

République Algérienne Démocratique et Populaire

Ministère de l'Enseignement Supérieure et de la Recherche Scientifique

Université Abderrahmane Mira

Faculté de Technologie



Département d'Automatique, Télécommunication et
d'Electronique

Projet de Fin d'études

Pour l'obtention du diplôme de Master

Filière : Télécommunications.

Spécialité : Réseaux et Télécoms.

Thème

Détection du visage et port du masque : Application pour covid-19

Préparé par :

ALLOU Linda

ADEL Yasmina

Examiné par :

Mr MEKHMOUKH Abdenour

M.C.A U.A.M. Bejaïa

Encadrant

Mr BENAMIROUCHE Nadir

M.C.A U.A.M. Bejaïa

Examineur

Mme GAGAOUA Meriem

M.C.B U.A.M. Bejaïa

Président

Année universitaire : 2020/2021

Remerciement

En premier lieu, on remercie "DIEU" de nous avoir donné tant de patience et de puissance de pouvoir réaliser ce modeste travail.

*On veut adresser nos remerciements à notre encadreur de mémoire Mr **A.MEKHMOUKH** Maître de conférences à l'université Abderrahmane-Mira de Bejaïa, pour sa grande disponibilité et ses précieux conseils tout au long de la rédaction de ce mémoire.*

On tient à remercier tout particulièrement notre famille qui nous a accordé la Liberté d'action et la patience nécessaires pour réaliser ce travail ainsi que toutes les personnes qui nous ont soutenues.

*Un spécial remerciement à **ABDELLI Lyes** qui nous à donner toujours le Soutient et l'aide.*

*En Fin, On Remercie Messieurs Les Membres Du Jury Mr **N.BENAMIROUCHE** et Mme **M.GAGAOUA** Qui Nous Ont Fait L'honneur De Siéger, Espérant Qu'ils Trouvent L'expression De Nos Profonds Respects Et Croire À Notre Sincère Gratitude.*



Dédicaces

Je dédie ce modeste travail aux êtres qui sont les plus chers:

À Mes très chers parents qui m'ont soutenu durant tout mon parcours étudiantin, qui ont fait de moi ce que je suis aujourd'hui, je leur serai éternellement reconnaissante.

À Mes chers frères Adel et Abdelmoumem, source de joie et bonheur.

À Mes chères sœurs Djazia, Hanane et Sonia, source de joie et bonheur.

À Mes plus belles tantes Zahia, Souria et Ghania

À Mes oncles Moussa, Farouk, Zoubir, Lounis et Taher

À Yasmina, chère amie avant d'être une binôme.

À tous mes amis qui me sont chers avec qui j'ai passé les plus belles années d'études qui resteront gravées à jamais : Yasmina, Khaled, Rosa, Lydia, Maria, Leila, Rafik, Lyes, Siham, Yasmine, Aissa, Jugurta, Dalil, Farah, Messilia.

À Mes meilleures amies Sabine et Samiha, qui m'ont soutenu durant toute la période de mon mémoire.

Linda

DEDICACE

Je dédie ce modeste travail à :

À mes chers parents, pour tous vos sacrifices, votre amour, votre tendresse, votre soutien et vos prières tout au long de mes études, aucun hommage ne pourrait être à la hauteur de l'amour dont vous ne cessiez de me combler. Que Dieu vous procure bonne santé et longue vie.

À mon cher frère Lamine et ma chère sœur Kahina Merci de si bien accomplir vos rôles de grand frère et grande sœur. Merci d'être là quand ça ne va pas. Merci de me prêter vos épaules quand j'en ai de besoin. Merci d'apaiser mes pleurs peu importe la situation; Merci de m'aimer telle que je suis, avec mes défauts et mes qualités et de me prouver à quel point vous tenez à moi de mille et une façons; Merci d'être là pour moi, sachez que je vous promets de toujours être là pour vous en retour.

À ma petite princesse Anya : le plus beau cadeau du monde. Tes petites mains, ton envie de parcourir le monde, ton enthousiasme, tes sourires, tes yeux brillants sont incomparables .Tu m'as apporté beaucoup de bonheurs .je t'aime
Princesse

À toute ma famille pour leur soutien tout au long de mon parcours universitaire,

À mon binôme Linda ALLOU et à toute sa famille ;

À mes amis(es) : Aissa , Anya ,Dalil , farid, fouzia, Jugurta, Linda, Lydia ,Lyes ,Lynda ,Leila, Maria, Nadjet, Rafik, Rosa, Sihem ,sofiane, Yasmine les expressions me trahissent, et ne peuvent exprimer mon attachement, mon amour et ma gratitude pour vous en témoignage de l'amitié qui nous unit et des souvenirs de tous les moments que nous avons passés ensemble, je vous dédie ce travail et je vous souhaite une vie pleine de santé et de bonheur.

Yasmina

Résumé :

La pandémie de coronavirus COVID-19 provoque une crise sanitaire mondiale. L'une des méthodes de protection efficaces consiste à porter un masque dans les lieux publics selon l'Organisation mondiale de la santé (OMS). Dans ce mémoire, un modèle utilisant l'apprentissage profond pour la détection de masques sera présenté. Le modèle proposé comprend trois volets : Le premier composant est la détection du visage, en suite le deuxième composant conçu pour l'extraction de caractéristiques. Tandis que le troisième composant est conçu pour le processus de classification de port du masques à l'aide de k plus proche voisin (KNN) et de machine à vecteurs de support (SVM).

Mot-clé: COVID-19, détection du visage, masque, HOG, Algorithme de Jones et viola, KNN, SVM.

Abstract:

The COVID-19 coronavirus pandemic is causing a global health crisis. One of the effective protection methods is to wear a mask in public places according to the World Health Organization (WHO). In this thesis, a model using deep learning for mask detection will be presented. The proposed model has three components : The first component is face detection, followed by the second component designed for feature extraction. While the third component is designed for the mask port classification process using k nearest neighbor (KNN) and carrier vector machine (SVM).

Keyword: COVID-19, face detection, mask, HOG, Jones and viola algorithm, KNN, SVM.

ملخص :

يتسبب وباء فيروس الكورونا في حدوث ازمة عالمية . و من بين وسائل الحماية الفعالة ارتداء قناع في الاماكن العامة وفقا لمنظمة الصحة العالمية . في هذه الاطروحة , سيتم تقديم نموذج باستخدام التعلم العميق للكشف عن القناع . يتكون النموذج المقترح من ثلاث مكونات : المكون الاول هو كشف الوجه , يليه المكون الثاني المصمم لاستخراج الميزة بينما تم تصميم المكون الثالث لعملية تصنيف ارتداء القناع باستخدام K اقرب جار و الة متجه الدعم.

الكلمات الرئيسية: الكورونا , كشف الوجه , القناع , خوارزمية HOG , خوارزمية SVM, JONES et VIOLA

KNN

Listes des figures

Chapitre I : Méthodes de détection de visage

Figure I.1 : Détection de visage.	4
Figure I.2 : Différents types de méthodes de détection de visage.....	4
Figure I.3 : Un visage typique de la méthode basée sur la connaissance.	5
Figure I.4 : Illustration de processeur de détection de peau.....	6
Figure I.5 : La détection de la couleur de peau.	6
Figure I.6 : L'algorithme de base utilisé pour la détection de visage.	8
Figure I.7 : Avant et après la détection du visage.	9
Figure I.8 : Types de caractéristiques de type Haar que Viola et Jones ont identifiées dans leurs recherches.	10
Figure I. 9 : Caractéristiques pseudo Haar intégrer dans un visage.	11
Figure I. 10 : La valeur de l'image intégrale au point (x,y).	12
Figure I. 11 : Image originale et image intégrale.	12
Figure I.12 : Calcul de la somme du rectangle D avec l'image intégrale.	13
Figure I.13 : Illustration de l'architecture de la cascade.	15

Chapitre II : Extraction des caractéristiques du visage.

Figure II.1 : Extraction des caractéristiques de l'histogramme LBP.	17
Figure II. 2 : Génération de code LBP.....	19
Figure II.3: Descripteur SIFT.	20
Figure II. 4 : Procédé du descripteur HOG.....	23
Figure II. 5 : Formation du HOG d'une image: (a) découpage en cellules et formation de l'histogramme par bloc, (b) vote d'un pixel selon l'orientation de ses gradients.	24

Chapitre III : Application pour le port du masque.

Figure III. 1 : Diagramme du système.	27
Figure III. 2 : Détection d'un ou plusieurs visages.	28
Figure III. 3 : Problème de variation de pose des visages (incliné et rotation).	29
Figure III. 4: Résultat de la détection des visages partiellement cachés.	30
Figure III. 5: Résultat de la détection à différent d'éclairage.	30
Figure III. 6: Image montrant la détection des faux positifs.	31
Figure III. 7 : Résultat de détection des visages d'animaux.....	31
Figure III. 8: Résultat dans le cas d'une image de la nature.....	32

Figure III. 9: Détection et extraction des caractéristiques du visage sans masque.....	32
Figure III. 10 : Détection et extraction des caractéristiques du visage avec masque.	33
Figure III. 11 : Extraction des caractéristiques HoG en 24x24 cellules.....	33
Figure III. 12: Extraction des caractéristiques avec en 8x8 cellules.	33
Figure III. 13 : Quelques échantillons de la base de données des visages sans masque.	35
Figure III. 14: Quelques échantillons de la base de données des visages avec masque.	36
Figure III. 15: Représentation de taux de détection de visage sans masque pour les classifieur KNN et SVM.....	37
Figure III. 16: Représentation de taux de détection de visage avec masque pour les classifieur KNN et SVM.....	39

Listes des tableaux

Tableau III.1 : Taux de détection de visage sans masque avec le classifieur SVM.....	36
Tableau III.2 : Taux de détection de visage sans masque avec le classifieur KNN.....	37
Tableau III.3 : Taux de détection de visage avec masque avec le classifieur SVM.	38
Tableau III.4 : Taux de détection de visage avec masque avec le classifieur KNN.	38

Liste des abréviations

ACP: Principal Component Analysis

AdaBoost: Adaptive Boosting

CVPR: Conference on Computer Vision and Pattern Recognition

DoG: Différence Of Gaussiens

HOG: Histogram of Oriented Gradients

IA: Intelligence Artificiel

KNN: K-Nearest Neighbors

LBP: Local Binary Pattern

SIFT: Scale Invariant Feature Transform

SVM : Support Vector Machine

Table des matières

Remerciement	
Dédicace	
Liste des figures	
Liste des tableaux	
Liste des abréviations	
Introduction générale.....	1
Chapitre I : Méthodes de détection de visage	
I.1.Introduction.....	3
I.2.Détection de visage	3
I.3. Méthodes de détection de visage	4
I.3.1. Approche basé sur les connaissances.....	5
I.3.2. Approche basée sur les caractéristiques invariantes	5
I.3.2.1. La couleur de peau	5
I.3.3. Approche basée sur l'appariement de gabarits (Template Matching)	6
I.3.3.1. Des faces prédéfinies de visages	6
I.3.3.2. Modèle déformable :	7
I.3.4. Approche basée sur l'apparence	7
I.3.4.1. Réseaux de neurones:	7
I.3.4.2. Eigen face (algorithme ACP).....	8
I.4.La méthode de Viola et Jones	9
I.4.1. Principe	9
I.4.2. Fonctionnement :	10
I.4.2.1. Fonctionnalités de type Haar :.....	10
I.4.2.2. Créer une image intégrale	11
I.4.2.3.Sélection de caractéristiques par boosting	13
I.4.2.4. Cascade de classifieurs	14

Conclusion.....	16
-----------------	----

Chapitre II : Extraction des caractéristiques de visage

II.1. Introduction	17
II.2. Motif binaire local (LBP: Local Binary Pattern).....	17
II.2.1. Concept de LBP.....	17
II.2.2. Etapes de méthode LBP :.....	18
II.3. Scale Invariant Feature Transform	19
II.3.1. Principe de base	20
II.4. Histogramme du gradient orienté (HOG).....	22
II.4.1.Construction du descripteur HOG	23
II.4.1.1. Calcul du gradient.....	23
II.4.1.2.Construction de l'histogramme	23
II.4.1.3. Formation et normalisation des blocs	24
Conclusion.....	25

Chapitre III : Application pour le port du masque

III.1.Introduction	26
III.2. Environnement de travail	26
III.3. Logiciel et le langage d'implémentation.....	26
III.4.Présentation de l'application.....	26
III.5. Implémentation et résultats	27
III.5.1. Détection du visage	27
III.5.1.1. Test sur des images positives	28
III.5.1.2. Test sur des images négatives.....	31
III.5.2. Extraction des caractéristiques	31
III.5.3. Classification.....	33
III.5.3.1. Classifieur SVM	34
III.5.3.2. Classifieur KNN	34

III.5.3.3. Création de la base d'entraînement	35
III.5.3.4. Base de données	35
III.5.3.5. Classification du visage sans masque.....	35
III.5.3.6. Classification visage avec masque :	37
Conclusion.....	39
Conclusion générale	40

Introduction générale

Le SARS-COV 2 (COVID-19) est un type de famille des coronavirus. Ce virus se propage principalement à cause des gouttelettes produites par les éternuements et la toux. Le lavage régulier des mains, le port d'un masque facial et le maintien de la distance sociale dans un lieu public peuvent être des mesures préventives pour arrêter la propagation du virus [1]. La pandémie de coronavirus COVID-19 provoque une crise sanitaire mondiale, qui généralise le port obligatoire du masque d'hygiène dans l'espace public. Le virus se propage par contact étroit et dans des zones surpeuplées.

Comme pour les systèmes de surveillance routiers et urbains, ou de détection de visages, la question du contrôle automatisé du port du masque se pose. D'un point de vue technique, une véritable révolution s'est opérée ces dernières années notamment dans le domaine de la reconnaissance d'objets grâce à l'apprentissage profond.

Les premiers systèmes de détection de visages ont été développés dans les années 70, ils sont efficaces dans peu d'applications, par exemple l'identification de photographie de passeport. Au début des années 90, plusieurs techniques ont été établies avec le progrès dans le codage de vidéo et la nécessité de l'identification de visage. Ces dernières années, différentes approches ont été développées pour résoudre le problème de détection de visages dans différents environnements et conditions [2].

En 2001 avec la publication de la méthode de Viola et Jones, la première méthode capable à détecter des visages en temps réel. La méthode devient standard est reprise et améliorée par de nombreux chercheurs. C'est également à partir des années 2000 que les méthodes basées sur l'apparence, qui apprennent des modèles de visage à partir d'un ensemble d'images d'apprentissage, se révèlent généralement supérieures aux autres types d'approches.

La détection de visage est la première étape vers des applications plus évoluées, qui nécessitent la localisation du visage. La détection de visage dans les images est une tâche principale dans plusieurs applications basées sur la vision, telles que les systèmes biométriques, la conférence vidéo et les interfaces homme-machine. Cependant, plusieurs problèmes difficiles causés par la diversité des variations, telles que les races humaines, l'illumination, les expressions faciales, les tailles des visages, les positions de la tête (rotations hors plan), les occlusions, les inclinaisons de la tête (rotations dans le plan) et ainsi de suite, empêchent les chercheurs de développer des systèmes efficaces de détection de visage. L'objectif de la détection de visage est de déterminer la présence ou non de visages dans une image et de donner, s'ils sont présents, les positions de ces visages et leurs étendus [3].

Notre travail, a pour objectif de réaliser un système pour le traitement d'images basé sur l'apprentissage dont le but est la détection du visage puis la reconnaissance du port ou non d'un masque sur une personne :

Chapitre I : Donne un aperçu général sur la détection de visage, en exposant les approches utilisées pour accomplir cette tâche.

Chapitre II : Détaille les méthodes d'extraction des caractéristiques dans une image d'une façon générale et plus précisément dans une image de visage. Le chapitre se concentre sur la description de la texture, la qualité d'image la plus utilisée pour exploiter et extraire l'information pertinente de l'image. Ensuite il examine les descripteurs des textures les plus utilisés pour la reconnaissance générale d'un objet et en particulier le visage.

Chapitre III : Consacré à la partie réalisation de notre système où nous présenterons les bases de données utilisées lors de la phase d'apprentissage et de test. Les résultats obtenus seront discutés dans ce même chapitre.

Enfin, la conclusion générale résumera les résultats obtenus ainsi que quelques perspectives sur les travaux futurs.

Chapitre I
méthodes de détection de
visage

I.1.Introduction

Ces dernières années, nous avons assisté à des avancées significatives dans les techniques capables de détecter et de reconnaître les visages. Nos caméras mobiles sont souvent équipées d'une telle technologie où l'on peut voir une boîte autour des visages. Bien qu'il existe des algorithmes de détection de visage assez avancés.

La détection des visages est la première étape vers des applications plus avancées, qui nécessitent le positionnement du visage, telles que la reconnaissance faciale, la reconnaissance des expressions faciales, l'évaluation de l'âge ou du sexe d'une personne, le suivi du visage ou l'estimation de la direction du regard et l'attention visuelle.

Nous vivons actuellement à l'ère de la technologie et nous travaillons dur pour distribuer Talents humains et capacités des machines. Le progrès scientifique et comme l'intelligence artificielle est utilisé pour gérer d'une manière optimale des systèmes et des équipements sophistiqués pour les aider à prendre des décisions pertinentes. Pour un tel résultat, il faut passer par la détection d'objets et de visages leur environnement naturel et réel.

Dans ce chapitre, nous allons présenter différentes méthodes de détection de visage.

I.2.Détection du visage

La détection de visage est un domaine de la vision par ordinateur consistant à détecter un visage humain dans des images fixes ou animées. C'est un cas spécifique de détection d'objet, où l'on cherche à détecter la présence et la localisation précise d'un ou plusieurs visages dans une image [4].

La détection de visage est une technologie informatique basée sur l'intelligence artificielle (IA). Elle peut être appliquée dans de nombreux domaines, notamment la sécurité, la biométrie, l'application de la loi, le divertissement et la sécurité personnelle, la vidéosurveillance, la robotique, le contrôle des interfaces homme-machine, la photographie, l'indexation d'images et la vidéo, la recherche d'images par contenu. La forte activité de recherche en détection de visage a également permis de faire émerger des méthodes génériques de détection d'objet.



Figure I.1 : Détection du visage [3].

I.3. Méthodes de détection de visage

La détection de visage, quelle que soit la technique utilisée, se base toujours sur une capture d'image (vidéo par exemple) ou sur une photo. Les caractéristiques du visage sont ensuite extraites selon la méthode choisie, puis enregistrées dans une base de données.

Dans cette section, nous passons en revue les techniques existant pour détecter les visages à partir d'une image fixe. La plupart des méthodes de détection de visage sont classés en :

- ✓ Approche basée sur les connaissances.
- ✓ Approche basée sur les caractéristiques invariantes.
- ✓ Approche basée sur l'appariement de gabarits.
- ✓ Approche basée sur l'apparence.

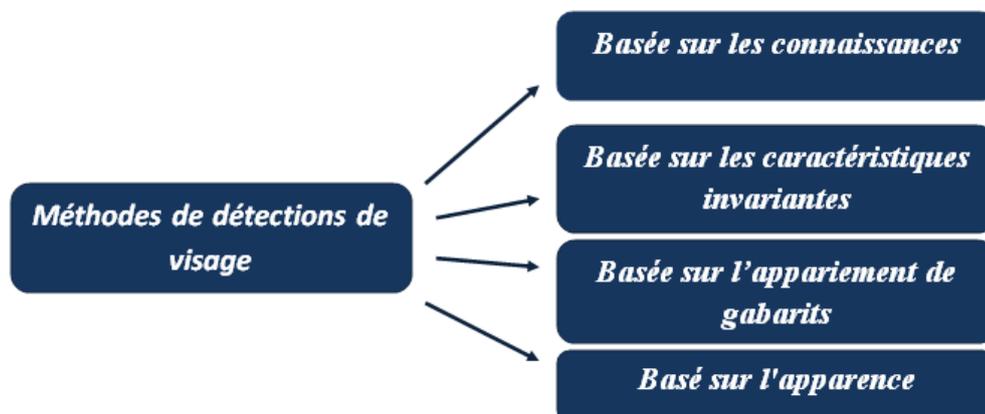


Figure I.2 : Différents types de méthodes de détection de visage.

I.3.1. Approche basée sur les connaissances

La méthode basée sur la connaissance dépend de l'ensemble des règles, elle est basée sur la connaissance humaine pour détecter les visages. Exemple, les visages sur les images ont souvent deux yeux qui sont symétriques, un nez et une bouche à certaines distances et positions les uns avec les autres. Le gros problème avec ces méthodes qu'il est difficile d'établir un ensemble de règles appropriée. Si les règles sont trop générales ou trop détaillées, il peut y avoir beaucoup de faux positifs. Cette méthode seule ne suffit pas : de nombreux visages ne peuvent pas être trouvés dans plusieurs images.

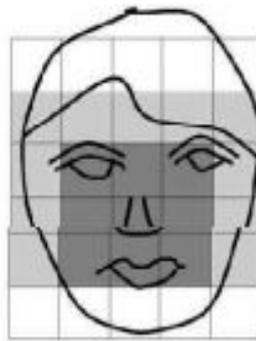


Figure I.3 : Un visage typique de la méthode basée sur la connaissance.

I.3.2. Approche basée sur les caractéristiques invariantes

Cette série d'algorithmes vise à trouver des caractéristiques structurelles même si le visage est dans des positions, des conditions d'éclairage ou des angles de vision différents. Le problème avec cette méthode est que la qualité de l'image sera gravement dégradée en raison de l'éclairage, du bruit ou de l'occlusion. Cependant, Il existe plusieurs propriétés ou caractéristiques invariables du visage dont le principal est la suivante :

I.3.2.1. La couleur de la peau

La couleur de la peau de l'être humain a été utilisée pour la détection des visages et sa pertinence a été prouvée comme caractéristique spécifique au visage. Le principe de cette méthode est basé sur l'information couleur pour la discrimination des pixels de peau ou non-peau. Chaque pixel d'une image couleur est codé dans un espace couleur (par exemple RGB ou YCrCb, ..).

Cette méthode se résume en trois étapes :

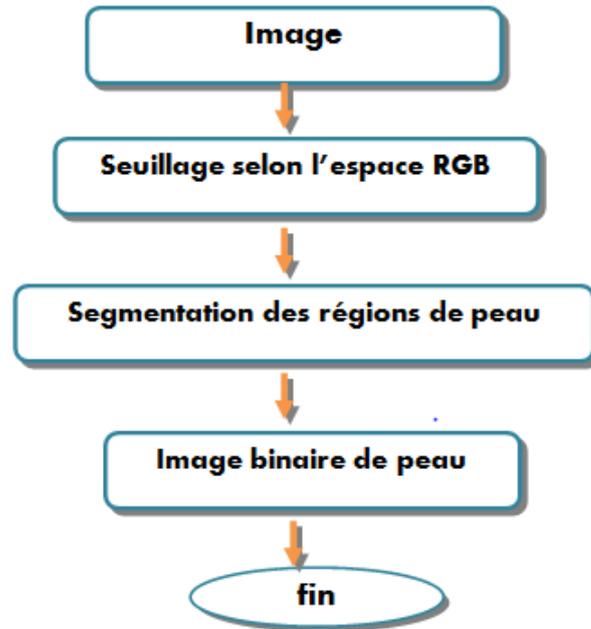


Figure I.4 : Illustration de processeur de détection de peau.

Cette méthode est caractérisée par la rapidité de traitement et par la simplicité de la décision. En effet le principe est simple et limité à la couleur de peau sans aucune considération des effets d'échelle et de position. Néanmoins, cette méthode affiche des détections des faux positifs et peut créer des conflits avec l'arrière-plan.



Figure I.5 : La détection de la couleur de peau [23].

I.3.3. Approche basée sur l'appariement de gabarits (Template Matching)

La détection de visages entiers ou de parties de visage se fait à travers un apprentissage d'exemples standard de visages. La corrélation entre les images d'entrées et les exemples enregistrés est calculée et utilisée pour la décision.

I.3.3.1. Des faces prédéfinies de visages

Cette technique est utilisée pour classer des objets, elle est très intéressante pour la détection de visage de par sa facilité d'application. Le principe de cette méthode est basé sur

une comparaison effectuée entre une image quelconque et un modèle prédéfini, dont le but est de calculer la corrélation pour aboutir à une décision par oui/non. La correspondance est faite pixel par pixel. Cette méthode a l'avantage d'être simple mais elle est sensiblement influencée par la variation d'échelle, de pose et de forme.

I.3.3.2. Modèle déformable

Cette approche est utilisée dans le but de modéliser les caractéristiques faciales qui s'adaptent élastiquement par rapport au modèle du visage présent. Dans cette méthode, les caractéristiques faciales sont décrites par des gabarits paramètres. Une fonction est définie pour relier les contours, les sommets et les angles dans l'image d'entrée, pour faire correspondre les paramètres sur les gabarits. La meilleure adaptation du modèle élastique est de trouver la fonction énergétique en minimisant les paramètres.

I.3.4. Approche basée sur l'apparence

La méthode basée sur l'apparence dépend d'un ensemble d'images de visage de formation des délégués pour découvrir les modèles de visage. Ces méthodes utilisent le même principe que présenté au point précédent mais se basent sur des modèles appris à partir d'un ensemble d'essai. En général, les méthodes basées sur l'apparence se basent sur des techniques d'analyse statistique et d'apprentissage pour trouver les caractéristiques appropriées des images de visage et de non-visage. Les caractéristiques apprises sont exprimé sous la forme d'un modèle de distribution ou d'une fonction discriminante puis utilisée pour les tests.

Ces méthodes présentent l'avantage de s'exécuter très rapidement et obtiennent généralement les meilleures performances mais demandent un long temps d'entraînement. Les méthodes appartenant à cette catégorie ont montré de bons résultats par rapport aux trois autres types de méthodes [5].

On peut citer parmi celles-ci :

- ✓ La méthode basée sur les réseaux de neurones de Rowley et al [22].
- ✓ La méthode de Schneiderman et Kanade basée sur un classifieur de Bayes naïf.
- ✓ Eigen faces (visages propres).
- ✓ Algorithme de Viola et Jones.

I.3.4.1. Réseaux de neurones:

Plusieurs méthodes utilisant les réseaux de neurones pour détecter les visages humains ont été proposées. La plus importante pour la détection des visages était développée par

Rowley, Baluja et Kanade. Leur méthode présente quelques similitudes avec le système de Sung et Poggio [14]. Le système fonctionne en deux étapes :

La première étape est un filtre qui reçoit en entrée une région de 20x20 pixels d'une image et plus une sortie allant de -1 à 1, signifiant la présence ou l'absence d'un visage. Pour détecter les visages plus grands que la taille de la fenêtre, l'image d'entrée est sous-échantillonnée à plusieurs reprises par un facteur de 1,2 et le filtre est appliqué à chaque taille. Après avoir appliqué l'étape de prétraitement appropriée à la fenêtre d'image. La fenêtre est ensuite passée à travers un réseau de neurones qui décide si la fenêtre contient un visage ou non. Le réseau a des connexions de rétine au niveau de la couche d'entrée ; Les champs récepteurs des unités cachées sont représentés sur la figure suivante.

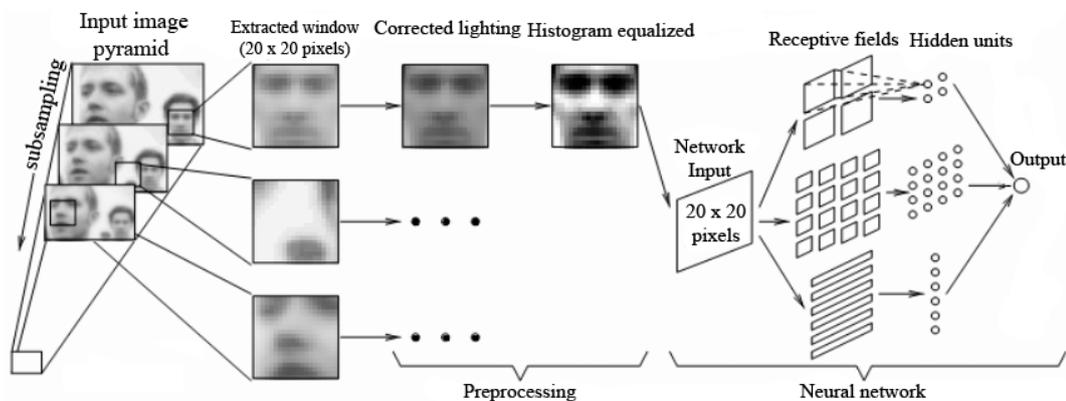


Figure I.6 : L'algorithme de base utilisé pour la détection de visage [22].

Le deuxième composant consiste à fusionner la détection de chevauchement et à arbitrer entre les sorties de plusieurs réseaux. Le réseau a une seule sortie de valeur réelle, qui indique si la fenêtre contient un visage ou non.

L'inconvénient de cette approche réside dans le temps de calcul qui ne permet pas souvent de faire des traitements en temps réel.

I.3.4.2. Eigen face (algorithme ACP)

La méthode Eigen face est très populaire, basée sur la technique ACP. Son principe est le suivant : étant exemple d'un ensemble d'images constituant des visages, il s'agit tout d'abord de trouver les composantes principales de ces visages. Ceci revient à déterminer les vecteurs propres de la matrice de covariance formés par l'ensemble des images exemples. Chaque visage exemple peut alors être décrit par une combinaison linéaire de ces vecteurs propres. Pour construire la matrice de covariance, chaque image de visage est transformée en vecteur. Chaque élément du vecteur correspond à l'intensité lumineuse d'un pixel. L'ACP est une technique rapide, simple et populaire dans l'identification de modèle, C'est l'une des

meilleures techniques. Les projections de l'ACP sont optimales pour la reconstruction d'une base de dimension réduite. Cependant, l'ACP n'est pas optimisé pour la séparabilité (discrimination) de classe.

I.4. La méthode de Viola et Jones

L'algorithme Viola-Jones est un cadre de détection d'objets développé par Paul Viola et Michael Jones en 2001. Ils ont publié la première version de leur papier le 13 juillet 2001. Leurs travaux sont inspirés de ceux de Papageorgiou, Oren et Poggio en 1998[6].

L'algorithme Jones-Viola basée sur l'apparence. C'est l'une des premières méthodes qui peut détecter efficacement des objets en temps réel ou sur l'image fixe. Il a été inventé à l'origine pour détecter les visages humains.

La méthode de Viola et Jones est l'une des plus connues et la plus couramment utilisée. Elle est considérée comme l'une des plus efficaces en détection de visage, et elle est devenue un standard dans le domaine de détection d'objet, en particulier pour la détection des visages et des personnes.

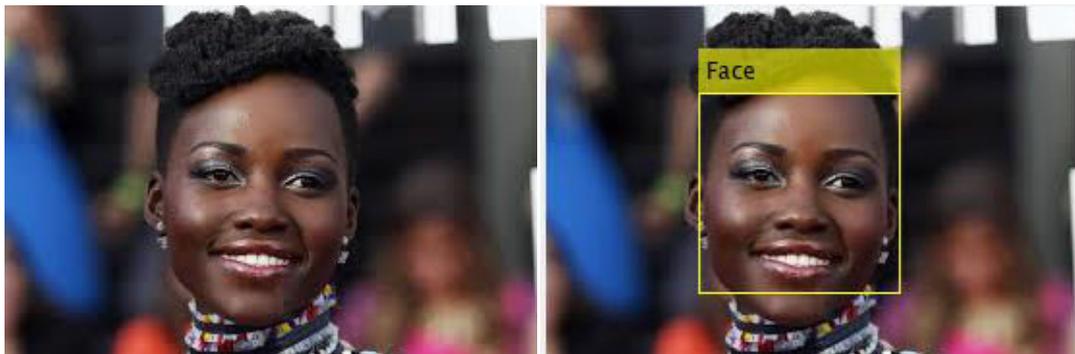


Figure I.7 : Avant et après la détection du visage.

I.4.1. Principe

La méthode de Viola et Jones consiste à balayer une image à l'aide d'une fenêtre de détection de taille initiale 24x24 pixels (dans l'algorithme original) et de déterminer si un visage y est présent. Lorsque l'image a été parcourue entièrement, la taille de la fenêtre est augmentée par un facteur multiplicatif de 1.25 et le balayage recommence, jusqu'à ce que la fenêtre fasse la taille de l'image.

Les caractéristiques sont calculées à toutes les positions et à toutes les échelles dans la fenêtre de détection. Un très grand nombre de caractéristiques par fenêtre sont ainsi générées (environ 160 000 caractéristiques), mais seules quelques-unes sont importantes pour

identifier un visage. Nous utilisons donc l'algorithme **AdaBoost** pour identifier les meilleures fonctionnalités parmi les 160 000 et former des classificateurs.

I.4.2. Fonctionnement

L'algorithme de Viola Jones comporte quatre étapes principales, sont les suivantes :

1. Sélection de fonctionnalités de type Haar.
2. Créer une image intégrale.
3. Sélection de caractéristiques par boosting.
4. Cascades de classificateurs.

I.4.2.1. Fonctionnalités de type Haar

Au 19^{ème} siècle, un mathématicien hongrois, Alfred Haar a donné les concepts d'ondelettes d'Haar, qui sont une séquence de fonctions « carrées » redimensionnées qui forment un ensemble, une famille ou une base d'ondelettes. Viola et Jones ont adapté l'idée d'utiliser des ondelettes d'Haar et ont développé les fonctionnalités dites du type Haar.

Les caractéristiques pseudo-Haar sont des caractéristiques utilisées en vision par ordinateur pour la détection d'objet dans des images numériques. Elles ont été utilisées dans le premier détecteur de visages en temps réel.

Il existe 3 types de caractéristiques du type Haar que Viola et Jones ont identifié dans leurs recherches :

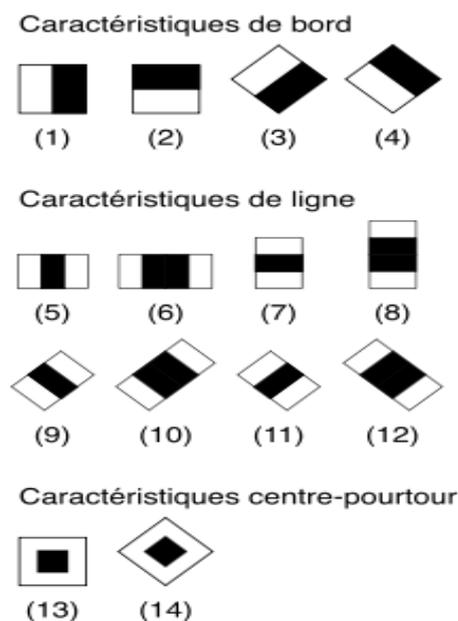


Figure I.8 : Types de caractéristiques de type Haar que Viola et Jones ont identifiées dans leurs recherches [26].

Ces fonctions nous aident à localiser certains objets dans l'image. Pour la détection des visages, nous convertissons l'image en une image en niveaux de gris, car aucune connaissance des couleurs n'est requise lors de la détection des visages dans l'image. Les images en niveaux de gris ne contiennent que des valeurs de niveau d'intensité. Si vous regardez de près ses caractéristiques Haar nous permettent de localiser les contours des images en niveaux de gris. Chaque fois que le niveau d'intensité diminue ou augmente, il y aura des bords, et ses caractéristiques nous permettent de les localiser.

Tous les visages humains partagent certaines propriétés universelles, comme la région des yeux, la région du nez et le contour du visage. Il existe des fonctionnalités connues qui fonctionnent très bien pour détecter les visages humains. (Voir la figure ci-dessus) :

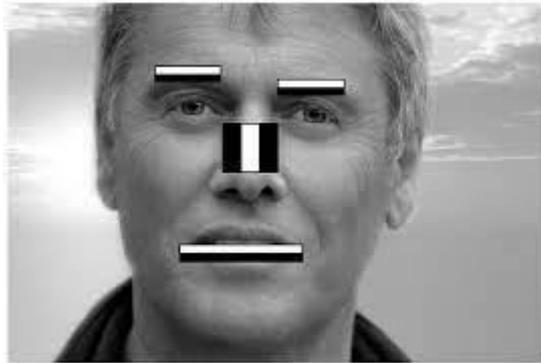


Figure I. 9 : Caractéristiques pseudo Haar intégrées dans un visage [24].

Ceci est juste une représentation de haut niveau de la façon dont ces fonctionnalités fonctionnent sur une image. En fait, pour détecter une seule caractéristique comme les yeux, il peut y avoir des centaines de variations des trois caractéristiques Haar qui devraient être satisfaites pour la passer comme un nez.

Comme il y a des centaines de caractéristiques sur un visage et pour détecter toutes ses caractéristiques, nous devons utiliser des centaines de variantes de caractéristiques Haar, cela devient assez coûteux en termes de calcul. Pour accélérer ce processus, le concept **Images Intégrales** est utilisé.

I.4.2.2. Créer une image intégrale

L'image intégrale, également connue sous le nom de table de surface sommée, est un algorithme permettant de calculer la somme des valeurs dans un sous-ensemble rectangulaire d'une grille [7]. Proposée à l'origine comme une méthode d'infographie en 1984 pour une

utilisation dans les mipmaps, elles sont surtout connues en vision par ordinateur depuis 2001 par leur reformulation dans la méthode de Viola et Jones.

La méthode de l'image intégrale est utilisée pour le calcul rapide et efficace des caractéristiques sur une image. C'est une représentation sous la forme d'une image de même taille que l'image d'origine, elle contient en chacun de ses points la somme des pixels situés au-dessus et à gauche du pixel courant.

L'image intégrale ii au point (x, y) est définie à partir de l'image i par :

$$ii(x, y) = \sum_{x' \leq x, y' \leq y} i(x', y') \quad (I.1)$$

Dans l'image intégrale, la valeur du pixel (x,y) est la somme des pixels au-dessus et à gauche du pixel (x,y) . (Voir la figure ci-dessous)

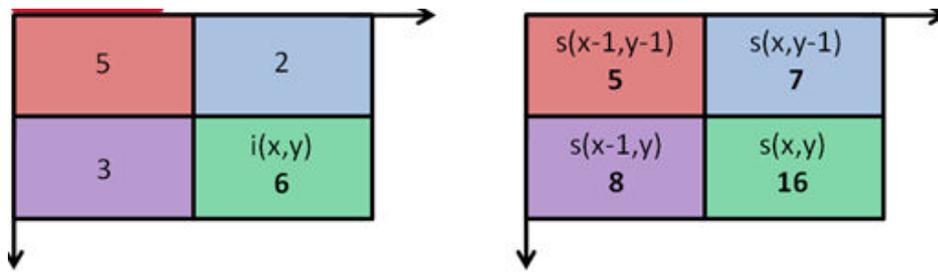


Figure I. 10 : La valeur de l'image intégrale au point (x,y) [5].

Dans la section précédente, nous avons vu comment faire pour calculer une valeur pour chaque caractéristique, nous devons effectuer des calculs sur tous les pixels à l'intérieur de cette caractéristique particulière de l'image originale pour obtenir l'image intégrale. (Voir la figure suivante)

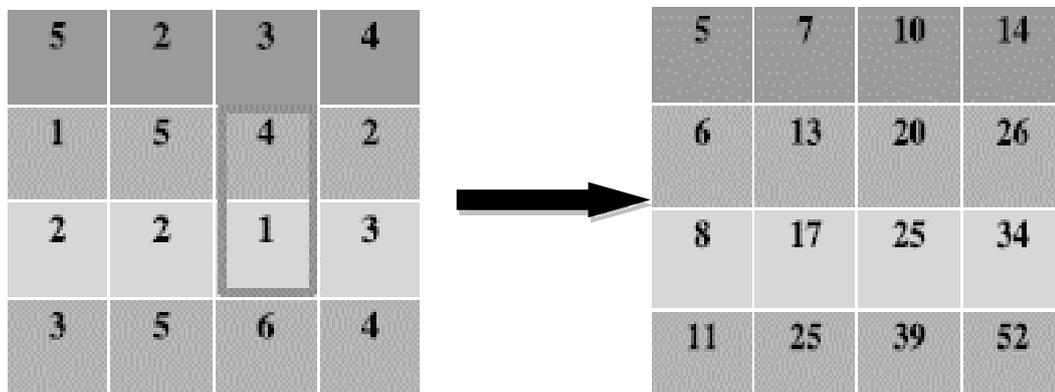


Figure I. 11 : Image originale et image intégrale [5].

En utilisant ces images intégrales, nous gagnons beaucoup de temps en calculant la sommation de tous les pixels d'un rectangle car nous n'avons à effectuer des calculs que sur quatre bords du rectangle. Voir l'exemple suivant pour comprendre.

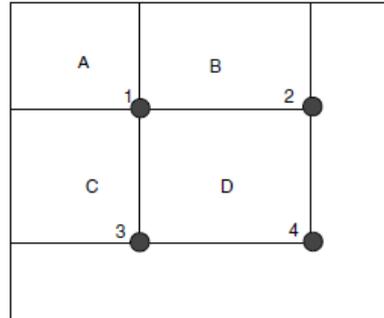


Figure I.12 : Calcul de la somme du rectangle D avec l'image intégrale [6].

La somme des pixels dans le rectangle D peut être calculée avec quatre opérations. La valeur de l'image intégrale au point 1 est la somme des pixels dans le rectangle A. La valeur à la position 2 est A + B, à la position 3 est A + C, et à la position 4 est A + B + C + D. la somme dans D peut être calculée comme $4 + 1 - (2 + 3)$ [6].

Avec cela, nous pouvons tous les substituer dans l'équation suivante :

$$i(x', y') = s(A) + s(D) - s(B) - s(C) \quad (\text{I.2})$$

Le concept d'image intégrale permet de calculer rapidement et efficacement la somme des valeurs de pixels dans une image donnée. Cela permet à son tour de vérifier rapidement un grand nombre de caractéristiques d'Haar. La méthode de l'image intégrale rend les calculs beaucoup moins intensifs et peut faire gagner beaucoup de temps pour n'importe quel modèle de détection faciale.

L'image d'entrée est "scannée" par un rectangle de dimension fixée pour localiser le visage. Le rectangle parcourt l'image dans son intégralité et puis son contenu est analysé pour tester s'il s'agit d'un visage ou non. Cependant pour un rectangle de 24 x 24 pixels, on peut compter environ de 160000 caractéristiques d'Haar et le traitement de tous ces caractéristiques d'Haar ne permettrait pas de détecter un visage en utilisant moins de ressources et rapidement donc pas de perte du temps pour l'exécution. On utilise une variante de la méthode de boosting pour résoudre le problème du temps d'exécution.

I.4.2.3.Sélection de caractéristiques par boosting

L'algorithme de boosting utilisé est en pratique une version modifiée d'AdaBoost, c'est un algorithme d'apprentissage automatique permettant de sélectionner le meilleur sous-

ensemble de fonctionnalités parmi toutes les fonctionnalités disponibles. Le boosting est une méthode pour trouver un classificateur très précis en combinant de nombreux classificateurs "faibles" [7]. Un classifieur faible est tout simplement un classifieur qui fonctionne mal, mais qui fonctionne mieux que de deviner au hasard.

Jones et Viola adaptent ce principe en assimilant une caractéristique à un classifieur faible, en construisant un classifieur faible qui n'utilise qu'une seule caractéristique. L'apprentissage du classifieur faible consiste alors à trouver la valeur seuil de la caractéristique qui permet de mieux séparer les exemples positifs des exemples négatifs. Le classifieur se réduit alors à un couple (caractéristique, seuil).

L'algorithme est de type itératif, à nombre d'itérations déterminées. À chaque itération l'algorithme sélectionne une caractéristique qui sera ajoutée à la liste des caractéristiques sélectionnées aux itérations précédentes, et le tout va contribuer à la construction du classifieur fort final. Cette sélection se fait en entraînant le classifieur faible pour toutes les caractéristiques et en sélectionnant celui avec l'erreur la plus faible sur l'ensemble d'apprentissage. L'algorithme tient également à jour une distribution de probabilité sur l'ensemble d'apprentissage, réévaluée à chaque itération en fonction des résultats de classification. En particulier, plus de poids est attribué aux exemples difficiles à classer, c'est-à-dire ceux dont l'erreur est élevée. Le classifieur « fort » final construit par AdaBoost est composé de la somme pondérée des classifieurs sélectionnés.

AdaBoost sert donc à booster un classifieur existant et a priori chaque classifieur faible possède le même espace d'entrée. Pour la variante d'AdaBoost de Viola et Jones, les classifieurs faibles ont comme donnée entrée une caractéristique de Haar différente. Cette variante d'AdaBoost de Viola et Jones est utilisée lors de l'apprentissage pour sélectionner les caractéristiques de Haar qui sont capables de détecter un visage et permet donc de résoudre le problème du nombre élevé de caractéristiques de Haar existant pour un rectangle de recherche.

I.4.2.4. Cascade de classifieurs

La méthode de Viola et Jones est basée sur une méthode de recherche exhaustive de l'image entière, qui teste l'existence d'objets à toutes les positions et à plusieurs échelles de la fenêtre. Cependant, cette méthode est coûteuse en temps de calcul. Une des idées clés de la méthode pour réduire ce coût est d'organiser l'algorithme de détection dans le classifieur en cascade.

Une cascade de classifieurs est un arbre de décision dégénérer dans laquelle, chaque étape est entraînée pour détecter un maximum d'objets intéressants tout en rejetant une certaine fraction des objets non-intéressants [8]. Elle est composée de plusieurs couches chacune étant la combinaison par boosting de caractéristiques rectangulaires.

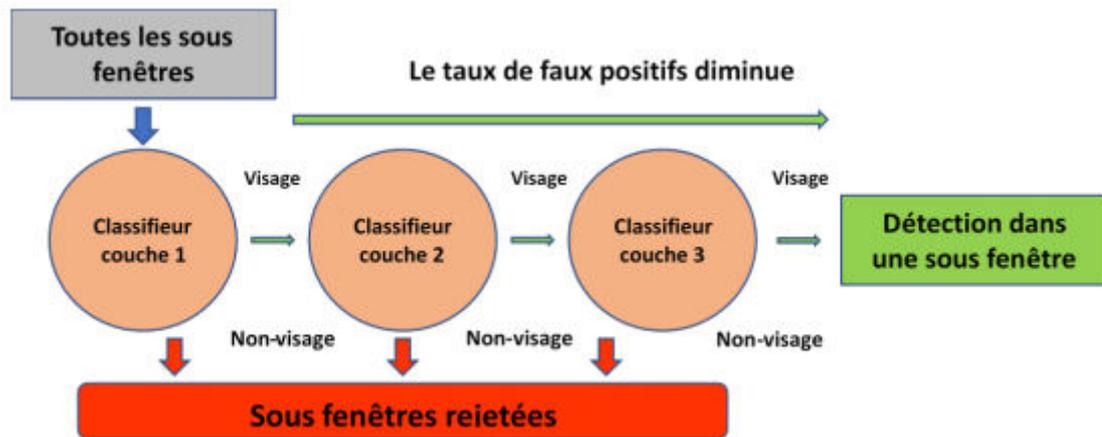


Figure I.13 : Illustration de l'architecture de la cascade [25].

- Si la décision est « fausse » le calcul va s'arrêter et il va forcer le détecteur de retourner. La décision est « vraie » l'image d'entrée passe au prochain classifieur dans la cascade.
- Si tous les classifieurs votent « vraie » alors l'entrée est classifiée comme un vrai exemple (il y a un visage). Si n'importe quel classifieur vote « faux » alors le calcul s'arrête et l'entrée sera classifiée comme fausse.

En pratique, la cascade est constituée d'une succession d'étages, chacune étant formée d'un classifieur fort appris par AdaBoost. L'apprentissage du classifieur de l'étage n est réalisé avec les exemples qui ont passé l'étage $n - 1$; ce classifieur doit donc faire face à un problème plus difficile : plus on monte dans les étages, plus les classifieurs sont complexes [9].

Le choix du nombre K d'étages est fixé par l'utilisateur ; dans leur méthode originale, Viola et Jones utilisent $K = 32$ étages. L'utilisateur doit également spécifier le taux de détection minimal d_i et le taux de fausse alarme maximal f_i à atteindre pour l'étage i . Le taux de détection de la cascade est alors donné par :

$$D = \prod_{i=1}^k d_i \quad (I.3)$$

Et le taux de fausse alarme par :

$$F = \prod_{i=1}^k f_i \quad (I.4)$$

En pratique, les taux d_i et f_i sont les mêmes pour tous les étages. Indirectement, ces taux déterminent également le nombre de caractéristiques utilisées par les classifieurs forts à

chaque étage : les itérations d'AdaBoost continuent jusqu'à ce que le taux de fausse alarme cible soit atteint. Des caractéristiques/classifieurs faibles sont ajoutés jusqu'à ce que les taux cibles soient atteints, avant de passer ensuite à l'étage suivant.

Conclusion

Dans ce chapitre nous avons présenté les différentes méthodes de détection de visage, en expliquant leurs principes de fonctionnement. On se basant sur quelques recherches, Nous avons divisé les méthodes en quatre catégories : méthode basée sur la connaissance, les caractéristiques invariables, Template Matching, méthode basée sur l'apparence.

Nous avons discuté de la méthode de Jones et Viola qui est basée sur l'apparence. Cette méthode est capable de détecter efficacement les visages et des objets en temps réel, sur une vidéo et sur des images. C'est l'une des méthodes les plus connues et les plus utilisées en particulier pour la détection de visages.

Chapitre II
Extraction des
caractéristiques du visage

II.1. Introduction

L'extraction des caractéristiques représente le cœur du système de reconnaissance, on extrait de l'image les informations qui seront sauvegardées en mémoire pour être utilisées plus tard dans la phase de décision.

Après avoir localisé un visage dans une image, la prochaine étape est l'extraction des caractéristiques. L'extraction des informations des visages détectés peut être obtenue des manières suivantes : Décrire les points d'intérêt liés aux traits du visage (comme les yeux), Sourcils, bouche, de cette façon, on obtient un vecteur représentant toutes les données.

Ce chapitre consiste à donner une description principalement pour les descripteurs et les filtres utilisés pour l'extraction des caractéristiques du visage. Nous allons décrire trois méthodes: LBP, SIFT et HOG.

II.2. Motif binaire locale (LBP: Local Binary Pattern)

L'idée principale derrière LBP est de décrire le voisinage des éléments d'image à l'aide de codes binaires. Cette méthode est généralement utilisée pour étudier leurs propriétés locales et identifier les caractéristiques des différentes parties de l'image, il a été proposé en 1996 par Ojala et al.

II.2.1. Concept de LBP

LBP est basé sur la division de l'image en plusieurs petites zones pour extraire les caractéristiques de l'image de chaque point ou pixel [10].

Son concept consiste pour chaque pixel à souiller ses huit voisins directs avec un seuil dont la valeur est le niveau de gris du pixel courant. Tous les voisins prendront une valeur 1 si leur valeur est supérieure ou égale au pixel courant et 0 si leur valeur est inférieure [11].

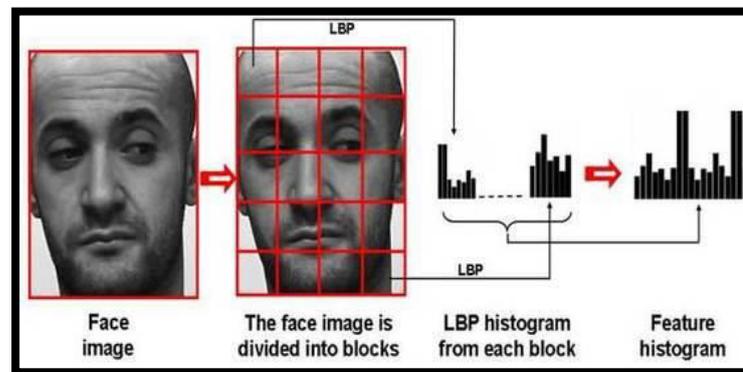


Figure II.1 : Extraction des caractéristiques de l'histogramme LBP [27].

Le code LBP du pixel courant est alors produit en concaténant ces 8 valeurs pour former un code binaire. On obtient donc, comme pour une image à niveaux de gris, une image des valeurs LBP contenant des pixels dont l'intensité se situe entre 0 et 255 [11].

Paramètres : le LBP utilise 4 paramètres :

- **Rayon** : le rayon est utilisé pour construire le motif binaire local circulaire et représente le rayon autour du pixel central. Il est généralement défini sur 1.
- **Voisins** : le nombre de points d'échantillonnage pour construire le motif binaire local circulaire. Gardez à l'esprit : plus vous incluez de points d'échantillonnage, plus le coût de calcul est élevé. Il est généralement défini sur 8.
- **Grille X** : le nombre de cellules dans le sens horizontal. Plus il y a de cellules, plus la grille est fine, plus la dimensionnalité du vecteur d'entités résultant est élevée. Il est généralement défini sur 8.
- **Grille Y** : le nombre de cellules dans le sens vertical. Plus il y a de cellules, plus la grille est fine, plus la dimensionnalité du vecteur d'entités résultant est élevée. Il est généralement défini sur 8.

II.2.2. Etapes de la méthode LBP

- Convertissez l'image en espace en niveaux de gris.
- Nous pouvons obtenir une partie de cette image sous la forme d'une fenêtre de 3x3 pixels.
- Il peut également être représenté sous la forme d'une matrice 3x3 contenant l'intensité de chaque pixel (0 ~ 255).
- Ensuite, nous devons prendre la valeur centrale de la matrice à utiliser comme seuil.
- Cette valeur sera utilisée pour définir les nouvelles valeurs des 8 voisins.
- Pour chaque voisin de la valeur centrale (seuil), nous définissons une nouvelle valeur binaire. Nous fixons 1 pour les valeurs égales ou supérieures au seuil et 0 pour les valeurs inférieures au seuil.
- Désormais, la matrice ne contiendra que des valeurs binaires (en ignorant la valeur centrale). Nous devons concaténer chaque valeur binaire de chaque position de la matrice ligne par ligne dans une nouvelle valeur binaire (par exemple 10001101).
- Remarque: certains auteurs utilisent d'autres approches pour concaténer les valeurs binaires (par exemple dans le sens des aiguilles d'une montre), mais le résultat final sera le même.

- Ensuite, nous convertissons cette valeur binaire en une valeur décimale et la définissons sur la valeur centrale de la matrice, qui est en fait un pixel de l'image d'origine.
- A la fin de cette procédure (procédure LBP), nous avons une nouvelle image qui représente mieux les caractéristiques de l'image d'origine.

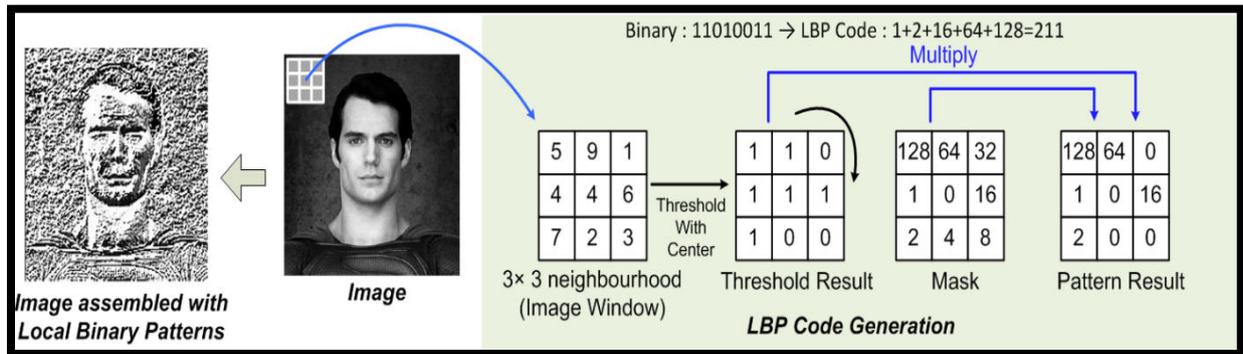


Figure II. 2 : Génération de code LBP [28].

Le processus global comprend les étapes suivantes (illustrées à la figure II.2) :

- 1) Seuil des valeurs dans un voisinage (fenêtre d'image) avec le seuil choisi en plaçant 1 où la valeur est supérieure ou égale au seuil et 0 sinon.
- 2) Multiplier la carte binaire résultante avec un masque prédéfini (généralement des puissances incrémentielles de deux).
- 3) Additionner les valeurs pour obtenir un code LBP de 8 bits.

Le résultat du traitement LBP est une image assemblée par les caractéristiques LBP.

II.3. Scale Invariant Feature Transform

Scale-invariant feature transform (SIFT), que l'on peut traduire par "transformation de caractéristiques visuelles invariante à l'échelle", est un algorithme de traitement d'image qui permet de détecter et de décrire les caractéristiques d'une image. Il permet de transformer une image en un ensemble de caractéristiques, chacune étant invariante aux transformations suivantes : translation, changement d'échelle, rotation et partiellement invariant au changement d'éclairage. Il a été développé en 1999 par le chercheur David Lowe, et le propriétaire du brevet est l'Université de la Colombie-Britannique [15].

II.3.1. Principe de base

La méthode proposée par Lowe comprend deux parties :

- ✓ Un algorithme de détection de caractéristiques et de calcul de descripteurs
- ✓ Un algorithme de mise en correspondance proprement dit.

Cet algorithme consiste à rechercher des points caractéristiques (souvent appelés "features points" ou points d'intérêt) sur une photo qui seront décrits chacun par des coordonnées, une orientation, une échelle, ainsi que 128 descripteurs. L'idée générale de SIFT est donc de trouver des points-clés qui sont invariants à plusieurs transformations de rotation, échelle, illumination et changements mineurs du point de vue (inférieur à 30 degrés) [15].

Les points d'intérêt sont des points particuliers des contours : des points isolés, des coins, des intersections.

L'étape fondamentale de la méthode proposée par Lowe consiste à calculer ce que l'on appelle les "descripteurs SIFT" des images à étudier. Il s'agit d'informations numériques dérivées de l'analyse locale d'une image et qui caractérisent le contenu visuel de cette image de la façon la plus indépendante possible de l'échelle [16].

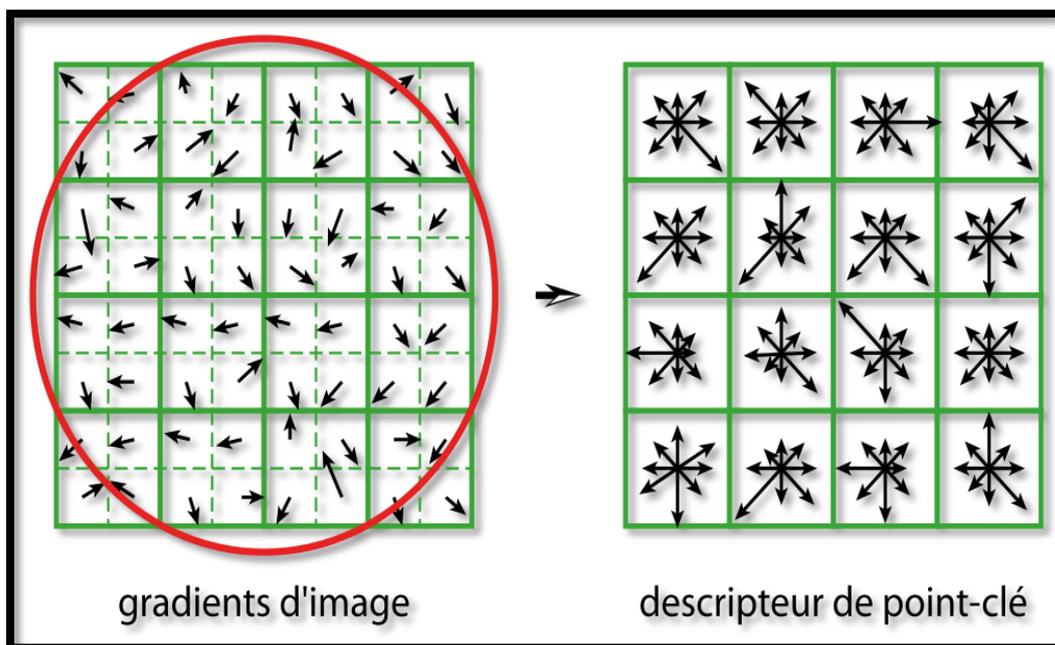


Figure II.3: Descripteur SIFT [29].

Le descripteur SIFT bien connu se compose d'histogrammes 3D de gradients d'image sur des coordonnées spatiales et des orientations de gradient.

À Gauche : Les dégradés de l'image sont pondérés par une fenêtre gaussienne, indiquée par le cercle rouge. La longueur des flèches correspond à la somme des grandeurs de gradient sur une direction donnée.

À droite : les gradients dans chacun des blocs spatiaux 4×4 sont collectés dans des histogrammes de 8 cases chacun pour former le descripteur SIFT 128D final.

Comme il est proposé par Lowe, les principales étapes pour calculer la SIFT des images à étudier sont : La première étape de l'algorithme est la représentation espace-échelle définie, par une convolution entre une fonction gaussienne paramétrable et une fonction f représentant l'image d'entrée. La détection des points d'intérêt présent dans l'algorithme SIFT se fait dans l'espace des échelles. Les emplacements des points d'intérêt sont définis comme étant l'extremum du résultat de la Différence de Gaussiennes (DoG), appliqué, dans l'échelle spatiale sur, une série d'image lissées et ré-échantillonnées [17].

L'espace d'échelle d'une image $L(x, y, \sigma)$ et le produit de convolution entre la fonction gaussienne $G(x, y, \sigma)$ avec l'image d'entrée $I(x, y)$ [17] :

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (\text{II.1})$$

Avec * le produit de convolution, G la fonction gaussienne.

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (\text{II.2})$$

Pour extraire des points d'intérêt stable dans l'espace échelle, Lowe a proposé d'utiliser l'extremum de la fonction de différences gessiennes, qui peut être calculée à partir de la différence de deux échelles séparées d'une constante multiplicative k. La fonction DoG est donnée par :

$$G(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (\text{II.3})$$

Cette première étape permet d'obtenir un grand nombre de points d'intérêt. Selon la méthode SIFT, une phase de filtrage de ces points est effectuée pour extraire les plus intéressants. Pour qu'un point soit sélectionné, il doit satisfaire deux conditions: La première, exige que la valeur $D(x, y, \sigma)$ soit supérieure à un certain seuil prédéfinie (fixée à 0.03 par Lowe), la seconde, s'intéresse à la matrice gaussienne de D :

$$M^L = \begin{bmatrix} \frac{\partial D^2}{\partial x} & \frac{\partial D}{\partial x} \frac{\partial D}{\partial y} \\ \frac{\partial D}{\partial x} \frac{\partial D}{\partial y} & \frac{\partial D^2}{\partial y} \end{bmatrix} \quad (\text{II.4})$$

A partir de cette matrice, le critère de sélection de Lowe est le suivant :

$$R^L = \frac{T_r(M^L s)}{\text{Det}(M^L)} \leq \frac{(r+1)^2}{r} \quad (\text{II.5})$$

$$M^L = e \frac{x^2+y^2}{2\sigma^2} \begin{bmatrix} \frac{\partial I^2}{\partial x} & \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \frac{\partial I^2}{\partial y} \end{bmatrix} \quad (\text{II.6})$$

Avec $\text{Det}(\cdot)$ le déterminant, T_r la trace et r un paramètre de critère. Une valeur élevée du paramètre r permet de s'assurer que le point considéré est un point d'intérêt.

Un point d'intérêt est défini par 5 paramètres $(x, y, \sigma, \theta, \vartheta)$, où les point (x, y) représentent les coordonnées dans l'image originale. Les paramètres (σ, θ) décrivent l'échelle et respectivement l'orientation. Le vecteur ϑ est le vecteur descripteur qui est calculé en utilisant son voisinage qui est divisé par un gris de 4×4 . Ensuite, le gradient est calculé sur chacune des 16 localisations de la grille puis est quantifié selon un histogramme à 8 orientations. La concaténation de ces éléments nous permet d'obtenir un vecteur de 128 éléments [17].

II.4. Histogramme du gradient orienté (HOG)

Afin d'explorer de nouvelles méthodes d'extractions de caractéristiques dans l'objectif d'améliorer les performances de reconnaissance de visages, pour un descripteur représentatif, le plus discriminatoire et le plus rapide en termes de temps de calcul consiste à appliquer les descripteurs HOG.

Les descripteurs HOG sont introduits par Navne, Dalal et Bill Triggs, chercheurs à l'INRIA de Grenoble, à la conférence CVPR de juin 2005 dans leurs travaux de détection des piétons [12].

L'idée essentielle derrière l'histogramme de gradient orienté c'est que l'apparence locale et la forme d'objet dans une image peuvent être décrites par la distribution d'intensité des gradients ou de direction des contours [13]. Le visage est détecté et subdivisée en blocs de tailles égales et ces derniers sont également subdivisés à leur tour en cellules. Pour chacune des cellules, une analyse des gradients des pixels est accomplie afin de former un

histogramme de gradient à neuf bandes. La combinaison de ces histogrammes représente alors le descripteur.

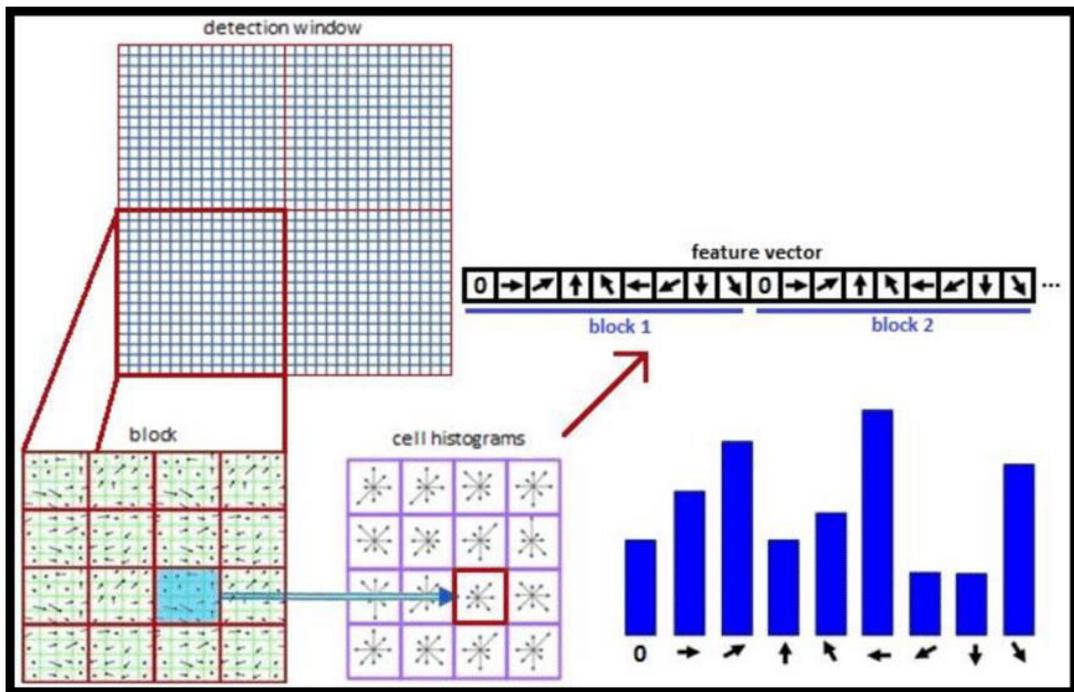


Figure II. 4 : Procédé du descripteur HOG [13].

II.4.1. Construction du descripteur HOG

II.4.1.1. Calcul du gradient

Une étape de prétraitement peut être effectuée avant le calcul du gradient, afin que les couleurs de l'image soient normalisées et une correction gamma correcte. Cette étape ne s'est finalement pas avérée nécessaire, la normalisation du descripteur lui-même s'avérant suffisante [13].

II.4.1.2. Construction de l'histogramme

Dans cette étape l'image est divisée en (8*8) cellules et on calcule l'histogramme d'orientation de gradient pour chaque cellule, qui vote alors pour une classe de l'histogramme, en fonction de l'orientation du gradient à ce point. Le vote du pixel est pondéré par l'intensité du gradient en ce point.

II.4.1.3. Formation et normalisation des blocs

Une étape importante est la normalisation des descripteurs afin d'éviter les disparités dues aux variations d'illumination. Cette étape introduit également de la redondance dans le descripteur. Pour cela, les auteurs regroupent plusieurs cellules dans un bloc, qui est l'unité sur laquelle est effectuée la normalisation.

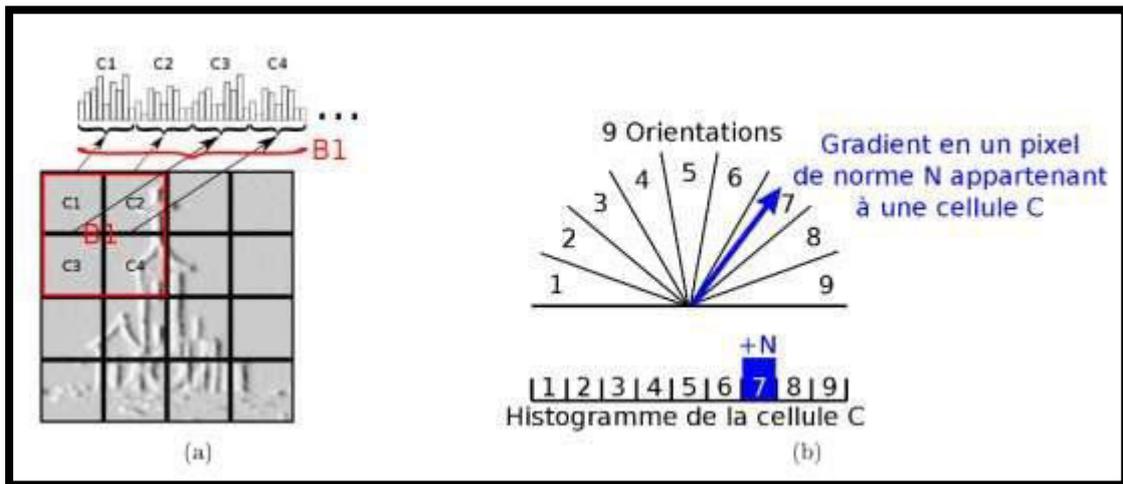


Figure II. 5 : Formation du HOG d'une image: (a) découpage en cellules et formation de l'histogramme par bloc, (b) vote d'un pixel selon l'orientation de ses gradients [17].

Les cases de direction sont définies par l'angle relatif par rapport à la direction de gradient dominante D de la région d'image. Enfin, combiner ensemble l'histogramme d'orientation Gradient de la zone de points d'intérêt pour former un vecteur de caractéristiques de taille 128 dimension. Le vecteur de HOG pour chaque cellule est normalisé en utilisant des vecteurs dans leurs blocs respectifs [17].

Il existe une corrélation entre le nombre de cellules de bloc et la précision de descripteur, si le nombre de cellules dans un bloc est élevé, le descripteur est plus discriminant. Mais la taille des cellules est alors plus petite. La taille des cellules doit cependant rester assez importante car, si les cellules sont très petites, les informations qu'elles contiennent sont trop peu nombreuses et donc peu pertinentes, de plus, un grand nombre de blocs donne un résultat final plus précis [18].

Conclusion

Dans ce chapitre nous avons brièvement discuté quelques techniques d'extraction de caractéristiques (la méthode de LBP, SIFT et le descripteur HOG) et leurs fonctionnements. Ces extracteurs sont souvent présentés dans le cadre d'une problématique donnée ou pour un type d'images spécifique.

Ce sont des technologies de base servant dans de nombreux domaines de vision par ordinateur. Ils peuvent servir d'entrée à un algorithme d'apprentissage supervisé en vue de leur classification. Cela peut notamment servir à détecter leur contenu et en particulier les objets présents dans l'image. Dans la même veine mais dans des domaines plus spécifiques, ils peuvent servir à détecter des visages ou des personnes.

Chapitre III

Application pour le port du masque

III.1.Introduction

Ce chapitre est essentiellement consacré à la conception et réalisation de notre application. La première partie de ce chapitre est une présentation des outils exploités pour le développement du logiciel tels que le choix du langage de programmation, l'environnement de programmation et le matériel utilisé. Tandis que la seconde partie nous déployons les approches théoriques utilisées dans le cadre de la conception de notre système de détection de visage et le port du masque.

III.2. Environnement de travail

On a utilisé :

- + Un ordinateur portable *Acer Extensa EX2519-C1M0*.
- + Windows 8.1 Professionnel.
- + Processeur Intel (R) Celeron (R) CPU N3060 @ 1.60GHZ 1.60 GHz.
- + Mémoire installée (RAM) 4,00Go.
- + Type du système : Système d'exploitation 64 bits, processeur x64.

III.3. Logiciel et le langage d'implémentation

Pour la réalisation de notre système nous avons choisi le langage de programmation MATLAB (R2021a).MATLAB est un environnement de calcul scientifique et de visualisation de données. Sa facilité d'apprentissage et d'utilisation (due à une syntaxe très claire) en ont fait un standard adapté pour les divers problèmes l'ingénierie. Parmi les raisons qui nous ont poussés à l'utiliser, on trouve :

- + Ses très nombreuses fonctions prédéfinies et prêtes à l'emploi.
- + Sa simplicité à l'implémentation et rapidité de calculs.
- + Sa fiabilité et sa robustesse.

MATLAB offre un certain nombre de fonctionnalités pour la documentation et le partage du travail. On peut intégrer le code MATLAB avec d'autres langages et applications, et distribuer les algorithmes et applications MATLAB.

III.4.Présentation de l'application

L'objectif principal de ce projet est la détection du visage puis du port ou non du masque sur des images fixes.

Notre système est divisé en deux parties :

✚ 1^{ère} partie : détection de visage

Utilisation d'un détecteur pré-entraîné (algorithme Jones et Viola) pour extraire automatiquement les visages sur les images.

✚ 2^{ème} partie : entraînement et classification des visages

Création de deux bases d'entraînement (masque et sans masque) avec l'extraction des caractéristiques des visages qui seront sauvegardé en mémoire pour être utilisé dans la partie test.

Adaptation d'un classifieur SVM et KNN pré-entraîné pour classifier automatiquement les visages portant un masque ou non.

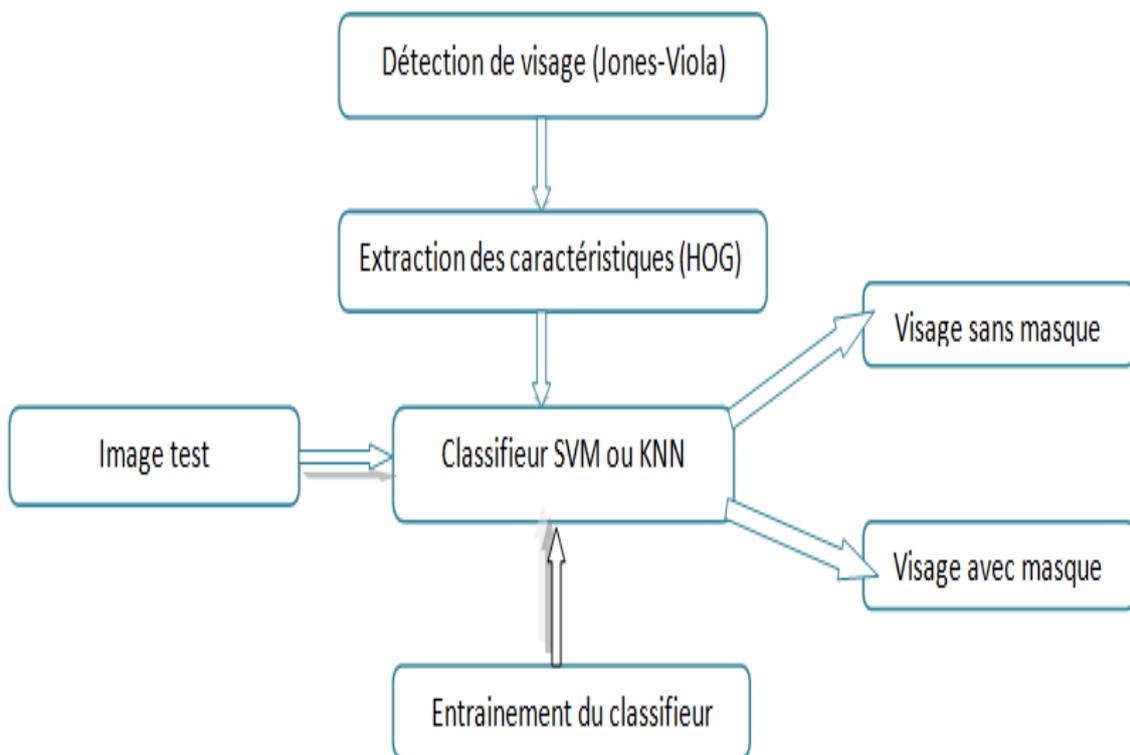


Figure III. 1 : Diagramme du système.

III.5. Implémentation et résultats

III.5.1. Détection du visage

Comme mentionné plus haut, la première étape pour réaliser notre application, est la détection du visage. Nous avons utilisé l'algorithme de viola-Jones pour cette tâche.

III.5.1.1. Test sur des images positives

✚ Visage en face

Le programme détecte bien les visages dans l'image contenant un seul ou plusieurs visages.

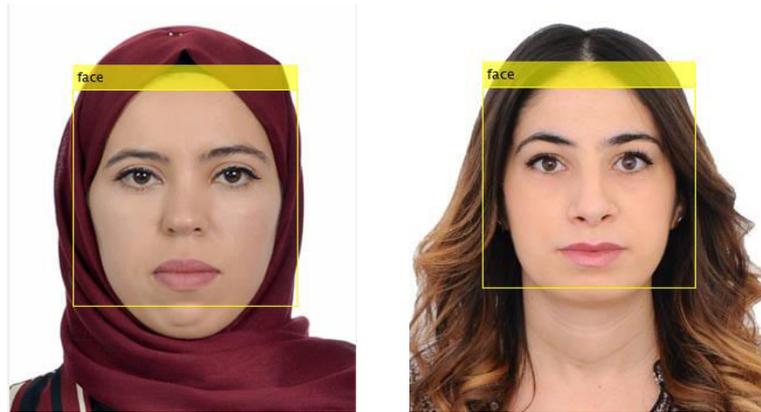


Figure III. 2 : Détection d'un ou plusieurs visages.

✚ Image contenant des visages à différent pose

Le programme ne détecte pas les visages quand les têtes présentées dans l'image sont penchées. Ce problème est dû au fait que tous les visages utilisés comme références regardent directement l'appareil photo.



Figure III. 3 : Problème de variation de pose des visages (incliné et rotation).

🚦 Image contenant des visages Occlus

Dans ce cas, l'algorithme s'exécute mal si un visage est partiellement caché par un autre ou par un objet quelconque.



Figure III. 4: Résultat de la détection des visages partiellement cachés.

🚦 Images à différentes conditions d'illumination

A cause de l'existence de changement de l'éclairage, un même visage avec la même expression faciale, l'algorithme donne un mauvais résultat.



Figure III. 5: Résultat de la détection à différent d'éclairage.

✚ Images fausses positive

La valeur des faux positifs apparaissent en quelques images. Généralement ils apparaissent dans des objets contenant une couleur presque exacte à celle de la peau du visage et lorsqu'une forme d'un objet quelconque et le visage contiennent une corrélation significative.



Figure III. 6: Image montrant la détection des faux positifs.

III.5.1.2. Test sur des images négatives

✚ Cas animal

La performance de l'exécution du programme est dégradée pour les images qui contiennent des animaux. Ce phénomène est dû à la similarité des caractéristiques des visages humains et quelques animaux.

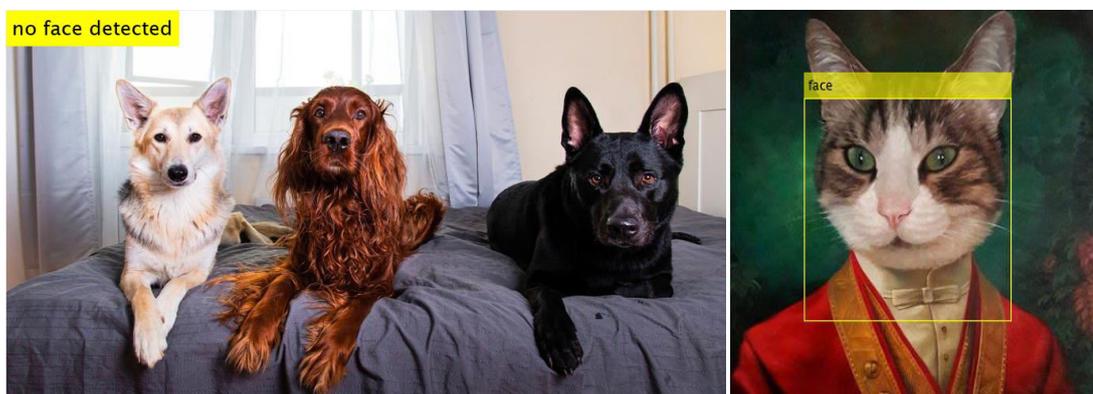


Figure III. 7 : Résultat de détection des visages d'animaux.

✚ Cas paysage

L'algorithme de Jones et Viola ne détecte aucun objet dans le cas d'une image de la nature. Donc il donne de bons résultats pour ce type d'image.



Figure III. 8: Résultat dans le cas d'une image de la nature.

III.5.2. Extraction des caractéristiques

La méthode utilisée dans ce projet est le descripteur de caractéristiques (Histogram of Oriented Gradients) (HOG). Ce dernier est un descripteur de caractéristiques utilisé dans la vision par ordinateur et le traitement d'images à des fins de détection d'objets. La technique compte les occurrences d'orientation de gradient dans des parties localisées d'une image. Un descripteur de caractéristique est la représentation d'une image ou d'un patch d'image qui simplifie l'image en extrayant des informations utiles et en jetant des informations superflues et forment un vecteur ou matrice. En suite Le vecteur de caractéristiques est introduit dans un algorithme de classification d'images et comparés à une base de données pour trouver la meilleure correspondance.

Dans cette étape, l'image est divisée en 32x32 cellules et un histogramme de gradients est calculé pour chaque 32x32 cellule.



Figure III. 9: Détection et extraction des caractéristiques du visage sans masque.

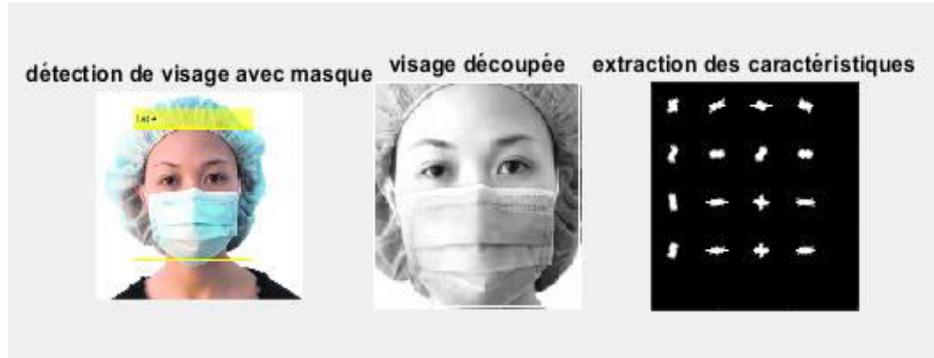


Figure III. 10 : Détection et extraction des caractéristiques du visage avec masque.

Nous faisons varier la dimension des cellules :

La figure III.11 illustre une détection du visage puis l'extraction des caractéristiques HoG d'un visage en 24x24 cellules.



Figure III. 11 : Extraction des caractéristiques HoG en 24x24 cellules.

La figure III.12 illustre une détection du visage puis l'extraction des caractéristiques HoG d'un visage en 8x8 cellules.

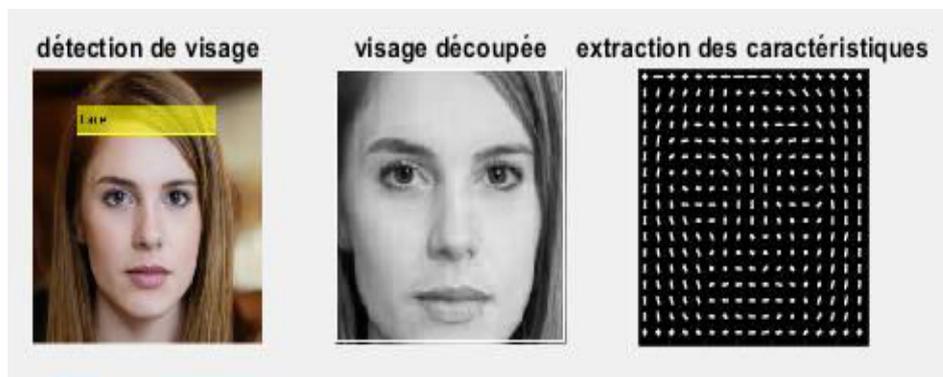


Figure III. 12: Extraction des caractéristiques avec en 8x8 cellules.

Nous avons remarqué qu'à chaque fois la dimension des cellules diminuent, le nombre des caractéristiques augmente et à chaque fois que la dimension des cellules augmente, le nombre des caractéristiques diminue.

III.5.3. Classification

La classification automatique est le processus qui permet d'analyser et d'organiser un ensemble de données selon leurs caractéristiques dans des classes de similarité. Elle se base principalement sur des représentations classiques de données dont les limites de traitement sont connues et qui dans la plupart du temps, demande un temps de calcul énorme [11].

Il existe deux catégories de classification : classification supervisée et classification non supervisée. Dans la première catégorie, les méthodes consistent à classer les objets à partir d'une base de données dite d'apprentissage, tandis que dans l'autre, les méthodes classent les objets sans avoir besoin de cette base de données. Dans la littérature scientifique, plusieurs méthodes de classification ont été présentées. Dans notre application nous allons utiliser deux classifieurs : **KNN (K plus proches voisins)**, **SVM (support vecteur machine)**.

III.5.3.1. Classifieur SVM

SVM est une série d'algorithmes d'apprentissage automatique qui peut résoudre les problèmes de classification, de régression et de détection d'anomalies ; leur but est de diviser les données en catégories en utilisant des frontières aussi simples que possible, afin que la distance entre les différents groupes de données et la frontière qui les sépare soit maximisée. Cette distance est également appelée "marge", donc SVM est appelé "séparateur de marge large", Et "vecteur de support" qui est la donnée plus proche de la limite «la frontière». La distance de la frontière avec le vecteur de support doivent être maximisés [19].

Afin de trouver cette fameuse frontière séparatrice, nous devons fournir des données d'entraînement SVM. Nous donnons à l'algorithme un ensemble de données dans lequel nous apprenons déjà ses deux classes. On entre alors dans la phase d'entraînement. Le SVM va déterminer la frontière la plus plausible.

Après la phase d'entraînement, le SVM a appris où est l'emplacement de la frontière à partir des données d'entraînement. Ainsi qu'après plusieurs phases d'entraînement, le SVM sait où placer la frontière pour de nouvelles données.

III.5.3.2. Classifieur KNN

L'algorithme KNN est l'un des plus simples de tous les algorithmes d'apprentissage automatique. Il s'agit d'un algorithme de classification supervisé, c'est-à-dire qu'il doit obligatoirement posséder un ensemble d'apprentissages auquel se référer durant tout le classement [20].

Pour trouver la classe d'un nouveau cas, cet algorithme se base sur le principe suivant : il cherche les k plus proches voisins de ce nouveau cas, ensuite, il choisit parmi les candidats trouvés le résultat le plus proche et le plus fréquent. Pour affecter un nouvel individu à une classe, l'algorithme cherche les k plus proches voisins parmi les individus déjà classés. Ainsi que l'individu est affecté à la classe qui contient le plus d'individus parmi les candidats trouvés [21].

Dans notre application nous avons utilisé l'hybridation entre deux classifieurs supervisés KNN et SVM en créant une base d'entraînement pour entraîner notre classifieur.

III.5.3.3. Création de la base d'entraînement

Afin de constituer la base d'entraînement pour les deux classifieurs, nous avons choisi quelques images parmi toutes les bases de données utilisées (20 visages pour chacune). Nous avons extrait les caractéristiques des visages puis formé la matrice caractéristique.

III.5.3.4. Base de données

Pour l'évaluation de notre système, nous avons utilisé quelques échantillons de la base de donnée ORL, d'autres bases de données existant et les photos de nos amis pour former notre propre base de teste.

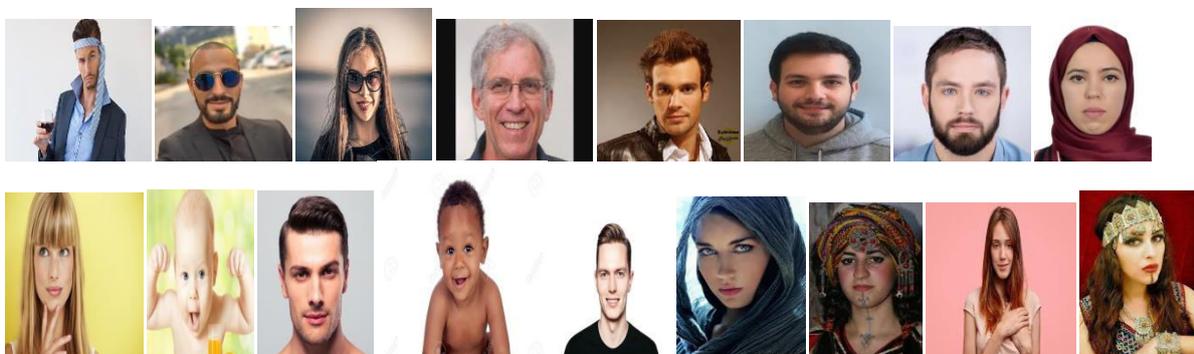


Figure III. 13 : Quelques échantillons de la base de données des visages sans masque.

Pour la base de donnée des visages portant un masque nous avons utilisé quelques échantillons existant sur Google et les photos de nos amis pour former notre propre base de teste.



Figure III. 14: Quelques échantillons de la base de données des visages avec masque.

III.5.3.5. Classification du visage sans masque

Cette partie s'intéresse à la classification des visages sans masque en utilisant deux classifieurs SVM et KNN. On obtient les résultats suivants :

✚ Classification SVM

Classifieur SVM	Blocs 8*8	Blocs 16*16	Blocs 24*24	Blocs 32*32	Blocs 64*64
100 images	88	90	88	85	82
Taux de détection(%)	88%	90%	88%	85%	82%
200 images	177	182	181	176	167
Taux de détection(%)	88,5%	91%	90,5%	88%	83,5%

Tableau III.1 : Taux de classification de visage sans masque avec le classifieur SVM.

Classification KNN

Classifieur KNN	Blocs 8*8	Blocs 16*16	Blocs 24*24	Blocs 32*32	Blocs 64*64
100 images	95	98	99	99	93
Taux de détection(%)	95%	98%	99%	99%	93%
200 images	189	196	196	197	191
Taux de détection(%)	94.5%	98%	98%	98,5%	95,5%

Tableau III.2 : Taux de classification de visage sans masque avec le classifieur KNN.

La figure III.15 illustre la variation du taux de classification du visage sans masque pour le classifieur SVM et le classifieur KNN en fonction de la dimension des cellules.

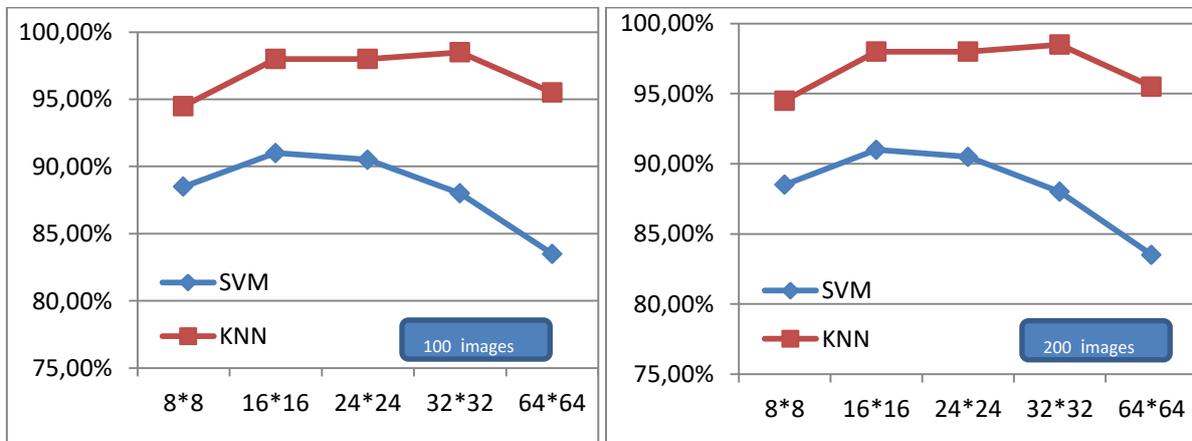


Figure III. 15: Représentation de taux de classification de visage sans masque pour les classifieur KNN et SVM.

D’après cette figure nous remarquons très bien que le taux de classification de classifieur KNN offre de meilleurs résultats par rapport au classifieur SVM.

III.5.3.6. Classification visage avec masque

Cette partie s’intéresse à la classification des visages portant un masque :

Classification SVM

Classifieur SVM	Blocs 8*8	Blocs 16*16	Blocs 24*24	Blocs 32*32	Blocs 64*64
100 images	97	99	100	99	93
Taux de détection(%)	97%	99%	100%	99%	93%
200 images	196	198	200	199	191
Taux de détection(%)	98%	99%	100%	99,5%	95,5%

Tableau III.3 : Taux de classification de visage avec masque avec le classifieur SVM.

Classification KNN

Classifieur KNN	Blocs 8*8	Blocs 16*16	Blocs 24*24	Blocs 32*32	Blocs 64*64
100 images	81	88	92	87	82
Taux de détection(%)	81%	88%	92%	87%	82%
200 images	179	184	190	185	178
Taux de détection(%)	89,5%	92%	95%	92,5%	89%

Tableau III.4 : Taux de classification de visage avec masque avec le classifieur KNN.

La figure III.16 illustre la variation du taux de classification du visage avec masque pour le classifieur SVM et le classifieur KNN en fonction de la dimension des cellules.

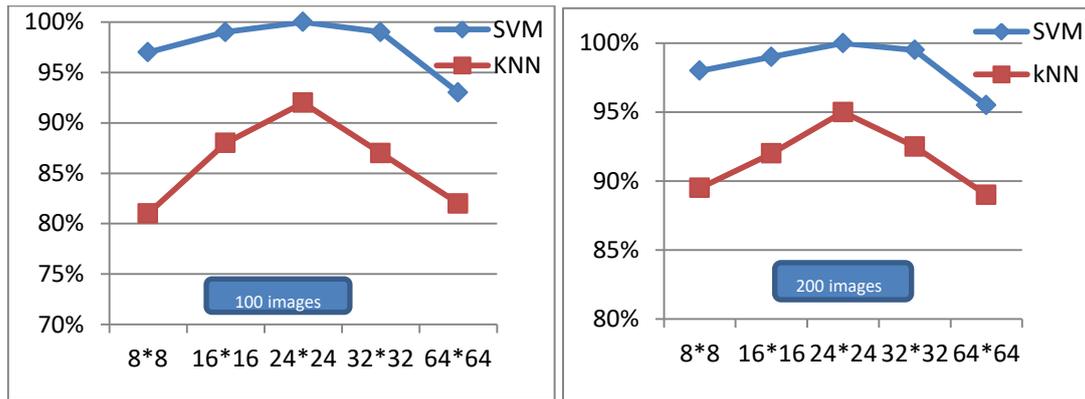


Figure III. 16: Représentation de taux de détection de visage avec masque pour les classifieurs KNN et SVM.

A partir des graphes obtenus, on constate que le taux de détection est égal à 100% pour la cellule 24*24. Le classifieur SVM offre de meilleurs résultats par rapport au classifieur KNN.

Conclusion

Dans ce chapitre, nous avons présenté les tests réalisés sur la méthode de Viola et Jones en vue d'une détection de visage. Nous avons testé l'application sur différentes images positives et négatives. Les tests ont été effectués sur les images originales non traitées.

Ensuite, nous avons utilisé le descripteur HOG pour extraire les caractéristiques des visages afin de constituer une base d'apprentissage pour la phase de classification.

Une série de tests est effectuée sur différentes images contenant des visages avec ou sans masque afin de conclure l'efficacité de notre application tout en constatant que le classifieur KNN est le plus performant pour la classification des visages portant le masque.

Conclusion générale

Dans ce mémoire, nous nous sommes intéressées au problème de détection de visage et le port du masque. Notre travail consiste à la mise au point d'un système destiné à reconnaître un visage masqué ou sans masque en utilisant la méthode des K Plus Proches Voisins et machine à vecteurs de support.

Nous avons donné un aperçu général sur les différentes méthodes de détection de visages puis abordé brièvement quelques descripteurs de fonctionnalités.

Nous avons exploité l'algorithme de Jones et Viola pour la détection de visage, et à l'extraction de la région du visage (région d'intérêt). L'algorithme d'extraction doit être efficace et doit fournir, sans aucune ambiguïté la région du visage dans l'image. Nous avons procédé par la suite à la constitution d'une base de données formée par des images contenant des visages avec ou sans masque. Un ensemble de tests est effectué sur des images en utilisant le descripteur de fonctionnalités HOG (Histogram of Gradient Orientation) qui réalise l'extraction des caractéristiques de chaque visage. Pour la phase de classification, deux classifieurs ont été utilisés : KNN et SVM.

D'après les résultats obtenus, l'algorithme des K Plus Proches Voisins offre une meilleure performance de classification.

Enfin, nous pouvons dire que ce travail reste ouvert pour des travaux de comparaison et/ou d'hybridation avec d'autres méthodes de classifications et d'autres types de descripteurs.

Les perspectives pour ce travail sont nombreuses : dans un premier temps, nous souhaitons développer un module de détection de visage et améliorer l'extraction de ses caractéristiques pour qu'elle soit automatique du fait que notre système ne traite que les vues de face ; qui représentent un domaine fixe. Il serait intéressant de définir un domaine variable portant sur la variété de conditions d'éclairage, de postures et d'expressions faciales. Une troisième perspective serait de rendre le processus de détection un système en temps réel.

Bibliographie

- [1]: S. Meivel, K. Indira Devi, S. Uma Maheswari, J. Vijaya Menaka. Real time data analysis of face mask detection and social distance measurement using Matlab. Department of ECE, M.Kumarasamy College of Engineering, India , Department of ICE, GRG Polytechnic College, India. Article, December 2020.
- [2] : KHAMMARI Mohammed. Détection et suivi de visages en temps Réel sur Flux Vidéo. Thèse Présentée en vue de l'obtention du diplôme De doctorat troisième cycle LMD. Université Mokhtar Badji- Annaba ,2016.
- [3] : BARKI Hicham. Détection et reconnaissance de visage. Pour l'obtention du diplôme magistère .Université Ferhat Abbas-Sétif U.F.A.S.(Algérie).
- [4] :S.Simeon. Analyse de video : création automatique d'un album photo. Article, mars 2010.
- [5]: Wenlong Zheng and Suchendra M. Bhandarkar. Face detection and haking using a boosted adaptative particle filter. Journal of visual communication and Imog Representation, 9-27, 2009.
- [6]:PAUL Viola , MICHEL Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. Article, 2001.
- [7]: Cha Zhang, Zhengyou Zhang, Sven Dickinson, Gerard Medioni. Boosting-Based Face Detection and Adaptation.Article.
- [8] :H.Faouzi. Apprentissage des réseaux D'ondelette beta base sur la théorie des frames : Application a la détection de visages. Mémoire de Magistère, université de Gabes, 2008.
- [9] : Paul Viola et Michael Jones.Robust Real-time Object Detection.IJCV, 2001.
- [10] : BEDOUI Loubna. Authentification de visages par la méthode d'analyse discriminante linéaire de Fischer. Ingénieur d'état en Automatique, Université Mohamed Kheider de Biskra, 2008.
- [11] :CHAA Mourad. SYSTÈME DE RECONNAISSANCE DE PERSONNE PAR DES TECHNIQUES BIOMÉTRIQUES. THESE. Université Ferhat Abbas –Sétif -1-, 2017.
- [12] :BENHALLOU Khadidja. Interface Design for Human Pose Estimation.Mémoire en vue de l'obtention du diplôme de magistère, Université des Sciences et de la technologie d'Oran « Mohamed BOUDIAF », 2015.
- [13]: N.Dalal et B.Triggs. Histograms of Oriented Gradients for Human Detection. INRIA Rhone Alps, Computer Vision and Pattern Recognition Conference, 2005.
- [14] :Ming-Hsuan Yang,Narendra Ahuja, and David Kriegman. A Survey on Face Detection Methods. Article , 1999.

- [15]: David G Lowe. Object recognition from local scale-invariant features. In Computer vision, 1999. The proceedings of the seventh IEEE international conference on, volume 2, pages 1150–1157. Ieee, 1999.
- [16] : David G Lowe, Distinctive image features from scale-invariant key points. International journal of computer vision. 60(2):91–110, 2004.
- [17] : BOHI Amine. Descripteurs de Fourier inspirés de la structure du cortex visuel primaire humain Application à la reconnaissance de navires dans le cadre de la surveillance maritime. Thèse pour obtenir le grade de Docteur en Informatique, Université de Toulon, 2017.
- [18] : Cyrille Migniot. Segmentation de personnes dans les images et les vidéos. Université de Grenoble, 2012.
- [19] : DJEFFAL Abdelhamid. Utilisation des méthodes Support Vector Machine (SVM) dans des l’analyse des bases de données. Thèse présentée pour l’obtention du diplôme de Docteur en sciences, 2012.
- [20] : Saravanan Thirumuruganathan. A detailed introduction to k-nearest neighbor (KNN) algorithm, Article, may 2010.
- [21] : H.Hilali. Application de la classification textuelle pour l’extraction des règles d’association maximales. Thèse de maitrise en informatique, université du Québec à Trois-Rivières, 2009.
- [22] : Henry A. Rowley, Shumeet Baluja, Takeo Kanade . Neural NetworkBased Face Detection . School of Computer Science, Carnegie Mellon University, Pittsburgh, PA 15213, USA, article ,1996.
- [23] : ABDELHAMID Elhassani. Segmentation de la couleur de la peau dans les images faciales : APPROCHE POUR REDUIRE LA FAUSSE DETECTION. Université Sidi Mohammed Ben Abdelah. Article, 2015.
- [24] : Paul A Viola and Michael J Jones. Detecting pedestrians using patterns of motion and appearance in videos, 2007. US Patent
- [25] : Abdallah Benzine, Abdelhak Loukkal, Daro HENG. Rapport pour le projet d’Apprentissage Avancé : Fish Conservancy Challenge. Article.
- [26] : Rainer Lienhart et Jochen Maydt. *An Extended Set of Haar-like Features for Rapid Object Detection*, In [IEEE ICIP](#), 2002
- [27] : Ahonen, T., Hadid, A. et Pietikainen, M. Face description with local binary patterns: Application to face recognition. IEEE transactions on pattern analysis and machine intelligence, 28(12):20372041.Article, 2006.

[28]: BENSENANE Hamdan. Sécurisation des systèmes biométriques. Université des sciences et technologies d'Oran, Mohamad Boudiaf. Thèse en vue de l'obtention du diplôme de doctorat en science, 2018.

[29] : S.PIERARD. Développement d'une méthode de représentation d'images sur base d'un vocabulaire visuel. FACULTÉ POLYTECHNIQUE DE MONS. Travail de fin d'études.