuré pour la localisation d'objets séquentielle

En utilisant une stratégie de recherche structurée en arborescence, une nouvelle méthode appelée Apprentissage par renforcement structuré en arborescence (Tree-RL) a été proposée pour localiser plusieurs objets dans une image. Tree-RL est une méthode basée sur la recherche arborescente où, à chaque étape, l'utilisation de RL décide de la division de la recherche arborescente. La recherche est sous forme de haut en bas et commence à partir de l'image entière jusqu'à la localisation d'un objet dans une feuille. L'état est défini comme étant le vecteur de caractéristiques de la fenêtre actuelle, le vecteur de caractéristiques de l'image entière et l'historique des actions entreprises. L'agent est autorisé à effectuer une action à partir de deux ensembles d'actions prédéfinis: l'un pour redimensionner la fenêtre actuelle en une sous-fenêtre et l'autre pour traduire localement la fenêtre actuelle. Dans un état donné, l'agent sélectionne les meilleures actions parmi les deux groupes d'actions, puis les actions sélectionnées sont effectuées dans deux branches distinctes d'une recherche arborescente. De cette manière, l'agent effectue simultanément une action de mise à l'échelle et une action de traduction locale à chaque état. Chaque chemin de la racine à une feuille de l'arbre de recherche fournit une recherche quasi optimale. Cela devrait améliorer la précision de

localisation pour les objets à différentes échelles. Pour calculer la récompense de chaque action, on utilise l'IoU. IoU calcule la proportion correspondant à la zone de chevauchement entre la fenêtre de l'agent actuel et le cadre de sélection fourni par la vérité au sol sur la zone de l'union de la fenêtre de l'agent et de la vérité au sol. Si l'action entreprise a pour résultat d'augmenter la valeur de l'IoU par rapport à celle-ci avant qu'elle soit prise, l'agent recevra une récompense positive. En raison des données d'entrée d'image continue de grandes dimensions et de l'environnement sans modèle, DQL est appliqué pour apprendre une stratégie optimale [23].

2.3.3 Localisation d'objet actif avec apprentissage par renforcement en profondeur

La localisation active des objets est [24] l'article principal sur lequel ce projet s'inspire. L'idée principale du document est de formuler la tâche de localisation d'objet en tant que processus de décision séquentiel dans lequel le terrain de jeu de l'agent est constitué d'images pour localiser des objets. L'agent va apprendre en se concentrant sur les régions denses dans les images où il pourrait y avoir des objets. L'agent doit déterminer les coordonnées des cadres de sélection liés aux objets d'une image. Plus spécifiquement, l'agent commence par analyser la scène entière et par réduire la fenêtre actuellement observée à un objet en effectuant une séquence d'actions. Dans l'article, la tâche de localisation d'objet est formulée en tant que processus de prise de décision séquentiel et peut donc être modélisée par le cadre MDP. Les états du MDP sont une représentation de la fenêtre de l'agent. Pour obtenir une représentation d'état, la fenêtre de l'agent est traitée par un modèle préformé et ses fonctionnalités sont extraites pour être utilisées, ainsi qu'un historique des actions prises par le passé. Le modèle pré-utilisé utilisé dans suit l'architecture proposée dans l'ensemble de données ImageNet 2012 et formée à cet effet pour la tâche de classification. Il est dit que l'utilisation d'un modèle de classification préformé peut aider à extraire des informations spatiales dans une scène et à réduire l'effort de formation. En ce qui concerne les actions, l'agent peut transformer sa fenêtre de huit manières différentes. Ils incluent le déplacement à gauche, à droite, en haut, en bas, la réduction et l'augmentation de la taille de la boîte, et le fait de rendre la fenêtre plus grosse ou plus grande. Il y a aussi une action de résiliation. L'agent apprend quand proposer la boîte en cours en tant qu'objet. Plus précisément, l'agent commence sa recherche en regardant l'image entière. Chaque action modifie la hauteur et la largeur de la fenêtre de l'agent d'un facteur fixe. De cette manière, contrairement à la méthode précédente, l'agent a quatre degrés de liberté pour changer l'emplacement de sa boîte. La liberté de déplacement de la fenêtre a rendu la méthode de localisation active différente des

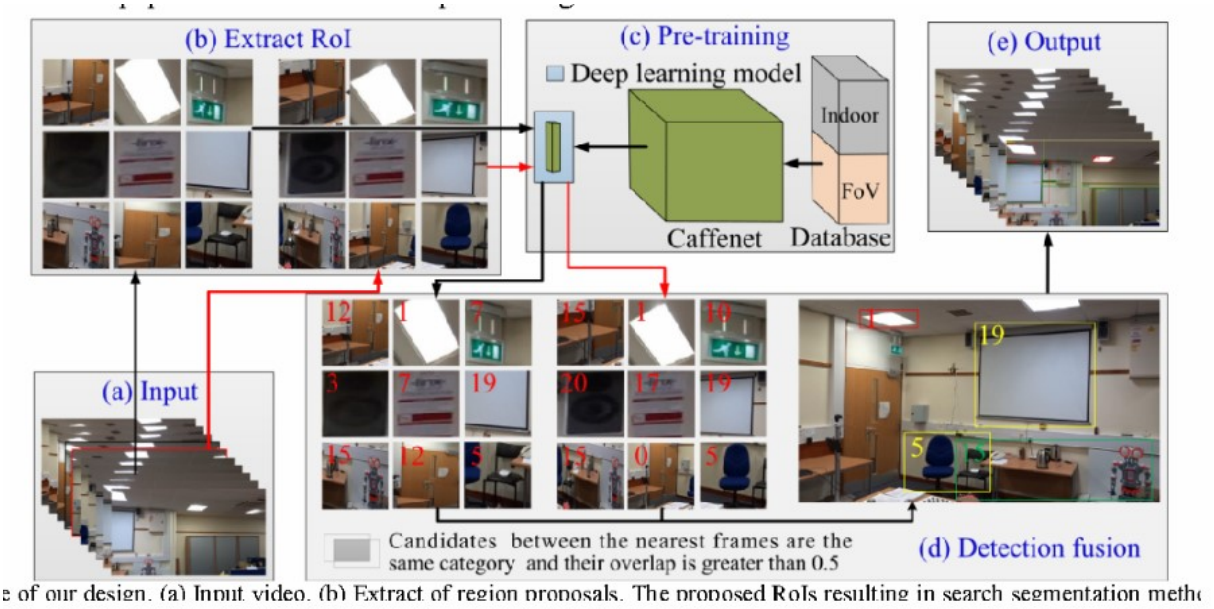
deux autres approches. Pour résoudre le problème de la conception d'un mécanisme de récompense, la localisation d'objet active utilise une fonction de récompense basée sur l'augmentation de l'IoU. Dans le cas d'une action augmentant l'IoU, l'agent obtiendrait une récompense positive +1 et sinon -1. Si l'agent localise avec succès un objet, il recevra +3. Enfin, pour obtenir une stratégie optimale de détection d'objet, DQL est utilisé pour résoudre le MDP, ce qui permet d'utiliser des entrées de grande dimension, c'est-à-dire des images.

2.3.4 Détection d'objets hiérarchique avec apprentissage par renforcement en profondeur

La détection d'objet hiérarchique avec apprentissage par renforcement en profondeur considère la tâche de détection d'objet comme un problème de décision séquentiel dans lequel l'environnement de l'agent est une image. Il formule le problème en tant que MDP et tente de le résoudre à l'aide de Q-learning. Contrairement à la «localisation d'objet active» et à la «détection d'objet hiérarchique», l'agent n'est pas libre de déplacer sa fenêtre lorsqu'il le décide par rapport à l'image d'entrée. Il existe plutôt un ensemble de fenêtres prédéfinies qui, lorsque l'agent effectue une action, sa fenêtre sera déplacé vers l'une des régions prédéterminées correspondant à l'action entreprise. L'agent commence par analyser l'image entière et décide où se concentrer parmi les fenêtres de sous-région. Il y a cinq fenêtres candidates pour l'agent, comprenant quatre quadrants de l'image plus une région centrale. En ce qui concerne les fenêtres de candidature, deux stratégies sont proposées pour déterminer les sous-régions. La première stratégie consiste à diviser l'image en sous-régions qui ne se chevauchent pas et la seconde à l'intersecté en plusieurs parties qui se chevauchent. En outre, les actions que l'agent peut effectuer sont classées en deux groupes dans lesquels l'agent peut décider de déplacer sa fenêtre et de modifier la région observée ou de mettre fin à la recherche d'objets. Semblable à d'autres travaux RL appliqués pour la détection d'objets, la fonction de récompense est basée sur l'IoU. Si une action entraîne une augmentation de l'IoU, l'agent recevra une récompense positive. De plus, avant qu'une fenêtre observée actuelle ne soit utilisée comme entrée dans Q-network, à l'aide d'un modèle pré-entraîné, les caractéristiques d'état sont extraites. Dans la détection d'objets hiérarchique, les modèles Image-Zooms et Pool45-Crops sont utilisés pour extraire des entités [25]. Chacun d'entre eux est utilisé séparément pour proposer un nouveau modèle. Après avoir extrait des entités en utilisant l'un ou l'autre des modèles, elles sont transmises à un réseau constitué d'une couche convolutive de 512 filtres de taille 7 × 7. La couche convolutionnelle est ensuite aplatie et sa dimension est réduite au nombre d'actions prédéfinies sur trois couches entièrement connectées.

2.3.5Reconnaissance d'objets indoor à l'aide d'un réseau de neurones convolutifs préformés

La reconnaissance d'objets indoor est une tâche clé pour la navigation intérieure dans les robots mobiles. Dans ce travail, un pipeline est proposé pour la détection d'objets d'intérieur basé sur le réseau de neurones à convolution (CNN). Avec la méthode proposée, on commence par préformer un modèle CNN hors ligne en utilisant à la fois un jeu de données public intérieur et un jeu de données vidéo privé (FoV). Cette opération est ensuite suivie d'un processus de recherche sélective pour extraire une région d'intérêt (RoI) après l'analyse de la vidéo d'entrée dans des images par cadre. Les ROI extraites sont ensuite classées en candidats à l'aide du modèle profond préformé et les candidats situés entre les images de trame les plus proches sont affinés à l'aide de la fusion par détection. Enfin, les images annotées sont fusionnées pour créer une vidéo en sortie.



e of our design. (a) Input video. (b) Extract of region proposals. The proposed RoIs resulting in search segmentation meth

Figure 2. 7: Exemple de Pipeline conçu pour la reconnaissance d'objet indoor[]

(a) Vidéo d'entrée. (b) Extrait des régions d'intérêts proposées. Les ROI (region of Interest) proposés résultant de la méthode de segmentation de la recherche sont redimensionnés en 256 × 256. (c) formation préalable. La base de données utilisée pour la formation est une base de données intérieure mélangée à des images vidéo. FoV est l'abréviation d'images de vidéos. (d) Fusion de détection. La détection est confirmée si la proposition entre les images les plus proches appartient à la même catégorie et si leur chevauchement est supérieur à 0,5. (e) Vidéo de sortie [26].

Les expériences montrent que cette conception est très efficace contre la détection d'objets d'intérieur.

Conclusion

Dans ce chapitre nous avons présenté des méthodes et des techniques de localisation. Nous avons ensuite étudié certaines méthodes basées sur les algorithmes de traitement d'image et aussi l'apport du Deep Learning à ce domaine de localisation et détection indoor. Divers articles ont également parlé d'études basées sur l'énergie et de modèles énergétiques pour quantifier et estimer la consommation d'énergie des applications par le biais de services et d'appels système sur les téléphones cellulaires. Bien que ces modèles soient peut-être généralisés, chaque application est unique et le même modèle pourrait ne pas être bon pour une application particulière et ce qui pousse les chercheurs donc à créer leur propre modèle d'énergie pour leurs applications. Certaines techniques ont utilisé un moniteur de puissance pour quantifier la puissance réelle consommation sur l'appareil. Il y a eu beaucoup de recherches d'optimiser la consommation d'énergie des interfaces sans fil telles que la 3G, le Wi-Fi, le Bluetooth, etc. Le Wi-Fi principalement utilisés pour la technique de navigation, dans ce cas les travaux se concentrent sur la modélisation énergétique robuste de l'interface Wi-Fi. D'autres travaux portent sur des systèmes de détection de position à efficacité énergétique visant à réduire la consommation de batterie élevée causée par les interfaces de localisation (WiFi, GPS, par exemple). Certains chercheurs proposent un mécanisme d'enregistrement à taux variable (VRL) qui désactive la journalisation de la localisation ou réduit le taux d'enregistrement GPS en détectant si l'utilisateur est immobile ou à l'intérieur. La plupart de ces travaux sur l'optimisation énergétique concerne la navigation en extérieur et rien sur la navigation en intérieur.

La plupart des systèmes ne parviennent pas à donner une précision de positionnement élevée avec un déploiement facile. Dans ce type de système, on n'a pas besoin d'extraire des fonctionnalités dans les images par rapport à d'autres méthodes basées sur la vision. Les méthodes d'apprentissage en profondeur ont seulement besoin d'attribuer des données de formation étiquetées à un réseau prédéfini; il apprend les fonctionnalités automatiquement.

Chapitre 3 :

Méthodologies de Localisation et Deep Détection

Introduction

Les algorithmes de localisation d'objets se composent de deux groupes: les algorithmes qui effectuent la localisation d'objets et la détection conjointement et les algorithmes qui se concentrent principalement sur la localisation des objets.

Dans cette section, nous passons en revue les algorithmes de détection d'objet, qui effectuent conjointement la détection et la localisation d'objet. La caractéristique commune de ces catégories d'algorithmes est qu'elles doivent traiter tous les sous régions d'une image afin de détecter des objets. En d'autres termes, ces algorithmes doivent analyser de nombreuses propositions de régions qui les ont ralenties. On trouve dans la littérature plusieurs méthodes de détection en profondeur d'objets (Objects Deep Detection). Toutes cette méthodologie est destinée aux applications localisation et/ou reconnaissance.

3.1 Modèle YOLO combinant la détection et la localisation

Le modèle You Only Look Once (YOLO) peut être traduit par « voir uniquement une fois » est un modèle qui consiste en un seul réseau de neurones qui peut être formé de bout en bout par rétro-propagation. Il fusionne les deux étapes des algorithmes précédents, la détection d'objet et la localisation, dans un modèle [27]. Cependant, YOLO formule le problème de détection d'objet différemment, en tant que tâche de régression qui sépare spatialement les cadres de sélection et associe les probabilités de classe. Le réseau est formé pour apprendre des représentations très générales d'objets. Il prend en entrée l'image entière et prédit les cadres de délimitation simultanément pour toutes les classes d'une image. L'image est divisée en une grille S × S, puis des boîtes englobantes B et des scores de confiance sont prévus pour ces boîtes. Si le centre d'un objet se trouve dans une cellule de la grille, celle-ci prédira les cadres de sélection B avec un score de confiance. Pour calculer le score de confiance, il faut calculer Intersection over-Union (IoU), qui correspond à la différence entre la boîte prévue et la vérité sur le terrain. De cette façon, le score de confiance qui est Probabilité (Object) × IoU peut être calculé. L'avantage de YOLO par rapport aux autres méthodes est sa rapidité. Il est si rapide par rapport aux autres méthodes, ce qui le rend idéal pour les applications en temps réel. Cependant, cela vient avec un compromis en termes de précision. YOLO fait plus d'erreurs de localisation. L'architecture du réseau est illustrée à la **figure 3.1**.

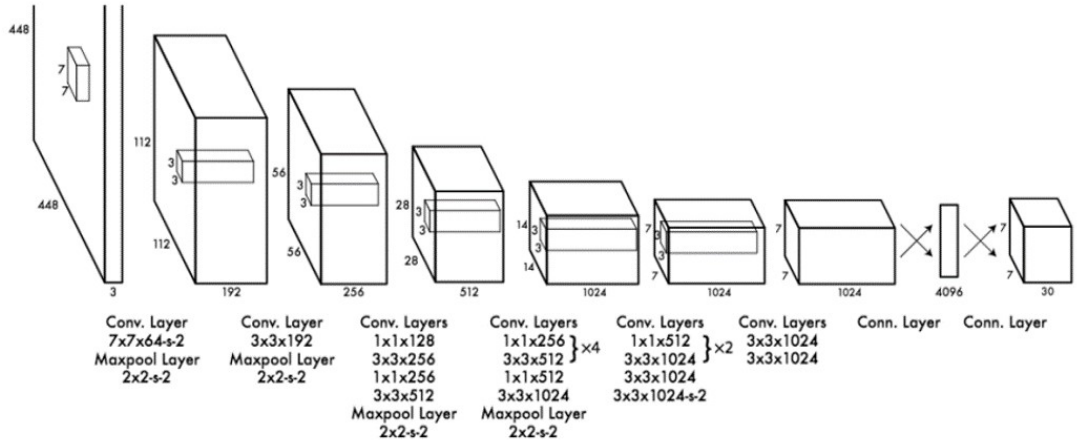


Figure 3. 1: Architecture du modèle YOLO [27]

La **figure 3.1** nous montre que l'architecture est formée de 24 couches de convolution et 2 couches FC.

3.2 Réseau de recherche d'architecture neuronale (NASNet)

Le réseau de recherche d'architecture neuronale Neural Architecture (Search Net (NASNet)) [28] suit un paradigme totalement différent pour la détection d'objets. L'idée de NASNet est d'apprendre l'architecture des réseaux de neurones. Alors que les chercheurs passent leur temps à trouver une conception d'architecture optimisée de réseau neuronal, NASNet applique un réseau de neurones récurrents (RNN) qui peut apprendre une architecture optimisée pour un objectif spécifique par rapport à un jeu de données. NASNet apprend combien de couches, de filtres et de neurones sont appropriés pour un problème donné. Le réseau est également capable d'apprendre de petits détails d'un réseau, tels que la hauteur des foulées (stride), la largeur des foulées et la taille du filtre dans le cas des CNN. Plus spécifiquement, NASNet est formé pour utiliser l'algorithme RL et la précision sur un jeu de données donné est utilisée comme signal pour former NASNet. Les auteurs de [28] ont créé une architecture de réseau neuronal apprise à l'aide de NASNet sur CIFAR-10, puis ont formé le réseau à Image-Net 2012. Ce modèle a ensuite été utilisé comme générateur de cartes de caractéristiques et empilé avec Faster R-CNN. Enfin, l'ensemble du pipeline a été recyclé avec l'ensemble de données COCO.

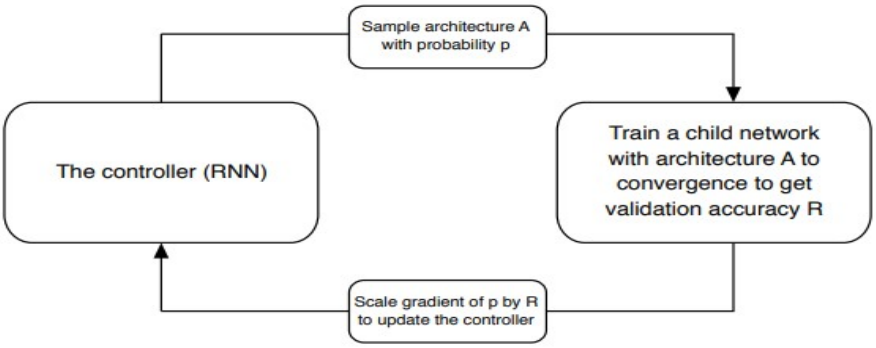


Figure 3. 2: Vue d'ensemble sur l'architecture NASNet [29]

L'art de la raison pour laquelle cet algorithme réussit est dû aux contraintes et aux hypothèses qu'il a formulées. L'architecture découverte par le NAS est formée et testée sur un jeu de données beaucoup plus petit que le monde réel. Ceci est fait parce que la formation sur quelque chose de grand, comme Image-Net, prendrait beaucoup de temps. Cependant, l'idée est qu'un réseau qui fonctionne mieux sur un jeu de données plus petit mais structuré de la même manière devrait également mieux fonctionner sur un jeu de données plus grand et plus complexe, ce qui est généralement vrai à l'époque de l'apprentissage en profondeur.

3.3 Réseau de convolution basé sur les régions (R-CNN)

Le réseau de convolution basé sur les régions (R-CNN) est le premier travail qui a appliqué la méthode d'apprentissage en profondeur dans les problèmes de détection d'objet. L'idée principale [30] est que l'algorithme trouve tous les objets dans une image en utilisant un algorithme de recherche exhaustive, puis classe les objets proposés en utilisant CNN. L'algorithme de recherche permettant de localiser des objets dans une image s'appelle recherche sélective. Cet algorithme de recherche a été conçu pour localiser des objets dans des images. L'algorithme de recherche sélective est capable de traiter avec une variété de conditions d'image. La base de l'algorithme de recherche sélective est l'algorithme hiérarchique de regroupement. En utilisant le groupement ascendant, l'algorithme de recherche sélective est capable de générer les emplacements d'objets à toutes les échelles. Le processus de regroupement se poursuit jusqu'à ce que l'image entière devienne une seule région. Les régions détectées sont ensuite traitées en utilisant divers espaces colorimétriques dotés de

propriétés d'invariance différentes, de mesures de similarité différentes et en faisant varier les régions de départ. La sortie de l'algorithme de recherche sélective est un ensemble de propositions de région pouvant contenir un objet. Le modèle R-CNN combine la recherche sélective et les méthodes CNN pour localiser et classifier les objets. Le R-CNN est composé de trois modules. Le premier génère un ensemble de régions de proposition utilisant la recherche de régions d'intérêt. Le second est un CNN pour extraire un vecteur de caractéristiques de 4096 dimensions de longueur fixe de chaque région. Le troisième module est un ensemble de classificateurs SVM linéaires dont l'entrée est le vecteur de caractéristiques et sa sortie est la probabilité d'appartenir à une catégorie d'objet. L'architecture de R-CNN est montrée dans la **figure 3.3.**

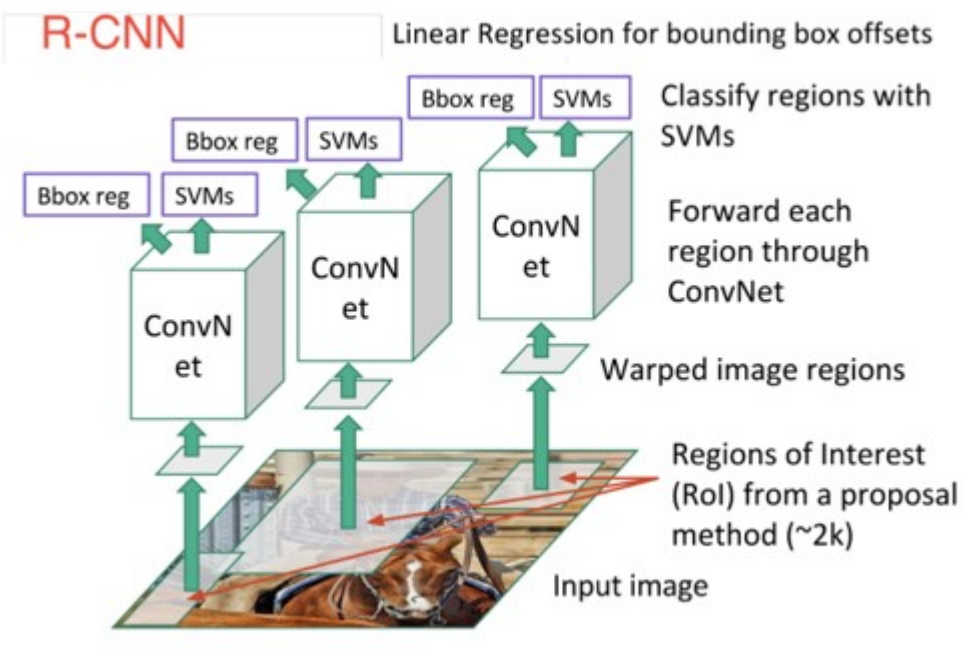


Figure 3. 3: Principe du réseau de convolution basé sur les régions (R-CNN) [31]

3.3.1 Fast R-CNN

Fast R-CNN [32] est une variante de R-CNN visant à accélérer la détection d'objets. R-CNN souffre de trois inconvénients majeurs :

1. La première est que l'algorithme consiste en plusieurs étapes qui sont apprises et réglées séparément.
2. La seconde est le temps de formation. Il est rapporté que pour 5K images de la VOC 2007, la formation prend 2.5 GPU-days.

3. Le dernier problème est au moment des tests, où il est nécessaire de réaliser des applications en temps réel, cependant, chaque image nécessite un traitement de 47 secondes.

Fast R-CNN est conçu pour réduire la quantité de calcul et de mémoire nécessaire à R-CNN en utilisant une perte multitâche pour entraîner l'ensemble du réseau en un seul passage et mettre à jour toutes les couches du réseau. Rapide R-CNN prend l'image entière en entrée et l'envoie au CNN principal. À l'aide de plusieurs convolutions et couches de regroupement, un vecteur de caractéristiques est extrait de l'entrée. Le vecteur de caractéristiques est utilisé par une couche de regroupement de régions d'intérêt pour extraire un vecteur de caractéristiques de longueur fixe pour chaque proposition d'objet. La méthode de recherche sélective est appliquée pour rechercher les régions RoI. Chaque vecteur de caractéristique est ensuite aplati pour être introduit dans des couches entièrement connectées qui génèrent finalement deux couches de sortie analogues :

1. Le premier est un vecteur de codage one-hot passé à travers une couche softmax pour indiquer la probabilité d'appartenir à K classes d'objets pour chaque objet de proposition.
2. La deuxième sortie est un vecteur à valeur réelle à quatre valeurs réelles pour chacune des K classes d'objets, qui code les coordonnées des boîtes de sélection prédites pour les objets détectés.

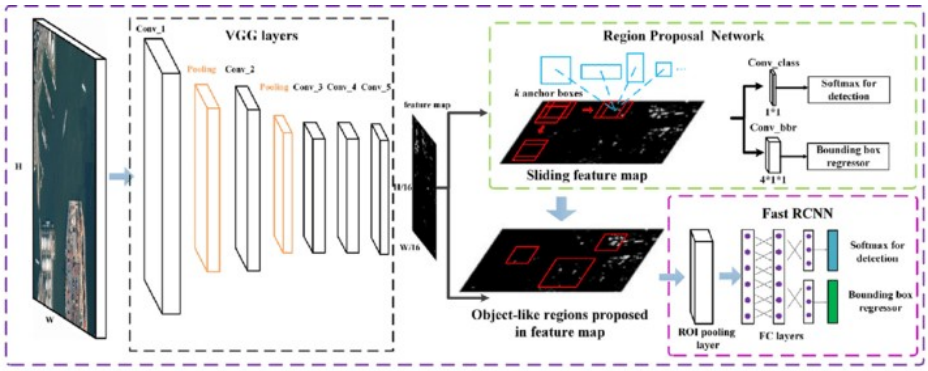


Figure 3. 4: Architecture du modèle Fast R-CNN [33]

3.3.2 Faster R-CNN

Faster RCNN [34] est une architecture de détection d'objets présentée par Ross Girshick, ShaoqingRen, Kaiming He et Jian Sun en 2015. Faster R-CNN est conçu pour remplacer l'algorithme de recherche sélective utilisé dans les versions

précédentes de R-CNN. Le problème avec la recherche sélective est qu'elle est coûteuse en calcul. Bien que Fast R-CNN ait introduit de nouvelles fonctionnalités pour réduire le temps d'entraînement et de test, la recherche sélective restait un goulot d'étranglement pour les algorithmes R-CNN. Dans R-CNN plus rapide, un nouveau réseau appelé réseau de proposition de région (RPN) a été introduit pour remplacer l'algorithme de recherche sélective. Ce réseau vise à proposer des régions qui seront utilisées ultérieurement par le réseau Fast R-CNN pour prévoir les boîtes englobant et détecter les objets. RPN utilise un modèle préformé sur le jeu de données Image-Net pour la classification. Plus précisément, le réseau RPN, qui est un réseau de convolution profonde qui propose des régions, prend une image en entrée et génère en sortie une carte de caractéristiques. La carte de caractéristiques est ensuite utilisée par un petit réseau. Le petit réseau prend en entrée une fenêtre glissante n × n sur la carte des caractéristiques. La sortie du petit réseau est une sortie équivalente, une couche de régression par case et une couche de classification par case. Sur chaque emplacement de fenêtre, le petit réseau prédit plusieurs propositions de région. Le nombre de propositions de région est défini par un paramètre appelé K. Les K régions proposées déterminent le nombre de zones de référence appliquées à tous les emplacements de fenêtre pour créer des propositions de région. Ces boîtes ont des échelles et des formats d'image différents pour capturer tous les objets possibles à la position de glissement actuelle et sont appelées ancres. De cette manière, il existe une ancre K pour chaque fenêtre coulissante. En utilisant des ancres, Faster R-CNN peut traiter plusieurs échelles et formats. La couche de classification de boîtes génère un vecteur de probabilités indiquant un score d'objectivité pour chaque boîte d'ancrage. Les boîtes d'ancrage détectées sont ensuite sélectionnées en fonction du score d'objectivité. Les boîtes d'ancrage dépassent un seuil prédéfini, puis sont acheminées vers Fast R-CNN. Il est à noter que Faster R-CNN fusionne le réseau RPN avec Fast R-CNN en utilisant un mécanisme appelé «attention mechanism». Le réseau RPN guide le réseau Fast R-CNN où chercher. Pour partager le calcul, les fonctions de convolution sont partagées entre RPN et Fast R-CNN. Le reste de l'algorithme est similaire à Fast R-CNN. Faster R-CNN est composé de 3 parties comme le montre la **figure 3.5.**

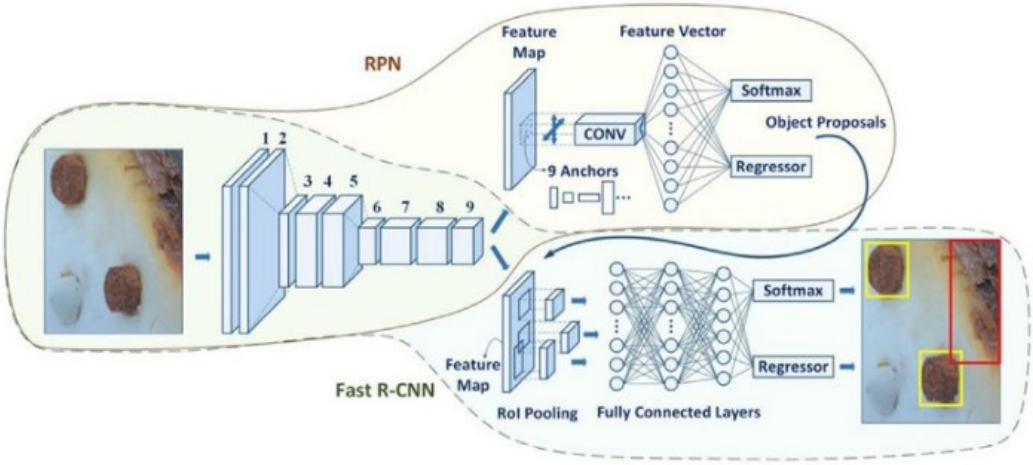


Figure 3. 5 : Architecture de Faster R-CNN [35]

3.3.3 Réseau de convolution basé sur le masque de région (Mask R-CNN)

Le réseau basé sur le masque de région CNN (Mask Region-based Convolutional Network (**Mask R-CNN**)) [36] étend Faster R-CNN en ajoutant une nouvelle partie à la reconnaissance du cadre de sélection afin de prédire un masque d'objet. Le masque R-CNN utilise les deux mêmes étapes que Faster R-CNN. Dans la première étape, Mask R-CNN adopte une architecture RPN identique, mais la deuxième étape vient avec une extension du Faster R-CNN d'origine. Dans la deuxième étape, outre les boîtes englobantes et les prédictions de classe, le masque R-CNN produit un masque binaire pour chaque RoI. La représentation de masque est utilisée pour démontrer la disposition spatiale de l'objet. En outre, la couche de regroupement RoI de Faster R-CNN, initialement héritée de Fast R-CNN, est également remplacée par une couche RoI Align. Le problème avec une couche de regroupement RoI traditionnelle est qu'elle quantifie un RoI à nombre flottant, ce qui provoque un malentendu entre le RoI et les entités extraites. Cette quantification n'affecte pas la classification mais a un impact négatif significatif sur la prédiction de masques précis au pixel près. L'idée d'ajouter un masque de pixels à l'algorithme signifie que la tâche de segmentation améliore la localisation et donc la classification. Pour atteindre cet objectif, trois fonctions de perte correspondant à chaque tâche sont définies et totalisées. L'erreur totale est ensuite utilisée pour optimiser et former le réseau.

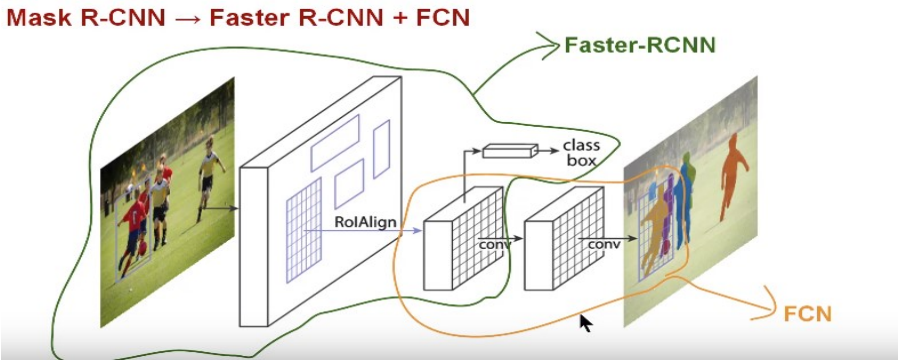


Figure 3. 6: 6Architecture du Mask R-CNN pour la Détection+ Segmentation [37]

3.4 Région basée sur réseau global de convergence(RFCN)

La méthode de région basée sur un réseau global de convergence **(Region-based Fully Convolutional Network)** utilise des algorithmes à plusieurs étages de R-CNN qui consistent à détecter des propositions de région et à reconnaître un objet dans chaque région. Le réseau entièrement convolutif basé sur la région (RFCN) [38] est un modèle comportant uniquement des couches de convolution pouvant être formées de bout en bout et se terminer par une propagation en arrière. Cela apporte l'avantage d'utiliser des réseaux résiduels (ResNual) pour la première partie de l'algorithme qui détecte les objets. La différence entre le RFCN et les algorithmes précédents réside dans le fait que chaque partie de l'algorithme peut être entraînée de bout en bout sur un réseau alors qu'avant, chaque étape des algorithmes consistaient en différentes parties et était entraînée séparément. Ce modèle peut prendre en compte simultanément la détection d'objet et la localisation d'objet. RFCN utilise le réseau RPN pour extraire les régions de proposition (intérêt) (RoI : Region of Interest). La dernière couche de RPN produit les cartes de caractéristiques, chacune détectant une catégorie d'objet à un emplacement spécifique. Après avoir extrait les propositions de région, RFCN les classe en catégories d'objet et en arrière-plan. La dernière couche de RFCN produit des cartes de scores sensibles à la position qui votent pour la position des objets dans une image. À la fin, une couche de regroupement RoI sensible à la position est utilisée pour agréger la sortie de la dernière couche de convolution et produit des scores pour chaque région RoI. La couche de regroupement RoI sensible à la position effectue un regroupement sélectif et renvoie un score de la banque de cartes de scores K × K. L'ensemble du processus est illustré à la **figure 3.7.**

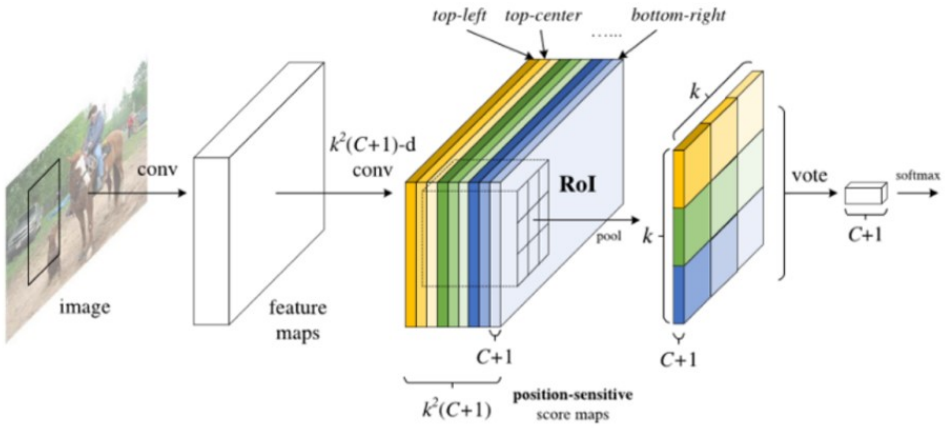


Figure 3. 7: Principe de RFCNN [39]

L'image d'entrée est liée au modèle ResNet, puis sa sortie est utilisée par le modèle RPN pour détecter le retour sur investissement et calculer les probabilités correspondant à chaque classe d'objet.

3.5Détecteur à coup (SSD)

Bien que les approches précédentes soient précises, elles sont trop lentes et trop coûteuses en calcul pour les systèmes embarqués et les applications en temps réel. Dans ces cas, la rapidité est souvent au détriment de la précision. Pour améliorer la vitesse, le détecteur à coup ou Single-Shot Detector en anglais, (SSD) a éliminé l'étape de proposition de région du pipeline de détection d'objet. Ainsi, le réseau de neurones n'a pas besoin de ré-échantillonner des entités pour produire des hypothèses de cadres de sélection. À l'aide d'un filtre de convolution, SSD prédit les catégories et les décalages d'objet dans les emplacements des boîtes englobantes. En outre, un autre filtre de convolution est utilisé pour effectuer la détection d'objet à différentes échelles. Les filtres sont appliqués sur les cartes de caractéristiques de la première partie du réseau de neurones. Cela conduit à un algorithme plus rapide et plus précis que les précédents. Semblable à YOLO et à RFCN, SSD propose un modèle consistant en un seul réseau de neurones à convolution qui peut être formé de bout en bout. Plus précisément, le SSD est basé sur un réseau de neurones à convolution à réaction qui produit un ensemble de boîtes englobantes et de scores pour la présence de classes d'objets [27]. La première partie du réseau neuronal appelée réseau de base suit une architecture standard (architecture VGG-16) et est responsable de l'extraction des caractéristiques. La deuxième partie du réseau produit un ensemble de prévisions. Ces prévisions comprennent les coordonnées prédites des boîtes englobantes,

notamment les coordonnées du centre, de la largeur et de la hauteur de la boîte. De plus, le réseau génère un vecteur de probabilités liées à la confiance pour chaque classe d'objets. De plus, deux autres méthodes sont utilisées pendant le temps d'entraînement. Pour conserver les zones les plus pertinentes, une méthode appelée suppression non maximale est utilisée, puis le résultat de celle-ci est consommé par la méthode Hard Negative Mining répertorie les cases négatives prévues en fonction du score de confiance et en sélectionne un sous-ensemble à utiliser pour le calcul de l'erreur. En effet, de nombreuses boîtes négatives sont prévues pendant la formation et pourraient avoir un effet destructeur sur la formation du réseau. L'architecture du réseau SSD est présentée dans la **figure 3.8**.

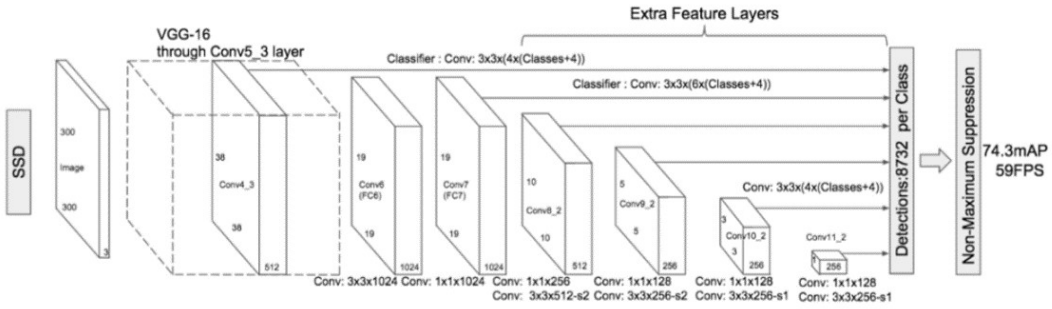


Figure 3. 8: Architecture du modèle SSD[40]

Le modèle SSD applique plusieurs couches aux cartes de caractéristiques générées par le réseau de base pour augmenter le nombre de cadres de sélection pertinents.

Conclusion

Ce chapitre propose une revue de littérature cohérente pour des algorithmes de détection d'objets qui appliquent en partie ou principalement les méthodes DL. Ces algorithmes traitent en premier lieu la localisation et détection et aussi la détection en profondeur. Les algorithmes de détection d'objets se composent de plusieurs étapes. Deux étapes communes dans chacune d'elles ont d'abord été de localiser un objet, puis de le classer en utilisant un classificateur, par exemple CNN. Cependant, le problème avec ces méthodes était le coût de calcul causé par la méthode utilisée pour produire des propositions de région et le fait que les algorithmes avaient plusieurs étapes pour traiter une image. En d'autres termes, toutes ces méthodes sont basées sur des classificateurs redéfinis pour la détection d'objets.

Chapitre 4 :

Conception du modèle Localisation et Deep Détection

Introduction

La localisation d'une personne et la détection de visage est une des tâches visuelles que les humains peuvent accomplir sans effort. Cependant, en termes de vision par ordinateur, cette tâche n'est pas facile. Une formulation générale du problème peut être définie comme suit: étant donnée une image statique ou une vidéo (séquences images), on veut détecter un nombre inconnu (s'il existe) de visages. La solution à ce problème implique la segmentation, l'extraction, et la vérification des visages et probablement des traits faciaux à partir d'un arrière plan non contrôlé. Comme processeur visuel d'entrée, un système de détection de visage devrait également pouvoir réaliser la tâche indépendamment de l'illumination, de l'orientation, et de la distance de la camera. Ce chapitre vise à fournir un aperçu de la recherche contemporaine en détection de visage et d'une manière structurée. Les auteurs dans [94] ont conduit une étude détaillée sur la recherche en reconnaissance de visage. Dans leur étude, plusieurs aspects, y compris la segmentation et l'extraction des traits, liées à la reconnaissance de visage ont été passés en revue. Une de leurs conclusions était que le problème de détection de visage a suscité étonnamment peu d'attention. Ceci a certainement changé au cours des dernières années comme on l'a montré dans le chapitre 3. La méthodologie de ce chapitre nous pousse à concevoir un modèle de localisation basé sur la détection en profondeur. C'est ce qui fait l'objet de notre chapitre 4.

4.1 Etude de la calibration des images

La calibration reste une phase primordiale et fondamentale pour la localisation [41].

4.1.1 Calibration d'un système

Dans cette phase nous déterminons la position et l'orientation des capteurs du système par rapport à un repère absolu ainsi que ses caractéristiques internes.

Figure 4. 1: Principe de calibration [41]

Par exemple :

1. **Calibration d'une caméra:** déterminer la transformation permettant de passer d'un objet défini dans un repère absolu à son image. Savoir passer du repère absolu au repère caméra.
2. **Calibration d'une table micrométrique:** une caméra observe un objet en rotation sur une table. Connaitre la transformation repère table au repère caméra.
3. **Recalage caméra/capteur:** un capteur de position et une caméra sont utilisés pour localiser la caméra. Comment les faire parler dans le même repère?

4.1.2 Calibration d'une caméra

Dans cette phase nous déterminons la transformation ponctuelle faisantpasser du point 3D exprimé dans un repère absolu à son image. Pour cela, nous devons :

1. Modéliser l'optique de la caméra (paramètres intrinsèques)
2. Déterminer la transformation repère absolu/ repère caméra (paramètres extrinsèques)

4.1.3 Modélisation d'une caméra

La figure 4.2 illustre la modélisation d'une caméra.

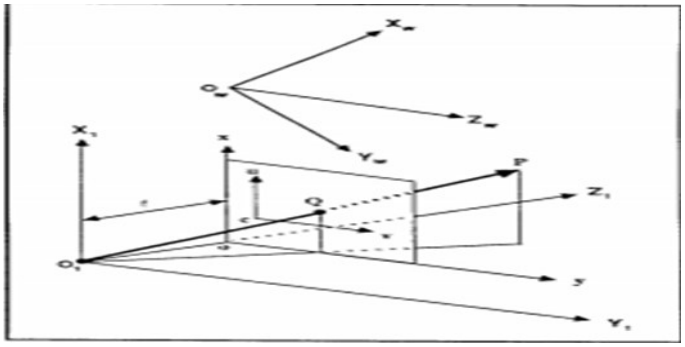


Figure 4. 2: Modélisation d'une caméra

- **Relation de base**

$P(X_1, Y_1, Z_1)$ dans $R_1 = (O_1, x_1, y_1, z_1)$

$Q(x, y)$ dans (O, x, y).

$$\frac{x}{f} = \frac{X_1}{Z_1} \quad \frac{y}{f} = \frac{Y_1}{Z_1} \tag{4.1}$$

Modèle de la rétine avec axes perpendiculaires :

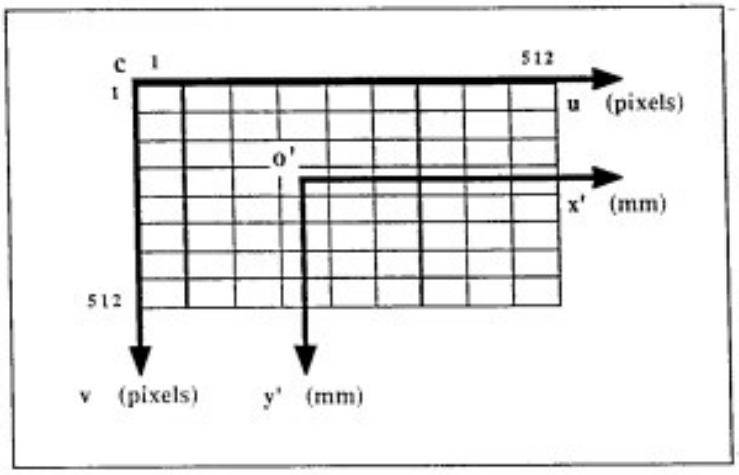


Figure 4. 3: Modèle de la rétine avec axes perpendiculaires

(u, v) mesurés en pixels (x, y) mesurés en mm.

Notations:

1/ k_u la dimension horizontale du pixel en mm.

1/ k_v la dimension verticale du pixel en mm.

(u_0, v_0) les coordonnées pixels de la projection de l'axe optique.

$$\begin{array}{rcl} x & = & \frac{u-u_0}{k_u} \\ y & = & \frac{v-v_0}{k_v} \end{array} \Rightarrow \begin{array}{rcl} u & = & k_u x + u_0 \\ v & = & k_v y + v_0 \end{array} \tag{4.2}$$

Relation entre points dans le repère caméra et coordonnées pixels

$$\begin{array}{rcl} u & = & k_u f \frac{X_1}{Z_1} + u_0 \\ v & = & k_v f \frac{Y_1}{Z_1} + v_0 \end{array} \Rightarrow \begin{array}{rcl} Z_1 u & = & k_u f X_1 + u_0 Z_1 \\ Z_1 v & = & k_v f Y_1 + v_0 Z_1 \end{array} \tag{4.3}$$

Seuls peuvent être détermine: $\alpha u = k_u f$, $\alpha v = k_v f$, u_0, v_0 (paramètres intrinsèques).

Ecriture matricielle en coordonnées homogènes:

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = \underbrace{\begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix}}_{K} \begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} \quad (4.4)$$

Le passage repère absolu/repère caméra est un déplacement définie par une rotation R et une translation T.

$$\begin{pmatrix} X_1 \\ Y_1 \\ Z_1 \end{pmatrix} = R \begin{pmatrix} X_w \\ Y_w \\ Z_w \end{pmatrix} + T = [R\ T] \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (4.5)$$

D'où

$$\begin{pmatrix} su \\ sv \\ s \end{pmatrix} = K \times [R\ T] \begin{pmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{pmatrix} \quad (4.6)$$

$$P = K \times [RT] \quad (4.7)$$

P est la matrice 3×4 de projection perspective. Elle est définie à un coefficient multiplicatif prés: P et λP déterminent les mêmes points projetés. Calibrer une caméra, c'est déterminer P.

- **Que fournit K?**

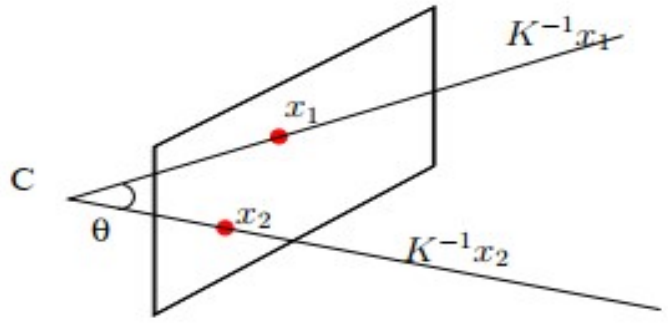


Figure 4. 4: Représentation de la transformation K

- K relie le point image à la direction du rayon issu de ce point.

- Si x est un point de l'image $d = K^{-1}x$ est la ligne de vue exprimée dans le repère euclidien de la caméra.
- Angle entre deux directions de vues:

$$\cos(\theta) = \frac{d_1.d_2}{||d_1||||d_2||} = \frac{(K^{-1}x_1)^t K^{-1}x_2}{\sqrt{(K^{-1}x_1)^t K^{-1}x_1}\sqrt{(K^{-1}x_2)^t K^{-1}x_2}} =$$

$$= \frac{x_1^t K^{-t} K^{-1} x_2}{\sqrt{x_1^t K^{-t} K^{-1} x_1}\sqrt{x_2^t K^{-t} K^{-1} x_2}} \quad (4.8)$$

- **P étant donnée, on calcule les intrinsèques**

Notation :

$r_i = [r_{i1}\ r_{i2}\ r_{i3}]$.

$$K \times [R\ T] = \begin{pmatrix} \alpha_u r_1 + u_0 r_3 & \alpha_u t_x + u_0 t_z \\ \alpha_v r_2 + v_0 r_3 & \alpha_v t_y + v_0 t_z \\ r_3 & t_z \end{pmatrix} \quad (4.9)$$

P est défini à un coefficient multiplicatif prés :

$$P = kK \times [R\ T]$$

Soit

$$\begin{pmatrix} l_1 & l_{14} \\ l_2 & l_{24} \\ l_3 & l_{34} \end{pmatrix}$$

D'où

$$\begin{pmatrix} l_1 = k(\alpha_u r_1 + u_0 r_3) & l_{14} = k(\alpha_u t_x + u_0 t_z) \\ l_2 = k(\alpha_v r_2 + v_0 r_3) & l_{24} = k(\alpha_v t_y + v_0 t_z) \\ l_3 = k r_3 & l_{34} = k t_z \end{pmatrix}$$

D'où $|k| = ||l_3||$

L'objet observé est devant la caméra → $t_z > 0$, le signe de k est déduit de l'égalité

$l_{34} = kt_z$

$$t_z = l_{34}/k \qquad r_3 = l_3/k$$
$$u_0 = \frac{l_1 . l_3}{k^2} \qquad v_0 = \frac{l_2 . l_3}{k^2}$$
$$\alpha_u = \sqrt{\frac{l_1 . l_1}{k^2} - u_0^2} \qquad \alpha_v = \sqrt{\frac{l_2 . l_2}{k^2} - v_0^2}$$
$$t_x = \frac{l_{14} - u_0 l_{34}}{k\alpha_u} \qquad t_y = \frac{l_{24} - v_0 l_{34}}{k\alpha_v}$$
$$r_1 = \frac{l_1 - u_0 l_3}{k\alpha_u} \qquad r_2 = \frac{l_2 - v_0 l_3}{k\alpha_v}$$

- **Calcul de P**

Les différents choix :

- Calcul de P à partir d'un objet 3D dont les points sont connus précisément et sont facilement identifiables dans les images (ex: Toscani & Faugeras)
- Calcul de P avec un objet plan dont la géométrie est connue. Les paramètres sont calculés à partir de l'observation de plusieurs vues de l'objet sans besoin de connaitre leur position (Zhang). Facilité de mise en œuvre!
- Calibrage à partir de points de fuite
- Calibrage à partir de rotations de la caméra

4.2 Etude de la localisation indoor

La vidéo et les dispositifs recevant des images d'une scène permettent d'effectuer 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. 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. 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, différentier 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 conclue 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.

4.2.1 Acquisition des images

Dans ses 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 et l'apprentissage profond, pour cette raison nous avons pensé à réaliser une acquisition avec deux Smartphones.

4.2.1.1 Acquisition par webcam

Une webcam est une caméra vidéoqui 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.

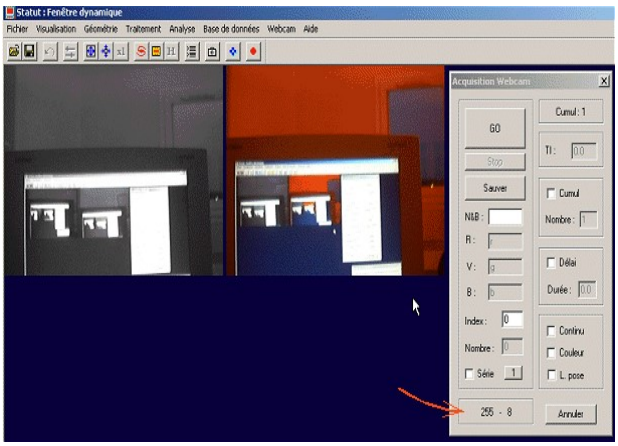


Figure 4. 5: Acquisition par webcam

4.2.1.2Acquisition 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. Certaines caméras IP 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.

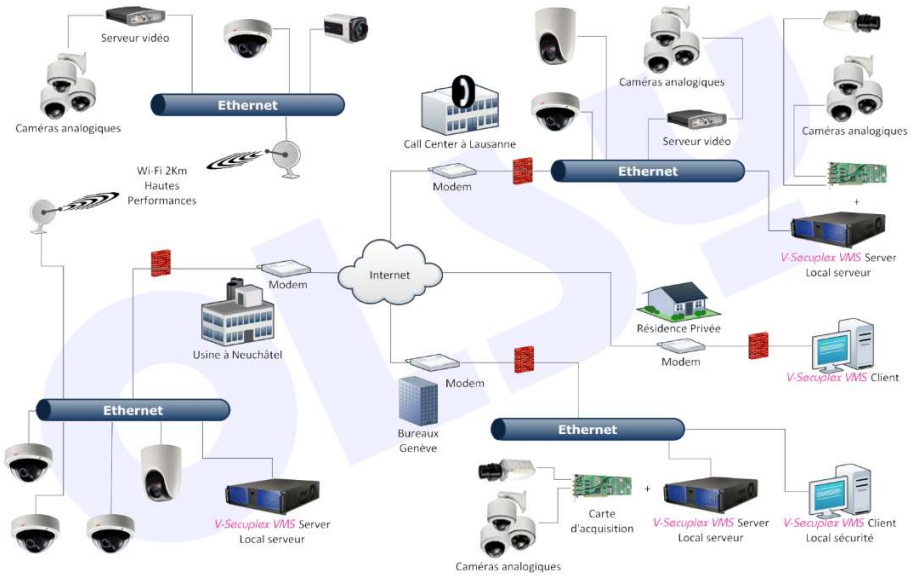


Figure 4. 6: Figure 4.2 Système IP [42]

Application utilisée de la caméra IP

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.

Figure 4. 7: Logo d'application

IP Webcam permet d'enregistrer des vidéos en Webcam, 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.

4.3 Utilisation de la détection Deep Learning (Faster R-CNN)

L'apprentissage en profondeur devient omniprésent. Grâce aux progrès récents des algorithmes d'apprentissage en profondeur et de la technologie GPU, nous sommes en mesure de résoudre des problèmes qui étaient considérés comme impossibles dans des domaines tels que la vision par ordinateur, le traitement du langage naturel et la robotique.

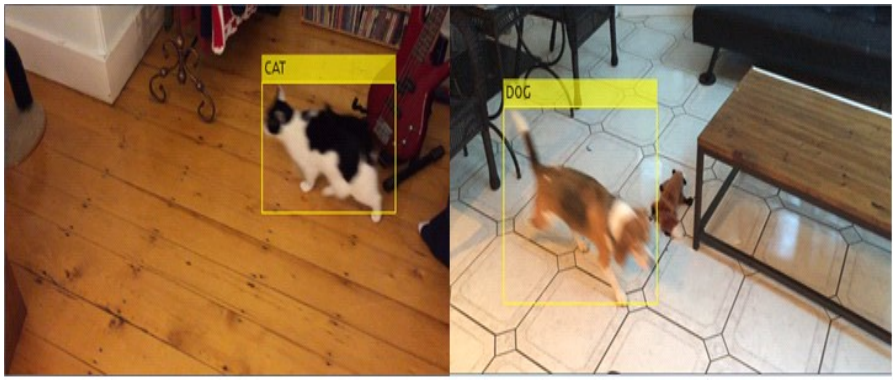


Figure 4. 8: Système de détection et de reconnaissance des animaux domestiques [43]

Comme nous l'avons déjà cité l'apprentissage en profondeur utilise des réseaux de neurones profonds qui existent depuis quelques décennies. Ce qui a changé ces dernières années, c'est la disponibilité d'importants jeux de données étiquetés et de puissants GPU. Les réseaux de neurones sont intrinsèquement des algorithmes parallèles et les GPU dotés de milliers de cœurs peuvent tirer parti de ce parallélisme pour réduire considérablement le temps de calcul nécessaire à la formation de réseaux d'apprentissage en profondeur. Dans ce chapitre, nous présenterons la manière de modéliser notre système afin de l'implémenter sur MATLAB pour développer un système de localisation et détection de

visage utilisant des réseaux de neurones à convolution profonde et des GPU. Ce qui fera l'objet du chapitre suivant.

4.4 Principe du Faster R-CNN

Dans notre travail nous utilisons le Faster R-CNN pour la détection de visage de personne dans une séquence. L'architecture de ce modèle est présentée par la **figure 4.4.**

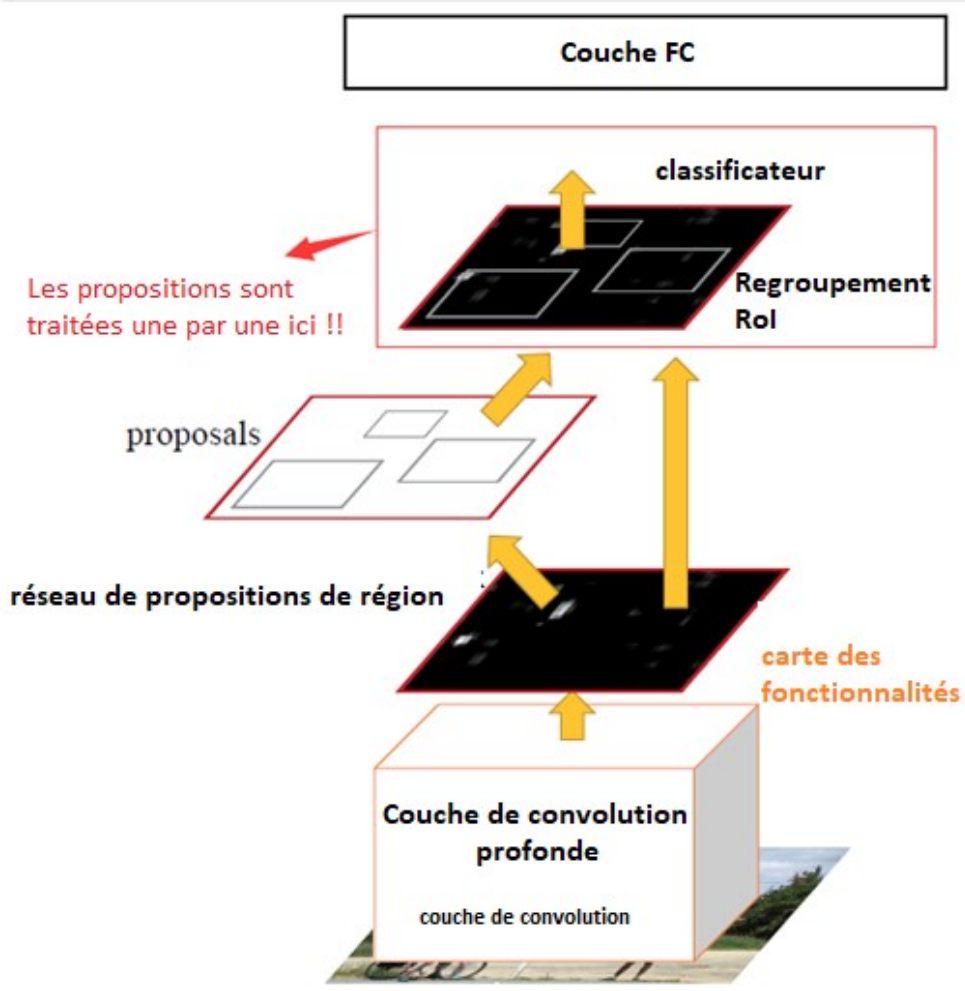


Figure 4. 9: Faster R-CNN modèle

L'idée principale est d'utiliser les dernières couches de conv (ou profondes) pour déduire des propositions de région. Faster-R-CNN est composé de deux modules :

1. *RPN (propositions de région):*qui donne un ensemble de rectangles basés sur une couche de convolution profonde

2. *Couche de regroupement Fast-RCNN RoI :* pour classifier chaque proposition et affiner son emplacement.

Nous décrivons le fonctionnement de Faster R-CNN comme ci-dessous, selon les six étapes suivantes :

1. Obtenir un réseau de neurones de convolution formés (Image-Net)
2. Obtenir des cartes de caractéristiques de la dernière couche (ou profonde) de convolution
3. Former un réseau de propositions de région qui décidera s'il y a un objet ou non sur l'image, et proposera également un emplacement de boîte
4. Donner des résultats à une couche personnalisée
5. Donner des propositions à une couche de regroupement de ROI (comme Fast RCNN)
6. Une fois que toutes les propositions sont remodelées à une taille de correctif, envoyez-les à une couche entièrement connectée pour poursuivre la classification

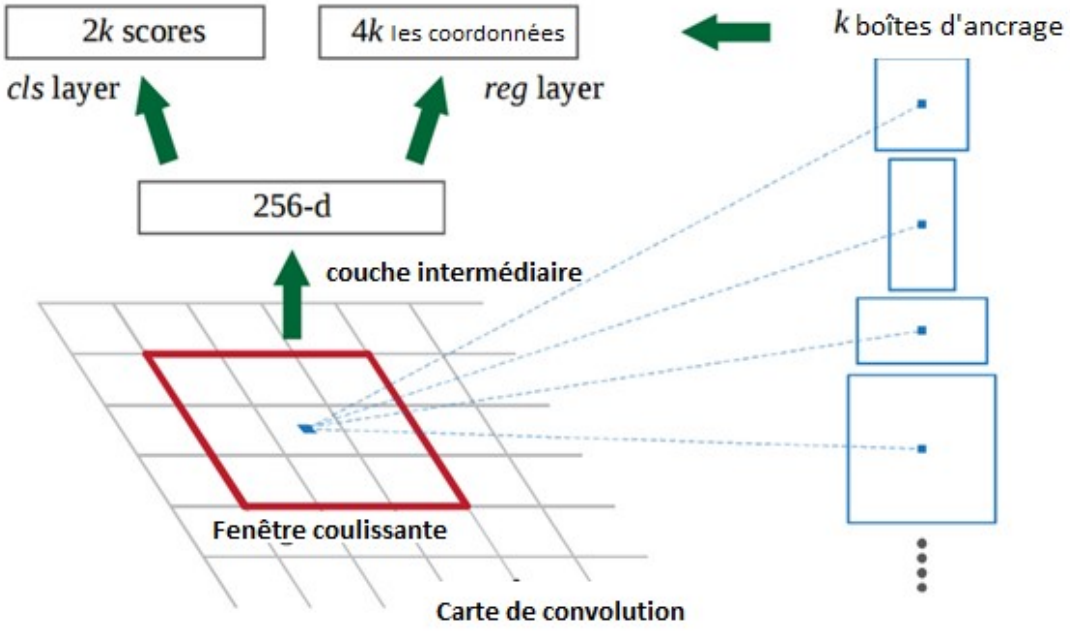


Figure 4. 10: Carte de la convolution

4.4.1Fonctionnement de RPN

Fondamentalement, le RPN glisse une petite fenêtre (3x3) sur la carte des fonctionnalités, qui classifie ce qui se trouve sous la fenêtre en tant qu'objet ou non, et donne également un emplacement du cadre de sélection. Pour chaque centre de fenêtre glissante, il crée k boîtes d'ancrage fixes et classe ces boîtes en tant qu'objet ou non.

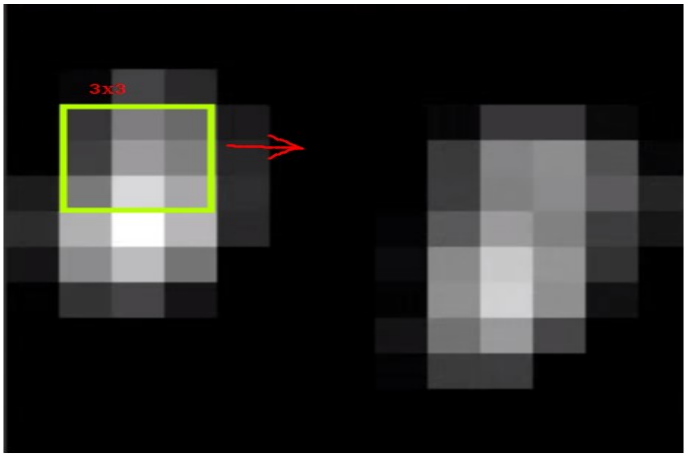


Figure 4. 11 : Recherche des fenêtres pour classifier

4.4.2 Entrainement de Faster R-CNN

Chaque réseau est formé séparément, mais nous pouvons également le former conjointement. Il suffit de considérer le modèle suivant :

1. Classification RPN (objet ou non objet)
2. Proposition de boîte englobant RPN
3. Classification Faster R-CNN (Classification des objets normaux)
4. Régression du cadre de sélection du Faster R-CNN (amélioration de la proposition)

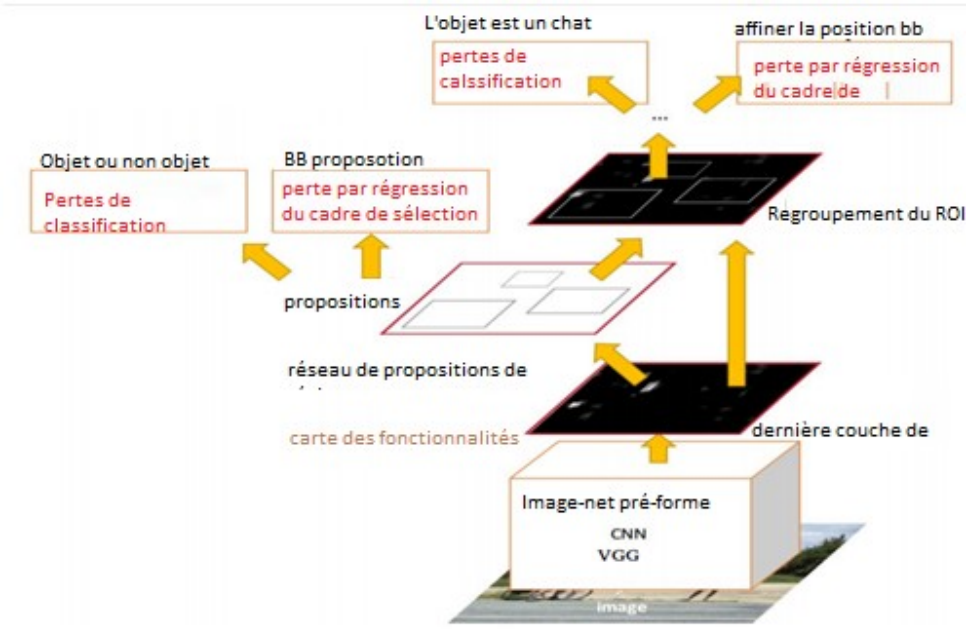


Figure 4. 12: Etapes d'entrainement de RCNN

3.4.4Résultat de Faster RCNN

Le meilleur résultat est maintenant Faster RCNN avec une couche ResNets 101.

	R-CNN	Fast R-CNN	Faster R-CNN
Test time per image (with proposals)	50 seconds	2 seconds	**0.2 seconds**
(Speedup)	1x	25x	**250x**
mAP (VOC 2007)	66.0	**66.9**	**66.9**

Figure 4. 13: Comparaison entre RCNN, Fast RCNN, Faster RCNN

Conclusion

Dans ce chapitre nous avons présenté une littérature cohérente pour la réalisation du système de localisation et les algorithmes de détection en profondeur d'objets qui appliquent en partie ou principalement les méthodes Deep Learning. Ces algorithmes ont été divisés en deux groupes de base, la détection d'objet et la localisation d'objet. Les algorithmes de détection d'objets se composent de plusieurs étapes. Deux étapes communes dans chacune d'elles ont d'abord été de localiser un objet, puis de le classer en utilisant un classificateur, par exemple. CNN. Cependant, le problème avec ces méthodes était le coût de calcul causé par la méthode utilisée pour produire des propositions de région et le fait que les algorithmes avaient plusieurs étapes pour traiter une image. En d'autres termes, toutes ces méthodes étaient des classificateurs redéfinis pour la détection d'objets. Cette étude nous permettra l'implémentation de notre modèle de localisation et détection en profondeur qui fera l'objet du chapitre 4.

Chapitre 5 :

Implémentation de la Localisation et Deep Détection

Introduction

Après des décennies de recherche, il n'existe toujours pas de produits omniprésents pour la localisation intérieure alors que la demande de services basés sur la localisation intérieure augmente rapidement dans les villes intelligentes. Les dernières années ont vu beaucoup de travaux sur la localisation intérieure. La plupart d'entre eux essaient de fournir un schéma largement utilisé pour la localisation en intérieur et d'obtenir des performances satisfaisantes comme le GPS dans les environnements extérieurs.

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 de 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. Dans ce chapitre nous présenterons l'implémentation du modèle de localisation indoor conçu. Ce modèle se base sur étape de localisation puis une détection en profondeur (Deep Détection). Pour assurer une bonne localisation, nous devons calibrer nos caméras.

5.1 Calibration de caméra

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. On peut 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.

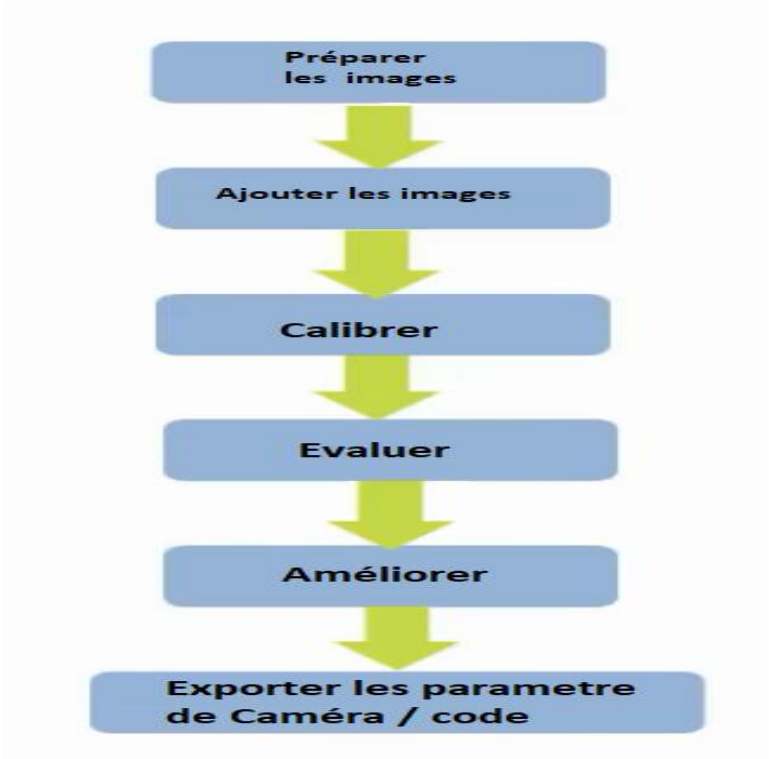


Figure 5.1: Etapes de calibration (étalonnage)

5.1.1 Installation et mise en œuvre de l'acquisition par Smartphones

Nous avons utilisé l'application Stéréo Camera Calibrator pour calibrer une caméra stéréo, que l'on a ensuite utilisé pour récupérer la profondeur des images. Un système stéréo se compose de deux caméras et un PC qui contient Matlab. 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.

Pour l'acquisition nous utilisons deux smart phones qui sont présentés sur la **figure 5.2**.

Figure 5. 2 : Photo des Smart Phones utilisés

Toutes les expériences de calibrage et localisation indoor sont effectuées sur un PC personnel dont les propriétés sont affichées sur l'écran se trouvant dans la **figure 5.3**.

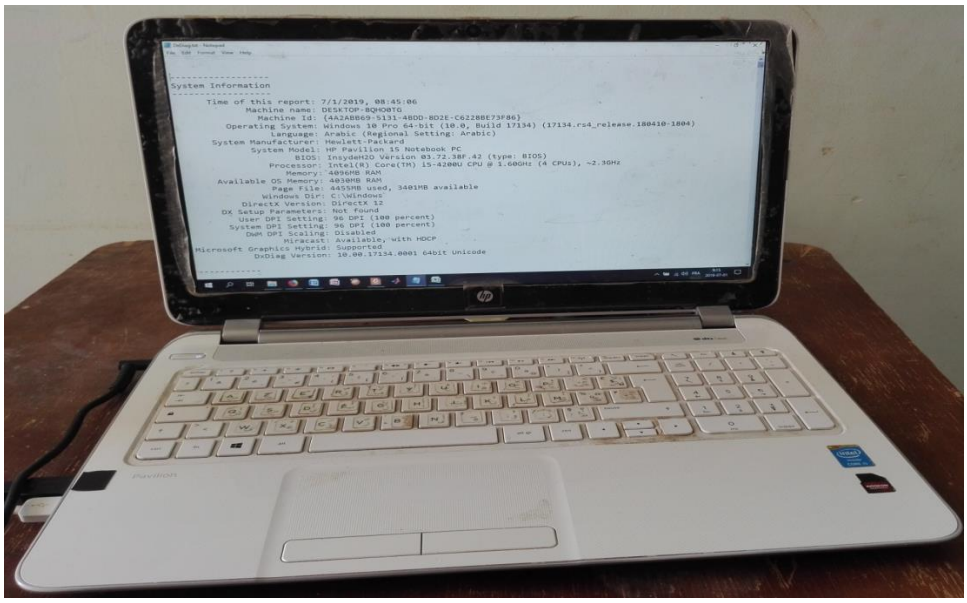


Figure 5. 3 :PC utilisé pour nos expériences

5.1.2Acquisition des images de calibration

Après la connexion des Smartphones (caméras) et le PC dans le même réseau nous prenons des photos avec différentes positions de damier comme le montre les images de la **figure 5.4** ci-dessous :

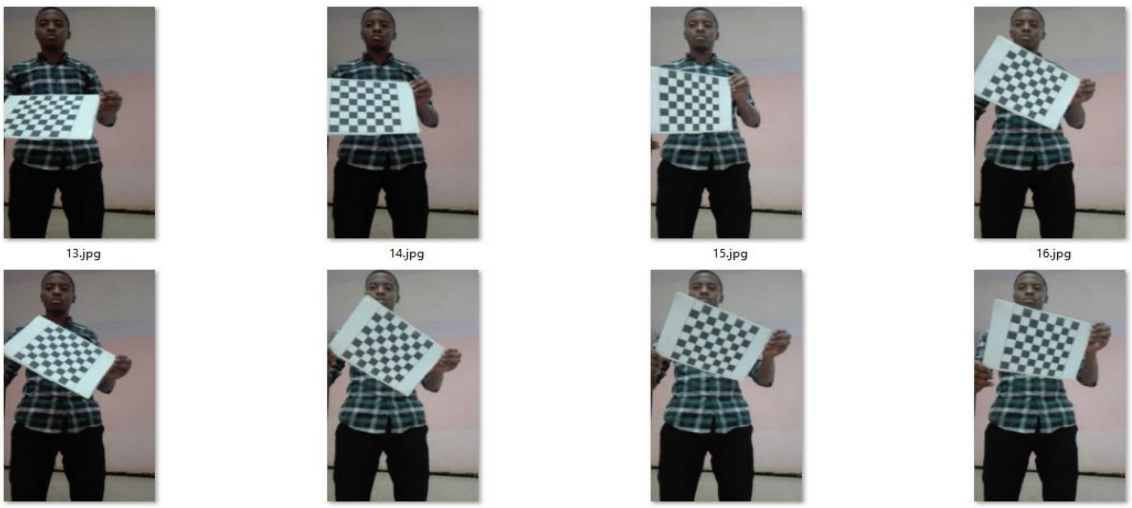


Figure 5. 4: Photos prises par la Caméra 1

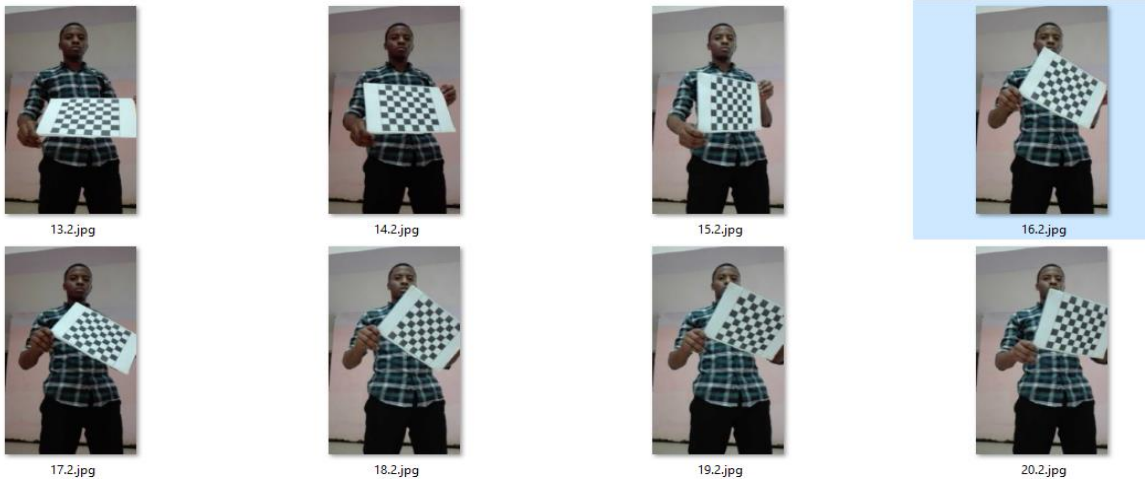


Figure 5.5 : Photos prises parla Caméra 2

- Ajouter les images au programme Matlab (Stereo Image Calibrator) : Après l'étape d'acquisition, nous procédons à l'ajout des images au programme Matlab (Stereo Image Calibrator) afin de réaliser la calibration des caméras. Cette opération est expliquée dans la figure 5.7

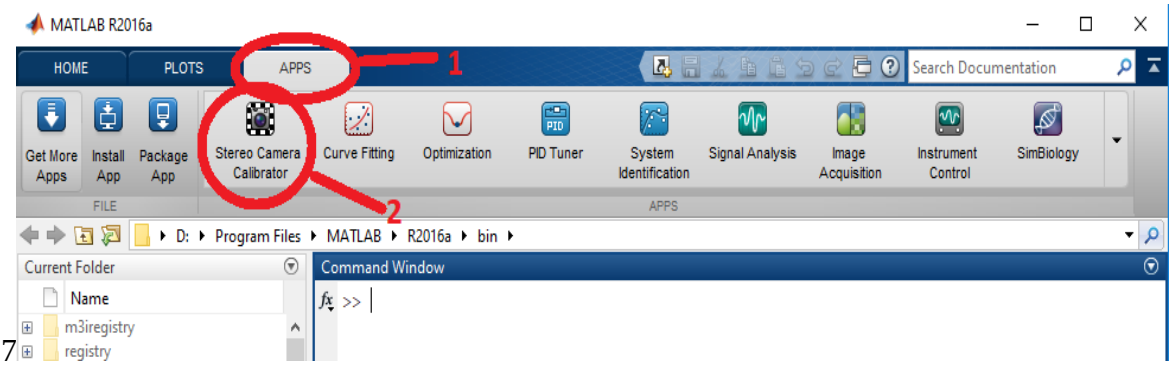


Figure 5.6: Ouverture de Stereo Image Calibrator

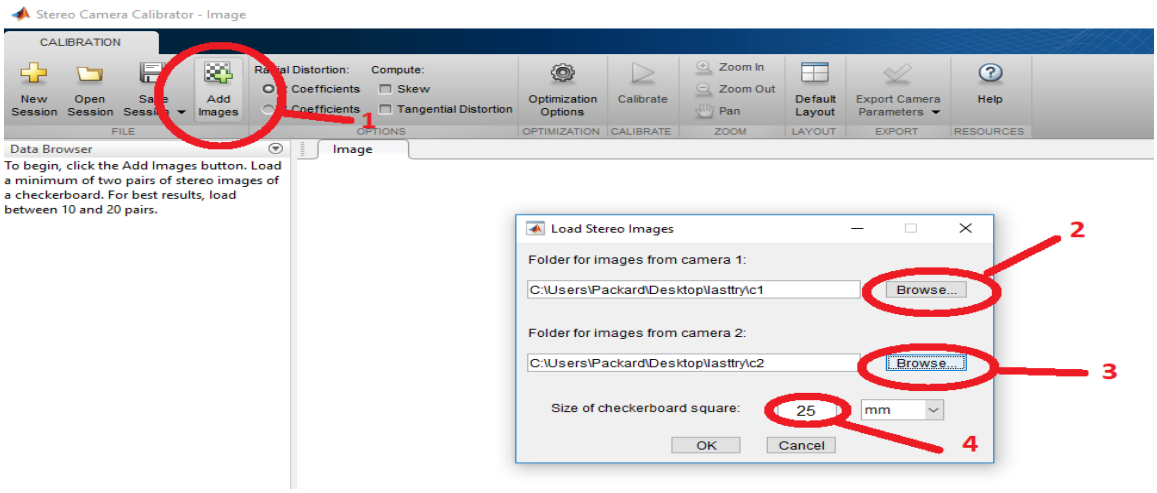


Figure 5.7: Opération d'Ajout des images

Pour bien garantir la calibration, nous devons suivre les étapes suivantes :

(1) Ajouter les images ;

(2) Entrer la direction de dossier des images capturées par caméra 1.

(3) Entrer la direction de dossier des images capturées par caméra 2.

(4) Entrer la taille de carrée.

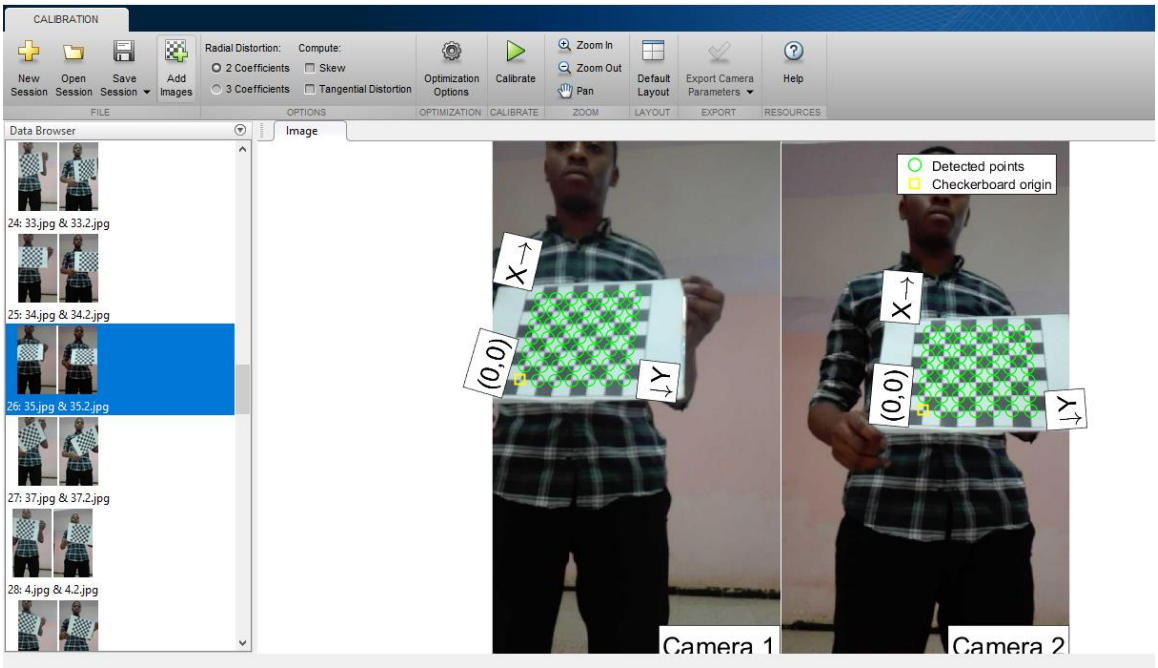


Figure 5.8: Importation des images par le programme

5.1.3 Calibrationdes images

Rappelons la définition du terme calibration ; ça correspond au fait d'étalonner, c'est-à-dire de confronter des données obtenues par des capteurs différents afin d'en tirer une information. Ce terme est un anglicisme souvent employé dans le monde scientifique.Tous les instruments ont tendance à dériver avec le temps et perdent leur précision d'étalonnage de façon périodique. L'ajustement assure que la précision reste au niveau requis et respecte des normes et un système de qualité.

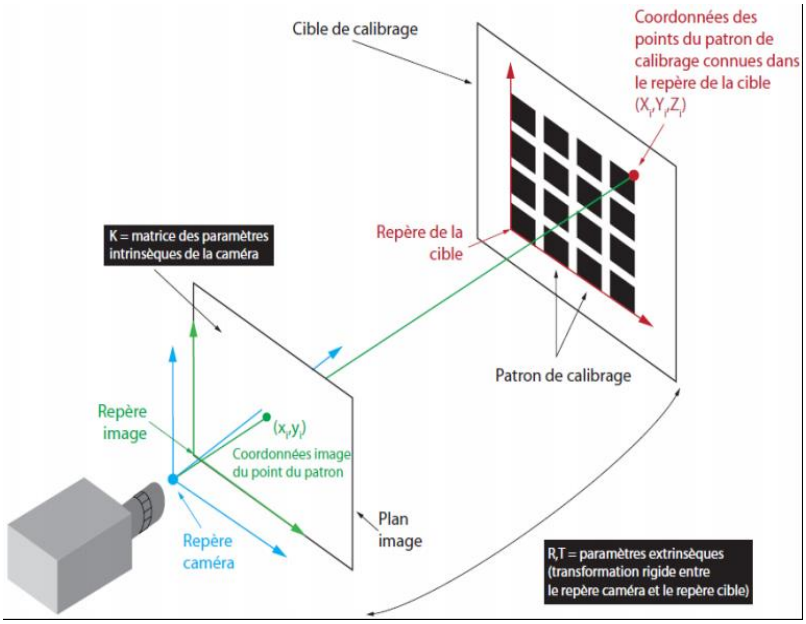


Figure 5.9: Principe générale de la calibration

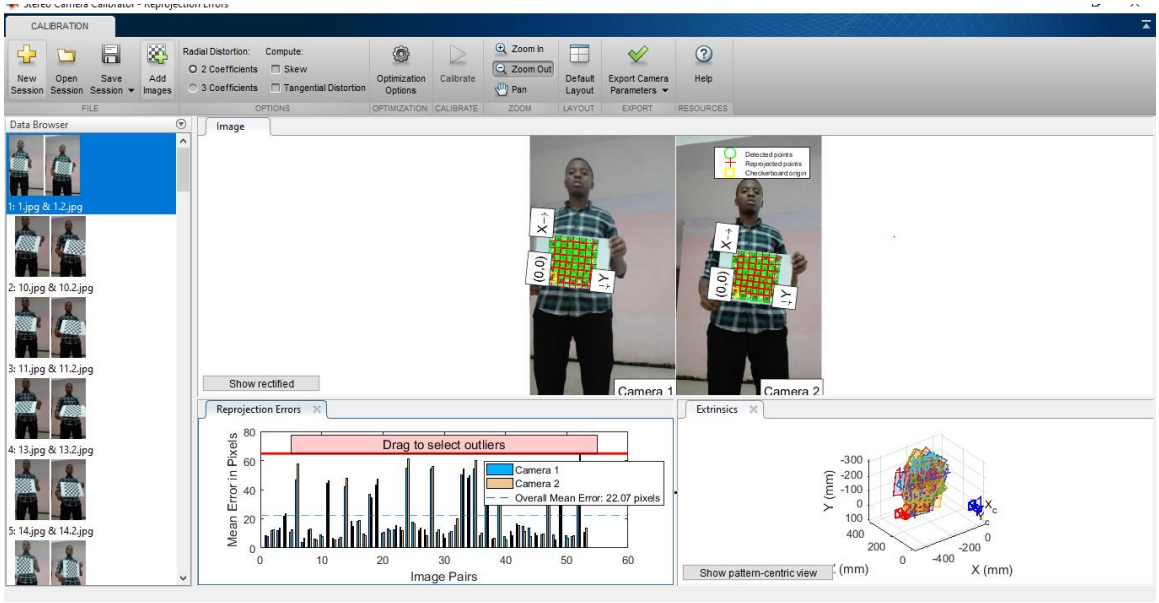


Figure 5.10 :Images calibrées

On peut minimiser les erreurs par glisser la ligne rouge dans la fenêtre de (Reprojectionerror) puis éliminer les images avec grands erreurs

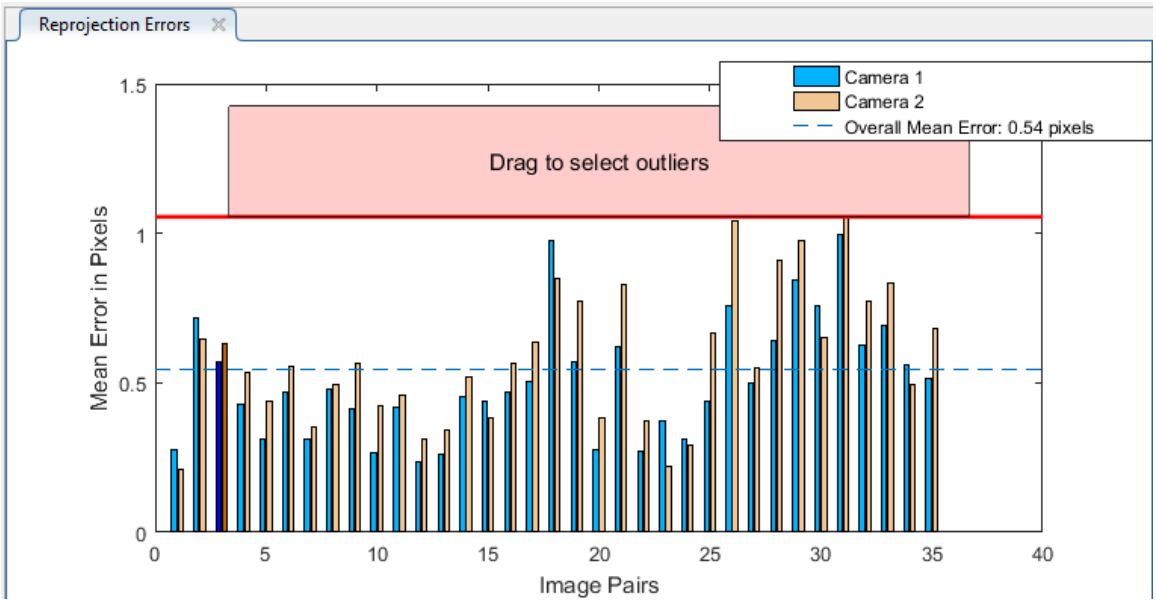


Figure 5. 11 : Evaluation des résultats de calibration

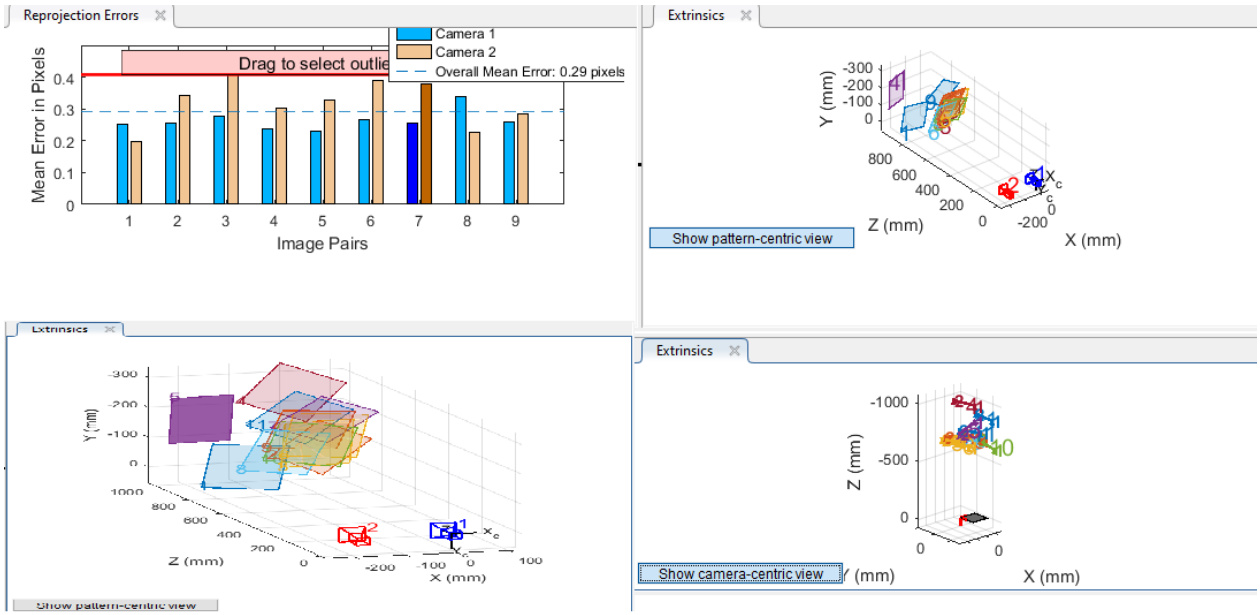


Figure 5.12: Résultats finaux

Les **figures 5.12 et 5.13** nous montrent l'opération de calibration des images. Cette opération reste fondamentale pour la localisation. L'erreur moyenne obtenue après calibration est égale à **0.3 pixels.**

- Exporterles paramètres de caméra et générer le script

1. **Paramètres intrinsèques**

La projection des coordonnées d'un point de l'espace sur l'image selon le modèle de sténopé est illustrée ci-dessous.

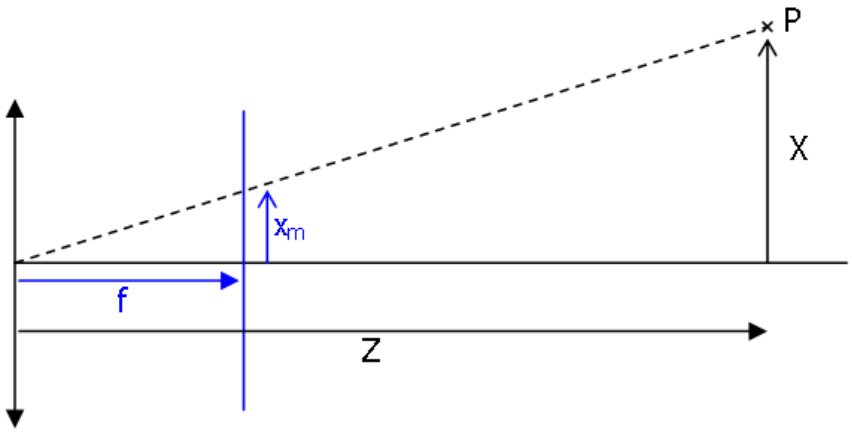


Figure 5.13: Projection centrale

Ainsi, les coordonnées (x_m, y_m, z_m) du point image de P dans le repère caméra sont :

$$x_m = f\frac{X}{Z} = fx$$

$$y_m = f\frac{Y}{Z} = fy$$

$$z_m = f$$

En informatique, on exprime en général les coordonnées des points d'une image à partir du coin supérieur gauche. On a donc les coordonnées du centre optique sur l'image comme paramètres supplémentaires à déterminer.Les relations s'écrivent matriciellement en faisant intervenir la matrice K (ou *M*), classiquement employée dans la littérature, dite de calibration interne :

$$K = \begin{pmatrix} f & 0 & i_1 \\ 0 & f & i_2 \\ 0 & 0 & 1 \end{pmatrix}.$$

2. Paramètres extrinsèques

Il s'agit simplement d'un vecteur de position de la caméra dans le repère de la scène et de trois vecteurs définissant l'orientation de celle-ci. Avec ces paramètres, on est alors capable de positionner les points de vue de l'objet les uns par rapport aux autres.Il est fréquent que les positions et orientations des caméras soient données relativement à la première caméra.

On peut faire l'estimation des paramètres intrinsèques et extrinsèques à partir de la matrice caméra$M$. Dans certaines applications, la connaissance de la matrice M est suffisante. Dans d'autres cas, on peut être intéressé à obtenir les paramètres intrinsèques et extrinsèques du modèle sténopé. Ces paramètres intrinsèques et extrinsèques peuvent être obtenus par plusieurs étapes de calcul sur la matrice M. Après implémentation du programme de calibration, nous pouvons afficher les paramètres de la caméra, ce qui est montré par la **figure 5.14.**

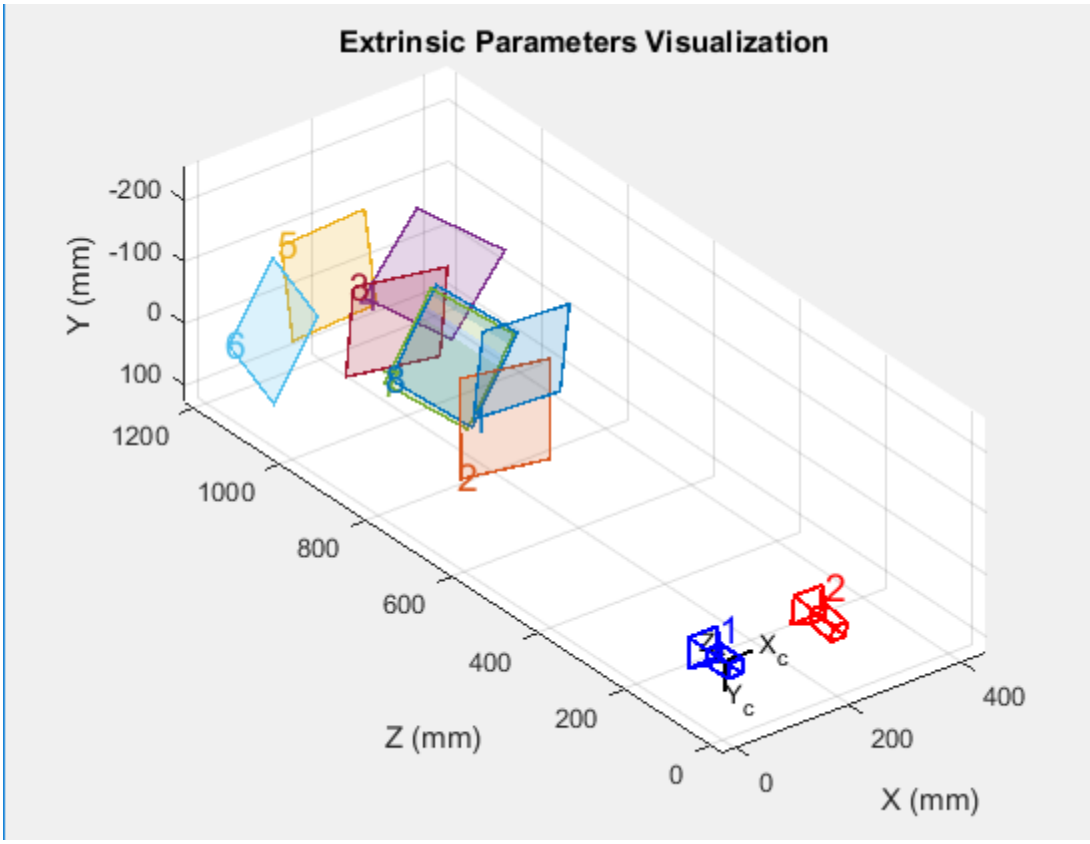


Figure 5.14: Visualisation de paramètres extrinsèques

D'après la **figure 5.14**, les paramètres extrinsèques visualisés montrent les captures de damier et ses positions par rapport les caméras.

Les erreurs moyennes de reprojection obtenues pour les paires d'images sont illustrées par la **figure 5.15.**

Figure 5.15 Erreur de reprojection moyenne pour les paires d'images

- Sauvegarder StereoParams.

```
stereoParams =

  stereoParameters with properties:

    Parameters of Two Cameras
         CameraParameters1: [1x1 cameraParameters]
         CameraParameters2: [1x1 cameraParameters]

    Inter-camera Geometry
         RotationOfCamera2: [3x3 double]
      TranslationOfCamera2: [458.8640 48.1803 -129.3291]
         FundamentalMatrix: [3x3 double]
           EssentialMatrix: [3x3 double]

    Accuracy of Estimation
     MeanReprojectionError: 22.0674

    Calibration Settings
                NumPatterns: 53
                WorldPoints: [49x2 double]
                 WorldUnits: 'mm'


estimationErrors =
```

5.2Localisation indoor

Après la calibration des caméras, nous mettons en œuvre notre modèle de localisation. Nous avons implémenté ce modèle de localisation indoor (lumière artificielle) et nous obtenons les résultats suivants montrés par les **figures 5.16, 5.17, 5.18et 5.19.**

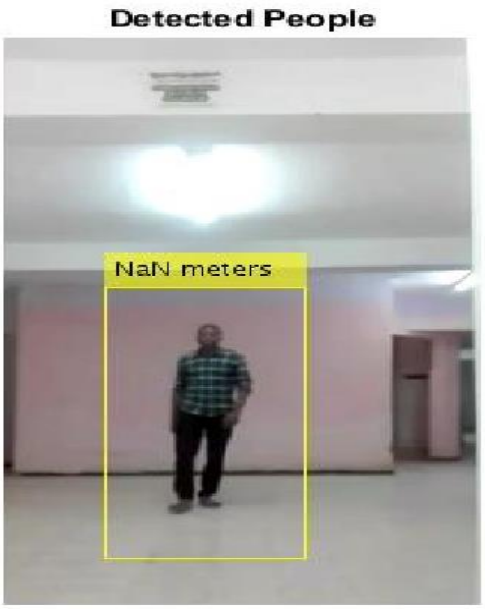
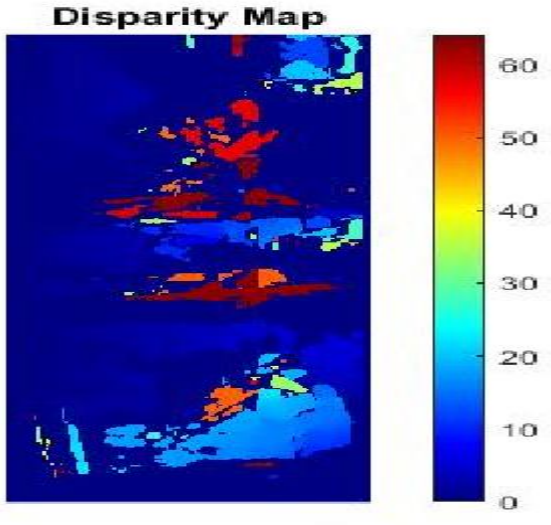


Figure 5.16Détection d'une personne**Figure 5.17**Rectification des vidéo-reconstruction 3D



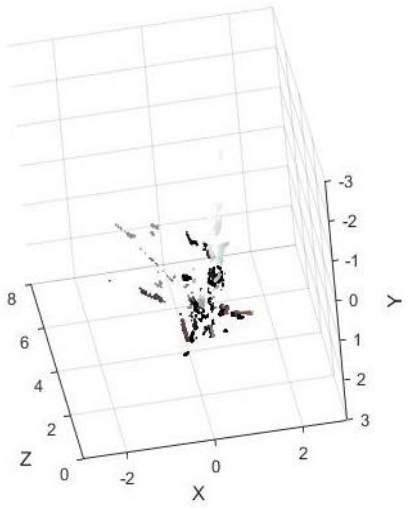


Figure 5.18 Carte de disparité**Figure 5.19**Reconstruction de la scène 3D

La **figure 5.18** représente la **carte de disparité** qui correspondant à la scène de la **figure 5.17** désignant ainsi une image numérique qui contient uniquement l'information sur les correspondances des points entre deux vues de cette même scène prises à deux endroits différents.

Dans la **figure 5.19**, les points visibles sur les images sont les projections des points réels qu'on peut alors situer sur des droites. Si deux ou plusieurs vues de l'objet sont prises, la position dans l'espace des points réels peut alors être obtenue par intersection de ces droites : c'est le principe de ***triangulation*** fondamental de toute reconstruction 3D à partir d'images.

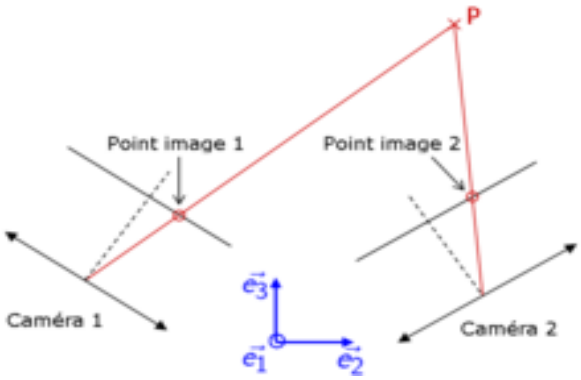


Figure 5. 20 Principe de triangulation

Pour obtenir, comme souhaité, les coordonnées des points de la scène, il faut cependant résoudre un certain nombre de problèmes :

- Problème de *calibration* ou comment se projettent les points de la scène sur l'image. Pour cela, le modèle de sténopé est utilisé et il est alors nécessaire de connaître des paramètres dits *intrinsèques* de la caméra (*distance focale*, centre de l'image...). Ensuite, il est nécessaire de connaître la position relative des caméras pour pouvoir déterminer les coordonnées des points de l'espace dans un repère de la scène non lié à la caméra. Ces paramètres, dits *extrinsèques*, sont la position et l'orientation de la caméra dans l'espace.
- Problème d'association (*matching*): il faut être capable de reconnaître et d'associer les points qui apparaissent sur plusieurs photos.
- Problème de reconstruction : il s'agit de déterminer les coordonnées 3D des points à partir des associations faites et des paramètres de calibration.

À cela s'ajoute un autre problème : la densité de la reconstruction. Une fois obtenues les coordonnées d'un certain nombre de points dans l'espace, il faut trouver la surface à laquelle appartiennent ces points pour obtenir un *maillage*, un modèle dense. Sinon, dans certains cas, quand on obtient un grand nombre de points, le nuage de points(voir **figure 5.20**) formé suffit à définir visuellement la forme de l'objet mais la reconstruction est alors clairsemée (*sparse*).

5.3Détection par Deep Learning

Pour la détection nous créons une base de données en utilisant les images à partir de la séquence localisée comme le montre la **figure 5.22,** ensuite nous lançons la détection basée sur le Faster RCNN. Les résultats de la localisation avant et après la détection se trouvent dans les figures **5.23, 5.24 et 5.25**.

Notre modèle de localisation et détection est implémenté et ce processus

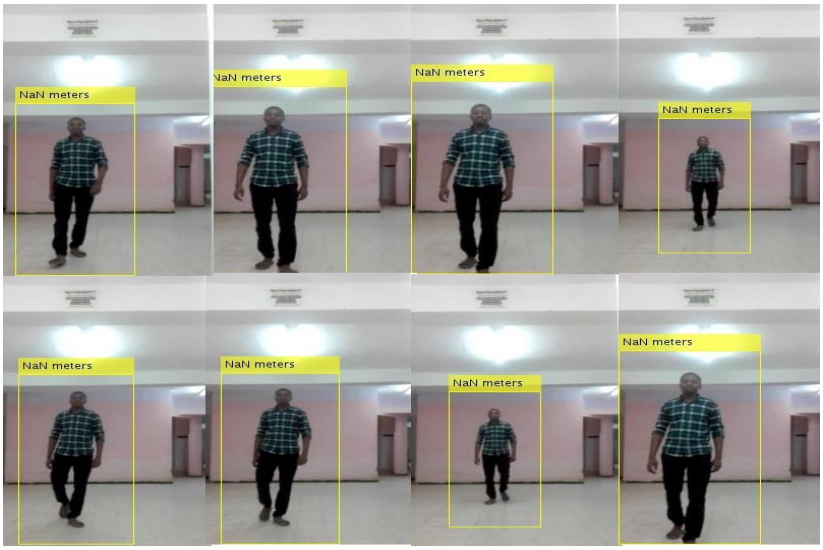


Figure 5. 21: Création de database

Figure 5. 22: Exemple de localisation de frame

Architecture du modèle Faster RCNN pour la détection

```
>> fasterrcnn

layers =

  11x1 Layer array with layers:

     1   ''   Image Input             32x32x3 images with 'zerocenter' normalization
     2   ''   Convolution             32 3x3 convolutions with stride [1  1] and padding [1  1  1  1]
     3   ''   ReLU                    ReLU
     4   ''   Convolution             32 3x3 convolutions with stride [1  1] and padding [1  1  1  1]
     5   ''   ReLU                    ReLU
     6   ''   Max Pooling             3x3 max pooling with stride [2  2] and padding [0  0  0  0]
     7   ''   Fully Connected         64 fully connected layer
     8   ''   ReLU                    ReLU
     9   ''   Fully Connected         2 fully connected layer
    10   ''   Softmax                 softmax
    11   ''   Classification Output   crossentropyex
*************************************************************************
```

Résultats de l'implémentation de Faster R-CNN

Avant de procéder à la détection des images, le modèle Faster RCNN doit être entrainé pour fixer ses paramètres. Pour cela, nous réalisons plusieurs apprentissages en utilisant la fonction ***trainingOptions***.

```
options = trainingOptions('sgdm', ...
      'InitialLearnRate', 1e-3, ...
      'MaxEpochs', 5, ...
      'VerboseFrequency', 200, ...
      'CheckpointPath', tempdir);
```

L'apprentissage du Faster RCNN nécessite plusieurs étapes.

1. **Expérience 1 :** *Pour MaxEpochs = 5etMiniBatchSize = 200*
   a. ***1ère étape de l'apprentissage :*** Dans cette étape 5 époques et 135 itérations sont utilisées. Les résultats se trouvent ci-dessous, avec RMSE = 0.79%, Accuracy = 100% et un taux d'apprentissage = 0.001% avec un temps d'apprentissage = 03 mn 52 s.

```
Training a Faster R-CNN Object Detector for the following object classes:

* Face

Step 1 of 4: Training a Region Proposal Network (RPN).
Training on single CPU.
|=========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     RMSE     |      Rate       |
|=========================================================================================|
|       1 |           1 |       00:00:01 |       50.00% |         0.89 |          0.0010 |
|       5 |         135 |       00:03:52 |      100.00% |         0.79 |          0.0010 |
|=========================================================================================|
```

b. 2ème étape de l'apprentissage : Dans cette étape le nombre d'époques est fixé à 5. Dans ce deuxième apprentissage le nombre d'itérations diminue et est égale 125. Les résultats se trouvent ci-dessous, avec RMSE = 0.47%, Accuracy = 90.22% et un taux d'apprentissage = 0.001%avec un temps d'apprentissage = 01 mn 42s.

```
Step 2 of 4: Training a Fast R-CNN Network using the RPN from step 1.
********************************************************************
Training a Fast R-CNN Object Detector for the following object classes:

* Face

--> Extracting region proposals from 27 training images...done.

Training on single CPU.
|=========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     RMSE     |       Rate      |
|=========================================================================================|
|       1 |           1 |       00:00:00 |       73.17% |         0.95 |          0.0010 |
|       5 |         125 |       00:01:42 |       90.22% |         0.47 |          0.0010 |
|=========================================================================================|
```

c. 3ème étape de l'apprentissage : Dans cette étape le nombre d'époques est fixé à 5. Dans ce troisième apprentissage le nombre d'itérations reprend sa valeur initial135. Les résultats se trouvent ci-dessous, avec RMSE = 0.72%, Accuracy = 100% et un taux d'apprentissage = 0.001%avec un temps d'apprentissage = 02 mn 42s.

```
Step 3 of 4: Re-training RPN using weight sharing with Fast R-CNN.
Training on single CPU.
|=========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     RMSE     |       Rate      |
|=========================================================================================|
|       1 |           1 |       00:00:01 |      100.00% |         0.93 |          0.0010 |
|       5 |         135 |       00:02:42 |      100.00% |         0.72 |          0.0010 |
|=========================================================================================|
```

d. ***4ème étape de l'apprentissage :*** Dans cette dernière étape le nombre d'époques est fixé à 5. Le nombre d'itérations est égal à 130. Les résultats se trouvent ci-dessous, avec RMSE = 0.36%, Accuracy = 93% et un taux d'apprentissage = 0.001%avec un temps d'apprentissage = 26s.

```
Step 4 of 4: Re-training Fast R-CNN using updated RPN.
********************************************************************
Training a Fast R-CNN Object Detector for the following object classes:

* Face

--> Extracting region proposals from 27 training images...load('Face.mat', 'T')
done.

Training on single CPU.
|=========================================================================================|
|  Epoch  |  Iteration  |  Time Elapsed  |  Mini-batch  |  Mini-batch  |  Base Learning  |
|         |             |   (hh:mm:ss)   |   Accuracy   |     RMSE     |       Rate      |
|=========================================================================================|
|       1 |           1 |       00:00:00 |       93.15% |         0.93 |          0.0010 |
|       5 |         130 |       00:00:26 |       96.61% |         0.36 |          0.0010 |
|=========================================================================================|

Finished training Faster R-CNN object detector.
```

***La meilleure performance d'apprentissage est obtenue pour la 4ème étape d'apprentissage avec un RMSE*=0.36% et un temps d'apprentissage = 26s.**

Donc, cette structure est retenue pour l'implémentation de la détection par Faster RCNN. Dans les expériences suivantes, nous gardons le même mode d'apprentissage et nous faisons varier les paramètres : *MaxEpochs* et le*MiniBatchSize*.

Résultats de localisation et détection

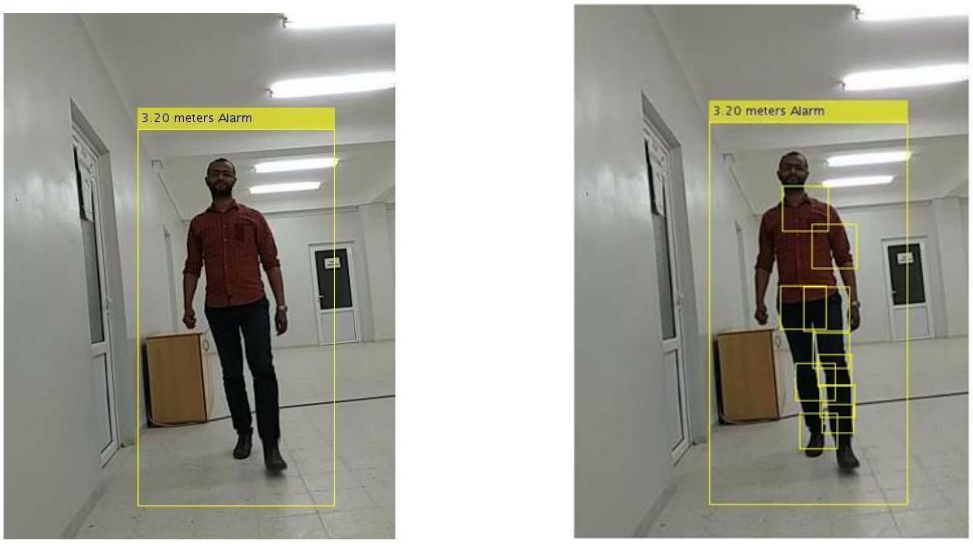


Figure 5. 23: Image localisée avant et après la détection Faster RCNN *(MaxEpochs = 5 et MiniBatchSize = 200)*

La figure **5.23**montre bien que notre approche détecte les parties de l'objet localisé et qui est la personne en mouvement dans ce frame d'une séquence vidéo.

2. Expérience 2 : *Pour MaxEpochs = 3etMiniBatchSize = 400*

Dans cette expérience 2, nous obtenons les résultats de localisation et détection données par la **figure 5.24.**

Figure 5. 24: Image localisée avant et après la détection Faster RCNN (*MaxEpochs = 3 etMiniBatchSize = 400*)

3. Expérience 3 : *Pour MaxEpochs = 10etMiniBatchSize = 400*

Dans cette expérience 3, nous obtenons les résultats de localisation et détection données par la **figure 5.25.**

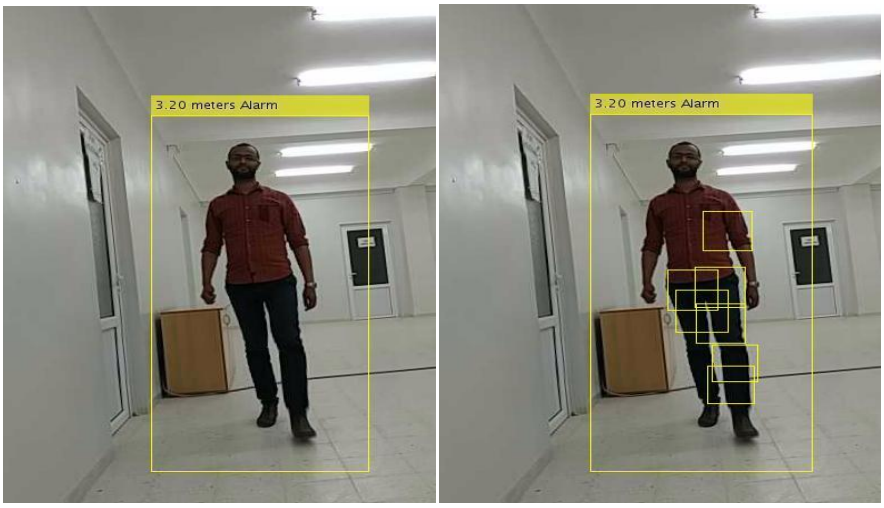


Figure 5. 25 : Image Localisée avant et après la détection Faster RCNN (*MaxEpochs = 10etMiniBatchSize = 400)*

Donc, les meilleurs sur le nombre de régions détectées est la première expérience pour (*MaxEpochs = 5etMiniBatchSize = 200)*

Conclusion

Les applications basées sur la localisation peuvent donc utiliser, dans un périmètre donné, une combinaison de solutions centralisées à large échelle, avec, potentiellement, d'autres services de localisationindépendants qui pourrait leur être composés dynamiquement.Une infrastructure logicielle locale de localisation telle quedécrite ci-dessus est une solution possible pour consolider cetteintégration au niveau le plus élevé, leur combinaisonaux niveaux inférieurs étant réalisable suivant le modèle.

Nous avons présenté les réseaux de proposition de région (Fast RCNN) pour une proposition de région efficace et précise.En partageant les fonctionnalités de convolution avec le réseau de détection. Notre méthode permet une détection d'objet unifiée, basée sur l'apprentissage en profondeur. Le Fast RCNN appris améliore également la qualité de la proposition de région et donc laprécision globale de détection d'objet.

Conclusion générale

Dans ce projet, un schéma de localisation en intérieur limité par des scènes a été proposé. Les résultats et les comparaisons ont montré que ce système présente des performances compétitives par rapport à la plupart des systèmes actuels. Nous avons résolu les problèmes de localisation indoor (à l'intérieur).

Ce projet présente des stratégies d'utilisation des fonctionnalités CNN pour la détection d'objets. Grâce au réglage plus précis de Faster R-CNN pour les images de requête qui incluent les mêmes objets à détecter, donc Faster R-CNN produit de meilleures représentations d'entités pour la détection d'objet. En concaténant et en normalisant une couche convolutionnelle moins profonde et une couche convolutionnelle plus profonde pour les RPN, Faster R-CNN permet de mieux identifier les images à basse résolution. Sans prendre en compte la consommation de temps, nous avons montré qu'il était possible d'améliorer considérablement les performances d'un système prêt à l'emploi. Dans les prochaines études, il faut étudier comment raccourcir le temps nécessaire à la mise au point pendant formation et comment améliorer encore l'efficacité de Faster R-CNN au cours des tests dans les suivantes études.

De plus, notre détecteur de personne sur les exemples de données est formé individuellement et avec succès en utilisant les méthodes rapides d'apprentissage en profondeur Faster RCNN.

Le détecteur de personne formé est testé sur les données de test et les résultats efficaces sont obtenus à partir de problème de détection de personne En outre, le succès le taux de détection du détecteur formé a été essayé de être maximisé autant que possible et expérimental des comparaisons d'analyse sont faites avec les résultats obtenu à partir des méthodes. La méthode proposée est détaillée ainsi que son implémentation et aussi les résultats expérimentaux.

En conclusion, nous avons présenté une procédure automatisée permettant de générer des données d'entraînement synthétiques pour les détecteurs d'objets CNN profonds. La procédure de génération prend en compte la géométrie et la segmentation sémantique de la scène afin de prendre des décisions éclairées concernant les positions et les échelles des objets. Nous avons utilisé deux détecteurs d'objets à la pointe de la technologie et démontré une augmentation de leurs performances lorsqu'ils sont formés avec un kit d'entraînement renforcé. En outre, nous avons également étudié l'effet de différents paramètres de génération et fourni quelques informations qui pourraient s'avérer utiles lors de futures tentatives de génération de données synthétiques pour les détecteurs d'objets d'entraînement.

Les applications basées sur la localisation peuvent donc utiliser, dans un périmètre donné, une combinaison de solutions centralisées à large échelle, 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 le plus élevé, leur combinaison aux niveaux inférieurs étant réalisable suivant le modèle.

Nous avons présenté les réseaux de proposition de région (Fast RCNN) pour une proposition de région efficace et précise. En partageant les fonctionnalités de convolution avec le réseau de détection. Notre méthode permet une détection d'objet unifiée, basée sur l'apprentissage en profondeur. Le Fast RCNN appris améliore également la qualité de la proposition de région et donc la précision globale de détection d'objet.

Fondé sur une étude de pointe de la localisation en intérieur, ce travail parmi les premiers à adopter, à notre connaissance, un apprentissage en profondeur pour extraire des informations sémantiques de haut niveau destinées à la localisation en intérieur. Cette méthode est similaire au mode cognitif au cerveau humain et présente un grand potentiel pour la recherche future.

References

[1] Frédéric E. Techniques et technologies de localisation avancées pour terminaux mobiles dans les environnements indoor. Mémoire de doctorat : Optique et Radio Fréquences : UNIVERSITE JOSEPH FOURIER - GRENOBLE.

[2] https://fr.wikipedia.org/wiki/G%C3%A9olocalisation

[3] https://eduscol.education.fr/orbito/system/navstar/gps1.htm

[4] LUBAC Bertrand. LOCALISATION PAR SATELLITES: LE SYSTEME GPS. Maître de conférences : Université Bordeaux

[5] https://fr.wikipedia.org/wiki/Gateway_GPRS_Support_Node

[6] Vectronix. WebSite, www.vectronix.ch

[7]http://mecaspa.cannesaeropatrimoine.net/MECANIQ/Techniques_inertielles/CENTRALE_INERTIELLE/CENTRALE_I ERTIELLE.htm

[8]https://www.google.com/search?biw=1600&bih=740&tbm=isch&sa=1&ei=ehoJXc3dNIueUKLggYAM&q=localisation+par+ultrason&oq=localisation+par+ultrason&gs_l=img.1.0.0i24.310355.338364..340277...0.0..0.146.146.0j1.....1....2j1..gws-wiz-img.crUtG9pd0do#imgrc=Nt6xKkHJQZ-GoM:

[9] N. B. Priyanta. The Cricket location-support system. *MOBICOM 2000*, August 2000

[10] AT&T Laboratories Cambridge. The Active Badge System. WebSite, http://www.uk. research.att.com/ab.html, 2002

[11] 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

[12] J. Hightower, G. Boriello, and R. Want. SpotOn : An indoor 3D location sensing techno logy based on RF Signal Strength. Technical Report UW-CSE 2000-02-02, University of Washington, February 2000

[13] Aeroscout, Entreprise Visibility Solutions. Website, http://www.aeroscout.com/

[14] André Günther and Christian Hoene. Measuring Round Trip Times to Determine the Distance between WLAN Nodes. TKN Technical Report TKN-04-16, Technical University Berlin - Telecommunication Networks Group, 2004.

[15] Rafiq Sekkal. Techniques visuelles pour la détection et le suivi d'objets 2D, Traitement du signal et de l'image. INSA de Rennes, 2014. Français

[16] https://twitter.com/khalidhamdan0/status/906692588220157952

[17] https://sergioskar.github.io/Localization_and_Object_Detection/

[18] https://www.linuxembedded.fr/2018/05/la-localisation-indoor-1/

[19] https://www.lesechos.fr/

[20] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. A brief survey of deep reinforcement learning. *arXiv preprint arXiv:1708.05866*, 2017

[21] Richard S Sutton, Andrew G Barto, et al. Reinforcement learning: An introduction. MIT press, 1998

[22] Mohammad Otoofi. Object localization using deep reinforcement learning, School of Computing Science Sir Alwyn Williams Building University of Glasgow

[23] Zequn Jie, Xiaodan Liang, Jiashi Feng, Xiaojie Jin, Wen Lu, and Shuicheng Yan. Treestructured reinforcement learning for sequential object localization. In Advances in Neural Information Processing Systems, pages 127–135, 2016

[24] Juan C Caicedo and Svetlana Lazebnik. Active object localization with deep reinforcement learning. In Proceedings of the IEEE International Conference on Computer Vision, pages 2488–2496, 2015.

[25] Miriam Bellver Buenoa, Xavier Giro-i Nietob, Ferran Marquesb, and Jordi Torresa. Hierarchical object detection with deep reinforcement learning. Deep Learning for Image Processing
Applications, 31:164, 2017.

[26] Xintao Ding et al, Indoor object recognition usingpre trainedconvolutional neural network, 23rd International Conference on Automation and Computing (ICAC)}, 2017, pages=1-6

[27] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 779–788, 2016.

[28] Barret Zoph and Quoc V Le. Neural architecture search with reinforcement learning. arXiv preprint arXiv:1611.01578, 2016.

[29] https://towardsdatascience.com/everything-you-need-to-know-about-automl-and-neural-architecture-search-8db1863682bf

[30] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015.

[31] https://towardsdatascience.com/deep-learning-for-object-detection-a-comprehensive-review-73930816d8d9?gi=60afdd2237bf

[32] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE international conference on computer vision, pages 1440–1448, 2015.

[33] Zhipeng Deng et al, Multi-scale object detection in remote sensing imagery with convolutional neural networks, ISPRS Journal of Photogrammetry and Remote Sensing, April 2018. DOI: 10.1016/j.isprsjprs.2018.04.003

[34] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, (6):1137–1149, 2017.

[35] https://towardsdatascience.com/faster-rcnn-object-detection-f865e5ed7fc4

[36] Kaiming He, Georgia Gkioxari, Piotr Dollar, and Ross Girshick. Mask r-cnn. In ´ Computer Vision (ICCV), 2017 IEEE International Conference on, pages 2980–2988. IEEE, 2017.

[37] https://medium.com/neuromation-blog/neuronuggets-segmentation-with-mask-r-cnn c76d363b67fb

[38] Jifeng Dai, Yi Li, Kaiming He, and Jian Sun. R-fcn: Object detection via region-based fully convolutional networks. In Advances in neural information processing systems, pages 379– 387, 2016.

[39] https://arxiv.org/pdf/1605.06409.pdf

[40] https://arxiv.org/pdf/1512.02325.pdf

[41] Marie-Odile Berger. Calibrage d'une caméra, 2014

[42] https://www.google.com/search?q=Ip+camera+acquisition&tbm=isch&tbs=rimg:CfJIQkjR8HZIjjE7l5GzBSI8GZlpOvPAheyncU0IIvpRuNt5YrX_1CWtRnPGaLkO3i2Ch3T4yMCx8Pq7jYwYXGyyoSCcTuXkbMFIjwEUmmjz4yfaKOKhIJZmWk688CF7IRKfUd4EwExY0qEgmdxTQgilG4xEPMWCfFnYybyoSCW3litf_1YJa1EV0h2NJycRa9KhIJGc8ZouQ7eLYR7lPTBaBfxNwqEgkKHdPjIwLHwxHmwVMK7d1ShioSCeruNjBj5cbLEalWt4fMilg&tbo=u&sa=X&ved=2ahUKEwiCsK_OvITjAhWG3OAKHf4CBbcQ9C96BAgBEBg&biw=1680&bih=890&dpr=1#imgrc=-kQ4QjB2OuXJlM:

[43] https://devblogs.nvidia.com/deep-learning-for-computer-vision-with-matlab-and-cudnn/