UNIVERSITÉ DE BLIDA 1

Faculté de Technologie

Département d'Électronique

THÈSE DE DOCTORAT

En Électronique

Spécialité : Automatique

Reconnaissance d'objets par l'histogramme de couleurs orientées

Par

Rabah HAMDINI

Devant le jury composé de :

M.Guessoum Abderrezak	Prof.	Université de Blida 1	Président
M.Belkhamsa Noureddine	MC A	Université de Blida 1	Examinateur
M.Ould Zmirli Mohamed	Prof.	Université de Médéa	Examinateur
M.Namane Abderrahmane	Prof.	Université de Blida 1	Directeur de thèse

Blida, février 2021

Dédicace

Le rêve du héros, c'est d'être grand partout et petit chez son père.

Victor Hugo...

À celui qui m'a aidé à découvrir le trésor de 'savoir', mon père M.HAMDINI Farid,

Ce travail est le vôtre . . .

À *ma très chère mère. . .*

Je t'aime beaucoup. . .

À mes futurs enfants inchallah..

Pour que votre papa sera un jour votre héros. . .

Remerciements

Merci à ALLAH avant tout, pour tout, et en tous les cas.

Voilà, je me retrouve enfin à écrire ma page de remerciements, clôturant ainsi cette grande aventure que furent les 4 années de mon doctorat. Lorsque je parcoure une thèse, je lis toujours la page des remerciements, même si force est de constater que je n'y ai jamais trouvé de quoi développer le moindre algorithme ! Mais c'est dans cet unique feuillet que transparaît le plus souvent le dur labeur d'un thésard, ses années de travail, d'échecs parfois, mais aussi de réussites, couronnées par ce dernier, mais ô combien important, diplôme !

Cette thèse de doctorat représente un chapitre important de ma vie. J'ai vécu au cours de ces dernières années des satisfactions et des souffrances, des rencontres et des départs, de beau et de décourageant aussi. Ce parcours, jamais linéaire, marque un changement de mes habitudes et de la façon de mes pensées. Avant d'exposer les résultats de mes travaux, je tiens à remercier toutes les personnes qui ont participé de près ou de loin le déroulement de cette thèse de doctorat.

Je souhaiterais remercier tout d'abord mon directeur de thèse, Pr. Namane Abderrahmane, de m'avoir permis de faire cette thèse. Je vous remercie pour votre soutien, vos conseils et votre investissement. Ces années de thèse ont été très enrichissantes intellectuellement et je suis reconnaissante d'avoir pu bénéficier de vos connaissances et de votre rigueur scientifique.

Je remercie aussi Pr. Guessoum Abderrezak pour avoir accepté la présidence du jury. Je remercie les membres du jury, Pr. Ould Zmirli Mohamed et Dr. Belkhamsa Noureddine pour avoir accepté d'être les examinateurs de mon mémoire et aussi pour leurs remarques et suggestions qui m'ont permis d'avoir de nouvelles perspectives dans mon travail.

Mes sincères remerciements vont surtout à l'inspiratrice de mon parcours universitaire, Dr.Nacira Diffellah. Merci pour votre disponibilité et pour la patience que vous m'avez accordés tout au long de ces années. Je garderais dans mon cœur votre générosité, votre gentillesse, vos précieux conseils et votre compréhension. Vous êtes pour moi un exemple à suivre dans ma vie.

Et pour finir, je vais remercier mes parents sans qui tout cela n'aurait même pas commencé.

Résumé

La reconnaissance des catégories des images est importante pour accéder aux informations visuelles des objets et des types de scènes. Dans cette thèse, nous proposons une nouvelle approche pour la catégorisation d'objet couleur en utilisant les informations puissantes fournies par la couleur. Cette approche est basée sur la combinaison de la constance des couleurs « Gray-Edge », la composante de teinte dans l'espace colorimétrique teinte, saturation et valeur ou HSV (en anglais pour Hue, Saturation, Value) et des idées de cellules et des bacs utilisés dans les descripteurs basés sur les histogrammes des gradients orientés ou HOG (en anglais pour Histograms of Oriented Gradients). Le descripteur orienté proposé profite de l'invariance des teintes contre le changement d'intensité lumineuse, le décalage d'intensité lumineuse et le changement/décalage d'intensité lumineuse, et résout son manque d'invariance contre le changement de couleur de la lumière en utilisant la constance de couleur Gray-Edge. De plus, l'utilisation de cellules et de bacs dans la méthodologie de construction de descripteur proposé a renforcé son invariance aux transformations géométriques et photométriques, et aussi augmente le taux de reconnaissance. Les classificateurs à machine de supports de vecteur ou SVM (en anglais pour Support Vector Machine) et classificateurs de plus proches voisins ou KNN (en anglais pour K-Nearest Neighbours), qui sont deux méthodes de forte classification connues pour leurs flexibilités et leurs pouvoirs de généralisation sont utilisés pour la classification. La méthode proposée est évaluée sur deux ensembles de données accessibles au public, dont la Bibliothèque d'images d'objets Columbia (en anglais pour Columbia Object Image Library coil-100) et la bibliothèque d'images d'objets d'Amsterdam ALOI (en anglais pour Amsterdam Library of Object Images).

Les tests ont montré non seulement les performances exceptionnelles de la méthode proposée dans cette thèse par rapport aux méthodes existantes en termes de taux de reconnaissance, mais aussi sa rapidité et sa capacité à optimiser l'utilisation des espaces mémoire.

Mots clés : Catégorisation d'objet, Reconnaissance d'objet couleur, HSV, Informations visuelles, SVM, K-NN,...

Abstract

Image category recognition is important to access visual information on the level of objects and scene types. In this thesis, we propose a new approach for color object recognition using the powerful information provided by the color. This approach is based on the combination of Gray-Edge color constancy, hue components in HSV (Hue, Saturation, Value) color space and cell and bin ideas used in the HOG (Histograms of Oriented Gradients) descriptors. The proposed oriented descriptor benefits of the invariance of hues against light intensity change, light intensity shift and light intensity change and shift, and solve its missing of invariance against light color change by using Gray-Edge color constancy. Moreover, the use of cells and bins in this proposed descriptor building strengthens its invariance to the geometric and photometric transformation and increases the recognition rate. The classifiers SVM (Support Vector Machine) and KNN (K-Nearest Neighbors) classifiers, which are two strong classification methods known for their flexibility and their power of generalization are used for the training and categorization steps. The proposed method is evaluated on two publicly available datasets including Columbia Object Image Library COIL-100 and The Amsterdam Library of Object Images ALOI. Tests have confirmed not only the exceptional performance of the proposed method compared to existing methods in terms of recognition rate, but also its rapidity and its optimization of using storage and memory spaces.

Key words : Object categorization, Color object recognition, HSV, Visual information, SVM, K-NN, ...

ملـخــص

التعرف على فئات الصور مهم للوصول إلى المعلومات المرئية حول طبيعة الكائنات وأنواع المشاهد. في هذه الأطروحة، نقترح نهجًا جديدًا لتصنيف الأشياء الملونة المصنوعة يدويا في الصور ذات الخلفية الموحدة باستخدام المعلومات القوية التي يوفرها اللون. يعتمد هذا النهج على مزيج من ثبات اللون "الحواف رمادية"، مكون الصبغة في عالم الالوان (صبغة، تشبع، قيمة) وكذلك فكرتي الخلايا والصناديق المستخدمة في واصف الرسوم البيانية للتدرجات الموجهة. يستفيد الواصف الموجه المقترح من ثبات الصبغة في مواجهة تغيرات شدة الضوء، تحول الضوء وتحول شدة الضوء، كما يعالج افتقارها إلى الثبات في مواجهة تغير لون الاضاءة باستخدام ثبات اللون "الحواف رمادية". بالإضافة إلى ذلك، أدى استخدام فكرتي الخلايا والصناديق في منهجية بناء الواصف المقترح إلى تعزيز قدرته على مواجهة التحولات الهندسية والضوئية وبالتالي زيادة معدل التعرف. للقيام بمهمة التصنيف قمنا بالاستعانة بتقنيتي شعاع الدعم الآلي وكي أقرب جار، وهما طريقتا تصنيف قويتان معروفتان بمرونتهما وقدرتهما على التعميم. تم اختبار الطريقة المقترحة على مكتبتي بيانات متاحتين للجمهور، وهما مكتبة كولومبيا ومكتبة أمستردام لصور الاشياء الملونة. أثبتت الاختبارات ليس فقط الأداء الاستثنائي للطريقة المقترحة في هذه الرسالة مقارنة بالطرق الحالية من حيث معدل التعرف، ولكن أيضًا سرعتها وقدرتها على تحسين استخدام مساحات الذاكرة.

الكلمات المفتاحية: الصور، الكائنات، اللون، الصبغة، شعاع الدعم الآلي، ...

Sommaire

Liste des figures

Liste des tableaux

CHAPITRE 1

Introduction générale

«Le secret d'un bon discours, c'est d'avoir une bonne introduction et une bonne conclusion. Ensuite, il faut s'arranger pour que ces deux parties ne soient pas très éloignées l'une de l'autre.»

George Burns

Introduction générale

L'image est aujourd'hui l'un des objets les plus importants de notre société. C'est un support d'information (journaux papier et télévisés, internet), publicitaire, artistique (cinéma, photographie) et social (Facebook, Instagram, etc.). De nombreux travaux visent donc à améliorer et faciliter son accessibilité, que ce soit son acquisition (appareil photo classique, appareil photo numérique, webcam), sa capacité de stockage et d'échange (formats, compression, etc.) ou son édition (des logiciels comme Photoshop ou Gimp peuvent corriger les yeux rouges ou recadrer une photo par exemple).

Ces dernières années, le domaine du traitement et de l'analyse d'images numériques s'est considérablement développé en générant une quantité importante de travaux de recherche. En effet, l'expansion récente des possibilités des ordinateurs a grandement facilité le traitement de masse de l'information numérique. De gigantesques quantités de calculs peuvent désormais être traitées dans des délais toujours plus courts. Des traitements plus complexes sont alors disponibles, ce qui ouvre considérablement les perspectives. Aujourd'hui, les systèmes de vision artificielle sont de plus en plus répandus et des caméras sont installées partout dans notre vie quotidienne.

Les systèmes de vision artificielle sont capables de détecter la présence d'une instance (reconnaissance d'objets) ou d'une classe d'objets dans une image numérique. Ils utilisent souvent l'apprentissage supervisé et ont des applications dans de multiples domaines, comme la recherche d'images par contenu ou la vidéosurveillance et bien d'autres applications... Pour les humains, voir est une tâche innée et on ne mesure souvent pas la difficulté à obtenir artificiellement les mêmes performances. Malgré les progrès de la vision par ordinateur, les systèmes développés sont bien en deçà des performances de l'œil et du cerveau humains.

1.1 Problématique

Les ordinateurs sont devenus omniprésents dans notre vie quotidienne. Ils exécutent des tâches répétitives, gourmandes en données et en calcul de manière plus efficace et précise que les humains. Ces capacités de précision et d'efficacité des ordinateurs ont incité les chercheurs à essayer d'étendre leurs capacités pour effectuer des tâches plus intelligentes telles que l'analyse de scènes visuelles ou de la parole, l'inférence logique et le raisonnement, en bref, des tâches de haut niveau que nous, les humains, exécutons inconsciemment des centaines de fois par jour avec tellement de facilité que nous ne réalisons même pas que nous les exécutons. Prenons l'exemple du système visuel humain, notre vie quotidienne est remplie de milliers d'objets allant des objets artificiels comme les voitures, les vélos, les bâtiments, les tables, les chaises aux objets naturels comme les moutons, les vaches, les arbres, les feuilles, les rochers, les montagnes et humains. Toute classe de données a une énorme variation de sous-classe, par exemple, «voiture» peut être utilisé pour désigner de nombreux véhicules à quatre roues, y compris diverses sous-catégories comme un camion, un camion, un bus et un minibus. Le type, la couleur et la perspective exacts d'une voiture ne sont pas pertinents pour décider qu'un objet est une voiture. De même, nous sommes en mesure de détecter des personnes dans une grande variété de conditions, indépendamment de la couleur ou du type de vêtement, de la pose, de l'apparence, des occlusions partielles, de l'éclairage ou de l'encombrement de l'arrière-plan. Les ordinateurs sont maintenant loin derrière les humains pour faire de telles analyses et inférences.

Un adulte peut généralement reconnaître plus de 10000 de catégories d'objets et un nombre beaucoup plus important d'instances d'objets [9]. Le processus de reconnaissance est souvent rapide, sans effort et robuste aux changements de point de vue, de luminosité et à des occultations d'une partie de l'objet, etc., l'opération d'apprentissage d'une nouvelle catégorie d'objets peut être effectuée à partir d'un petit nombre d'images de cet objet, et l'apprentissage d'une nouvelle instance d'ob-

jet peut être effectué à partir d'une seule image de cet objet.

Il est très utile de programmer des ordinateurs pour qu'ils soient également capables de reconnaître de la même manière des catégories d'objets ou des instances d'objets. Ce domaine de recherche est appelé reconnaissance visuelle par ordinateur. Après une phase d'apprentissage, où certaines images de la catégorie d'objet ou de l'instance sont fournies, l'ordinateur prédit si une image contient ou non cette catégorie d'objet (reconnaissance de catégorie d'objet) ou cet objet spécifique (reconnaissance d'instance d'objet).

Pour y parvenir, l'un des objectifs des chercheurs travaillant dans le domaine de la vision par ordinateur et de l'intelligence artificielle a été de permettre aux ordinateurs d'analyser et d'interpréter des images ou des vidéos. L'une des tâches principales est la détection de différentes classes d'objets dans les images et les vidéos. Une telle capacité aurait de nombreuses applications, par exemple dans l'interaction homme-machine, la robotique, l'analyse automatique du contenu multimédia numérique personnel ou professionnel, les processus de fabrication automatisés et les véhicules autonomes intelligents. Mon travail de thèse se déroule dans ce cadre complexe. Il s'agit de développer un système capable de reconnaître les instances d'objets à partir d'images provenant d'une simple caméra.

Les descriptions d'images et les fonctions de classification associées doivent respecter un cahier des charges lié à la nature des images considérées et au fonctionnement du système de reconnaissance d'objets.

1.1.1 Cahier des charges lié à la nature des images

— **Pouvoir discriminant**

L'utilisation de fonctions de comparaison de descripteurs d'image doit permettre de distinguer le cas où l'image requête et l'une des images candidates qui forment un couple d'images différentes du cas où elles forment un couple d'images similaires.

— **Invariance aux translations, rotations de l'objet**

Les images considérées contiennent un seul objet qui n'occupe pas toujours la même position spatiale. Dans le cadre de notre travail, Il est donc essentiel que les descripteurs soient insensibles aux traductions d'un objet d'une image à une autre ainsi qu'aux rotations dans un plan perpendiculaire à l'axe optique de la caméra. Finalement, La restriction que nous nous imposons aux mouvements possibles d'un objet entre deux acquisitions d'images est que chaque élément de surface de l'objet observé dans l'une des images doit apparaître dans l'autre image qui représente le même objet.

— **Invariance à la résolution spatiale de l'image**

Les objets peuvent être placés à différentes distances de la caméra. Cette caméra peut être équipée d'objectifs de focales différentes, dans ce cas, la résolution spatiale des images considérées peut ne pas être constante.

— **Invariance aux Modifications d'illumination**

Les changements d'éclairage modifient également l'apparence des objets. Un changement uniforme sur toute l'image peut être annulé en rectifiant la luminosité et le contraste aux valeurs standard, mais dans un cas réel, les changements ne sont pas uniformes, par exemple lorsque le soleil se déplace, lorsque les rideaux s'ouvrent ou lorsqu'une lampe s'allume.

— **Invariance aux Déformations**

Certains objets sont déformables, ce qui génère d'importants changements d'aspect entre deux états. On peut considérer de petites apparitions / disparitions (comme une antenne radio rétractable sur une voiture), des déformations articulées, qui subissent une marionnette articulée (chaque partie non articulée garde la même apparence, mais l'apparence général change) ou des déformations élastiques (visage qui fait une grimace).

— **Invariance aux Fonds chargés**

La position des objets dans les images est souvent déterminée par un

rectangle de délimitation, aussi appelé région d'intérêt (ROI(en anglais pour Region of Interest)). Ce rectangle contient l'ensemble de l'objet, ainsi que les pixels qui proviennent du fond. Lorsque cette toile de fond change, le contenu du rectangle englobant change également, ce qui est une nouvelle source de variabilité.

— **Invariance à l'Occultation**

Les occultations d'une partie des objets sont une grande source de variabilité dans l'apparence des objets, car les apparences typiques d'une partie des objets sont remplacées par l'objet occultant.

1.1.2 Cahier des charges lié à l'exploitation d'un système de reconnaissance d'objets

— **Temps de réponse à une requête**

L'utilisateur d'un système de reconnaissance d'objets sélectionne une image de requête qui contient un objet qu'il recherche dans une base de données d'images candidates. Cette base de données est constituée d'images à partir desquelles les pixels de l'objet ont été préalablement extraits afin de calculer leurs descripteurs. Le calcul des descripteurs des images candidates est effectué hors ligne de sorte que pour chaque recherche, seul le descripteur de l'image de requête est calculé en ligne. Le temps nécessaire au calcul des descripteurs d'images n'est donc pas un facteur crucial à minimiser à tout prix. En revanche, la comparaison des descripteurs d'images de requête et de ceux des nombreuses images candidates se fait en ligne. Le temps nécessaire à cette comparaison influence le système de reconnaissance d'objets.
Si l'apparence des objets dans les images ne changeait jamais, il serait très facile de les reconnaître. Il suffirait d'utiliser un comparateur pixel à pixel dans une région de l'image à une image de référence de catégorie connue pour savoir s'il peut être du même objet (différence nulle) ou non (différence non

nulle). Mais l'apparence des objets sur les images varie, et cette variation rend la tâche difficile.

— **Place mémoire occupée par les descripteurs**

Lorsque les descripteurs des images candidates sont évalués hors ligne, ils sont stockés sur le disque prenant en charge le système d'indexation. De plus, le temps nécessaire pour évaluer la fonction de comparaison de deux descripteurs est souvent directement lié à l'espace mémoire occupé par ces descripteurs. Il est donc intéressant de privilégier les descripteurs occupant le moins d'espace mémoire possible.

1.2 Contributions

Nos contributions sont établies à différents niveaux :

— Dans le cadre de l'étude des systèmes de reconnaissance d'objets, nous proposons une taxonomie permettant de révéler les différents avantages et inconvénients des différentes méthodes existantes.

— Sur un plan plus théorique, nous proposons une chaîne perceptive et optimale (en termes de temps de calcul et d'allocation mémoire) pour la reconnaissance d'objets, basée sur la combinaison de la constance des couleurs, des composantes de teinte dans l'espace colorimétrique HSV (Hue, Saturation, Value) et l'idée de cellule et de bin utilisées dans les descripteurs HOG (Histogram of Oriented Gradient). Le descripteur orienté proposé tire parti de l'invariance de teinte contre le changement d'intensité lumineuse et le décalage d'intensité lumineuse, et résout son manque d'invariance contre le changement de couleur de la lumière en utilisant la constance de couleur Gray-Edge. De plus, l'utilisation de l'idée des cellules dans la construction de descripteur proposé renforce son invariance face aux transformations géométriques et photométriques et augmente le taux de reconnaissance.

— D'un point de vue expérimental, nous caractérisons l'influence de chaque

étape de notre algorithme sur ces performances et son efficacité. Les résultats obtenus montrent que notre système proposé peut être à la base d'une application qui remplacera l'utilisation des codes barres dans les magasins commerciaux (centres commerciaux, pharmacies, etc.).

1.3 Plan de la thèse

Le plan de la thèse suit la construction du système développé, à savoir :

— Le deuxième chapitre de cette thèse est consacré à une étude bibliographique sur les différents éléments de base nécessaires pour définir les couleurs des surfaces et leurs apparences. Dans cette partie, nous nous intéressons aux différents phénomènes physiques qui sont à l'origine d'une couleur à savoir l'absorption, la dispersion, la diffusion, l'interférence et la diffraction. Nous présentons également les méthodes de contrôle de la couleur ainsi que les techniques qui permettent de passer de la couleur perçue à une couleur mesurée que l'on peut communiquer et reproduire facilement.

— Le chapitre trois porte sur une revue bibliographique complète des systèmes de vision artificielle dans le contexte de la reconnaissance d'objets. Dans ce chapitre, nous fournissons un historique général avant de présenter la structure commune des algorithmes. Cette structure implique des descripteurs d'images qui sont présentés et classés par grandes familles. Enfin, ce chapitre se termine par une analyse statistique traversant les applications finales et les descripteurs d'images afin d'identifier les plus courantes.

— Le chapitre quatre décrit la méthode de construction du descripteur proposé pour la reconnaissance d'objets. Ce descripteur est d'abord testé sur les images de la bibliothèque des images d'objets Columbia (COIL (en anglais pour Columbia Object Image Library)), puis il est testé sur les images de la bibliothèque des images d'objets Amsterdam (ALOI (en anglais pour Amsterdam Object Image Library)).

— Dans la conclusion générale, la méthode développée et les résultats obtenus sont analysés et critiqués. Enfin, nous proposons nos perspectives qui, nous l'espérons, nous permettront d'améliorer et de compléter ce travail.

CHAPITRE 2

Couleur

« La vraie philosophie est de voir les choses telles qu'elles sont.»

George Louis Buffon

Résumé

Dans ce chapitre nous présentons la perception et la mesure des couleurs. Nous détaillerons notamment le phénomène de constance chromatique, à savoir notre capacité à identifier visuellement la couleur d'un objet. Et ce quel que soit l'éclairage utilisé. Par la suite, nous décrirons la formation d'une image numérique couleur et plus particulièrement les conséquences de modifications des conditions d'éclairage sur les couleurs des objets.

Contents

Les couleurs des matériaux et leurs apparences visuelles ont fait l'objet d'un grand intérêt et reçoivent une importante attention dans diverses activités industrielles et commerciales. La notion de couleur est liée à la perception et à l'interprétation subjective de la personne. La colorimétrie est un ensemble de mesures, des outils et de méthodes qui permettent de repérer la couleur. Elle s'intéresse à déterminer les trois paramètres caractérisant une couleur isolée, à savoir : la clarté, la teinte et la saturation. Elle permet de créer à l'aide des relations entre la lumière et les matériaux colorés, un espace colorimétrique muni d'une métrique, afin de pouvoir repérer et reproduire les couleurs.

2.1 Définition de la couleur

La couleur est une classe d'apparence comme la forme, la brillance, la transparence et la texture. Chaque classe est définie par un caractère qualitatif présent dans la perception [104]. Viénot définit la couleur comme étant un attribut perceptif, subjectif, élaboré dans notre système visuel à partir de la lumière renvoyée par l'objet et son environnement [114]. La notion de la couleur est donc indissociable de la vision. Elle est conçue dans le système visuel composé de l'œil et du cortex visuel. De fait, la couleur n'a pas de réalité physique. C'est le résultat d'une interprétation subjective des signaux visuels opérés par le cortex : c'est une sensation [49].

2.2 Perception de la couleur

D'une façon générale, la perception de la couleur d'un objet fait intervenir trois éléments (figure2.1) :

— L'interaction lumière/matière, qui forme le stimulus de couleur.

— L'œil, qui focalise le stimulus de couleur et le projette sur sa partie photosensible, la rétine.

— Le système d'interprétation, qui permet d'identifier une couleur grâce à ses

différents attributs.

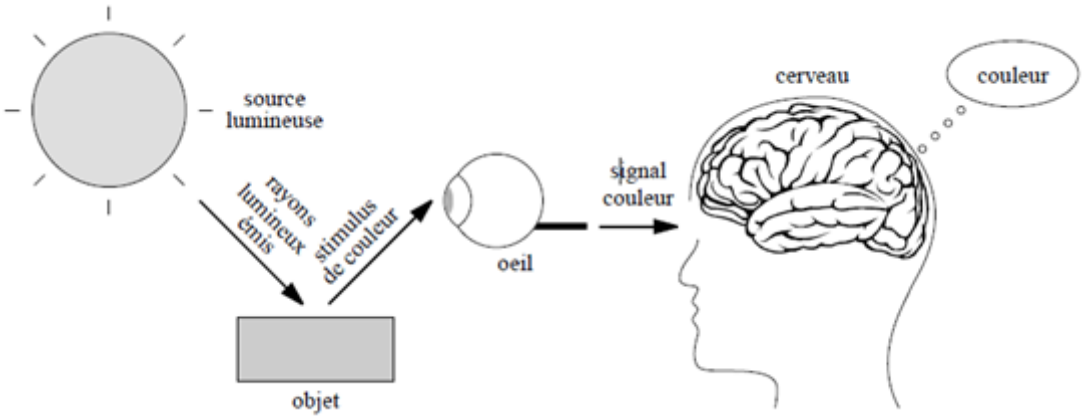


FIGURE 2.1 – Perception humaine de la couleur (source : [72]).

Dans cette section, nous allons voir comment la lumière interagit avec les surfaces pour produire les ondes électromagnétique qui est à la base de la vision. Nous aurons aussi un aperçu du système visuel humain qui capte ce signal.

2.2.1 Stimulus de la couleur : interaction lumière/matière

Le stimulus de la couleur émis par une surface élémentaire d'un objet dépend de la source lumineuse qui éclaire l'objet ainsi que du matériau qui compose cette surface élémentaire.

2.2.1.1 Lumière

On appelle lumière la partie visible d'un vaste groupe de radiations qui vont des rayons cosmiques aux ondes radar. Toutes ces ondes sont de même nature (électromagnétiques) et se déplacent dans le vide à la même vitesse : environ 300000 km/s. Elles diffèrent les unes des autres selon leurs longueurs d'onde et l'énergie qu'elles transportent (figure2.2).

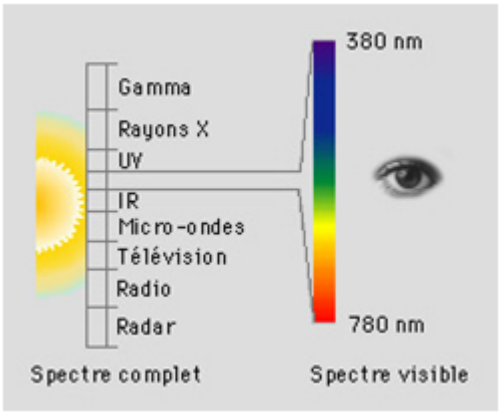


FIGURE 2.2 – Spectre complet et spectre visible de longueurs d'onde (source : [12]).

La lumière consiste en un ensemble de particules élémentaires de masse nulle appelées photons qui se comportent comme des ondes dans une certaine mesure, mais aussi comme des particules [10]. Les photons irradient à partir de leur source vers l'extérieur et traversent l'air suivant une trajectoire rectiligne. La quantité de photons qui arrive sur une surface par unité de temps est appelée luminance. Un photon se déplace à la vitesse c, qui dépend du milieu, l'onde est associée à une fréquence f. La fréquence et la vitesse d'un photon déterminent sa longueur d'onde λ :

$$\lambda f = c \tag{2.1}$$

L'énergie E de chaque photon est reliée à la fréquence par l'équation de Planck :

$$E = hf \tag{2.2}$$

avec $h \approx 6.626 * 10^{-34}$ joules.seconde, la constante de Planck.

2.2.1.2 Types de sources lumineuses

Les sources d'émission de photons correspondent généralement à des corps chauds comme le soleil, du feu ou l'élément d'une lampe à incandescence.
Dans la matière, la production d'énergie lumineuse se fait grâce aux électrons. Ces derniers occupent des orbitales très précises au sein de l'atome mais si on excite

l'atome par un apport d'énergie, par exemple de la chaleur, les électrons absorbent cette énergie et sautent sur des orbitales supérieures.

Les positions d'excitation sont très instables et dès que l'apport d'énergie cesse les électrons ont tendance à reprendre spontanément leur position d'origine en restituant leur surplus d'énergie sous la forme de photons. L'énergie des photons ainsi émise, donc leurs longueurs d'onde, varie en fonction de l'importance du (saut) effectué par l'électron pour rejoindre son orbitale stable. Comme chaque électron possède de nombreuses orbitales d'excitation, un même atome peut émettre des photons de longueurs d'onde différentes. À quelques exceptions près, les rayons lumineux ne sont pas constitués de photons de même longueur d'onde. Une source lumineuse émet généralement une quantité de photons définie par tranche de longueur d'onde. Ces quantités prises sur la totalité du spectre visible forment la distribution de puissance spectrale (DPS) de la source de lumière qui sert à évaluer l'efficacité lumineuse et la chromaticité d'une source de lumière.

Une autre caractéristique d'une source lumineuse est sa température de couleur, c'est-à-dire la température à laquelle il faudrait porter un corps noir pour obtenir une répartition spectrale d'énergie identique à celle de la source. Elle est exprimée en Kelvin (K). Certaines sources ont été normalisées par la CIE (Commission internationale de l'Éclairage), sous le nom d'illuminant standard, parce qu'elles correspondent à des conditions d'illumination courantes ou intéressantes [3]. Il faut faire attention à la distinction entre illuminant et source. En effet la notion de source fait référence à un objet physique qui émet de la lumière (une lampe ou le soleil), alors que le terme illuminant fait référence à une répartition spectrale d'énergie particulière, non nécessairement obtenue par une source. Un illuminant normalisé se caractérise par sa répartition spectrale relative d'énergie, notée $E(\lambda)$. Il s'agit d'une normalisation à 100 de la répartition spectrale d'énergie pour une longueur d'onde particulière, en général $560nm$ pour la plupart des illuminant (voir figure 2.3). Les principaux illuminant normalisés de la CIE sont référencés ci-dessous dans la figure 2.3 pour la répartition spectrale d'énergie de certains d'entre eux :

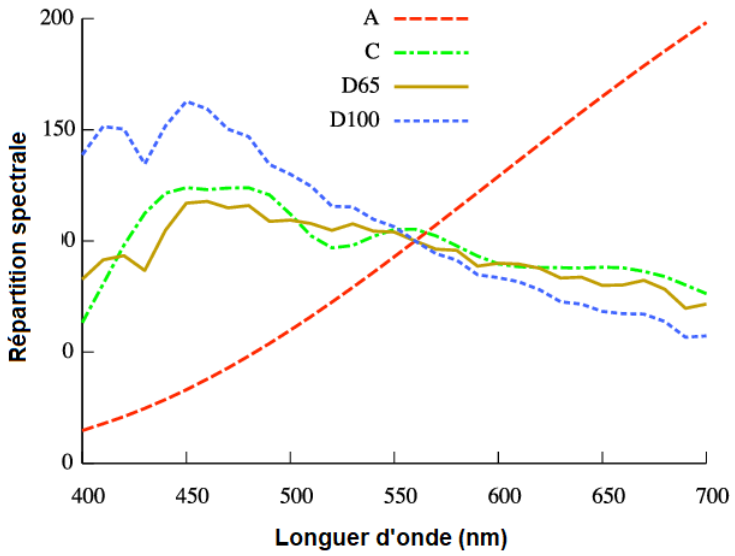


FIGURE 2.3 – Répartition spectrale relative d'énergie des illuminant standards A, C D_{65} et D_{100} (source : [25]).

— **Illuminant** A : lumière émise par un corps noir porté à la température de 2856 K. Il est destiné à représenter une source lumineuse produite par une lampe à filament de tungstène.

— **Illuminants** B**,** C **et** D : ils simulent la lumière du jour. L'illuminant B représente la lumière du soleil à midi avec une température de couleur de $4874K$. Il n'est plus en usage actuellement. L'illuminant C représente la lumière moyenne du jour avec une température de couleur de $6774K$. L'illuminant D représente différentes lumières naturelles du jour : l'indice associé à l'illuminant est sa température de couleur et représente un moment de la journée particulier. L'illuminant D_{65} $(6500K)$ est le plus fréquemment utilisé, car il correspond à une lumière naturelle en plein jour en zone tempérée. C'est un réglage standard dans l'industrie du cinéma et la production audiovisuelle. Les illuminants D_{50}, D_{55}, D_{75} sont aussi très utilisés. Même si ces illuminants sont facilement caractérisables mathématiquement, leur inconvénient majeur reste qu'ils sont difficiles à reproduire par une source artificielle.

— **Illuminant** E : illuminant équi-énergétique (lumière d'énergie constante). Il ne correspond à aucune source réelle et ne présente qu'un intérêt théorique.

— **Illuminant** F : la série d'illuminant F (notés de $F1$ à $F12$) correspond à la

lumière émise par différentes lampes fluorescentes. Les plus utilisés sont les illuminant $F2$ (lampe fluorescente standard), $F7$ (lampe fluorescente à bandes larges) et $F11$ (lampe fluorescente à trois bandes étroites).

2.2.1.3 Matériau

On peut distinguer deux sortes de matériaux : des matériaux opaques (métal, bois, etc.) ou des matériaux transparents et translucides (gélatine, verre, eau....). Les premiers renvoient la lumière vers l'observateur par réflexion et les seconds par transmission [11].

Un objet apparaît vert sous une lumière blanche, s'il réfléchit principalement les radiations de longueur d'ondes moyennes ($500nm$ à $570nm$) et s'il absorbe principalement les radiations de courtes et de grandes longueurs d'onde ($380nm$ à $500nm$ et $570nm$ à $780nm$). Le rapport du flux lumineux réfléchi par l'objet au flux réfléchi dans les mêmes conditions par le diffuseur parfait en réflexion définit ce que l'on appelle «facteur de réflectance» [50] :

$$R(\lambda) = \frac{L_r(\lambda)}{L_{dr}(\lambda)}(\%) \tag{2.3}$$

L_r et L_{dr} sont la quantité d'énergies lumineuses réfléchies respectivement par l'objet et le diffuseur parfait par réflexion. Ce facteur est égal au rapport de la partie de flux total réfléchie par diffusion au flux incident dans le cas d'une mesure avec un spectrophotomètre à sphère intégrante (CIE. 1978). L'ensemble des valeurs du facteur de réflexion diffusée pour différentes longueurs d'onde λ du spectre visible constitue un spectre de réflexion diffusée. Ceci permet de caractériser un objet coloré par sa courbe de réflectance.

2.2.1.4 Interaction lumière / matière

Les objets colorés ne génèrent pas de lumière. Si nous voyons leur couleur, c'est que les rayons lumineux entrent en contact avec la matière composant l'objet, et que

celle-ci en absorbe une partie, puis réfléchit et/ou transmet l'autre partie. Ce sont les rayons lumineux réfléchis que nous percevons. Ainsi un objet blanc réfléchit et transmet dans toutes les directions toutes les radiations de la lumière, alors qu'un objet noir absorbe toutes les radiations de la lumière. L'interaction lumière/matière produit trois phénomènes distincts : la réflexion, l'absorption et la transmission (figure 2.4).

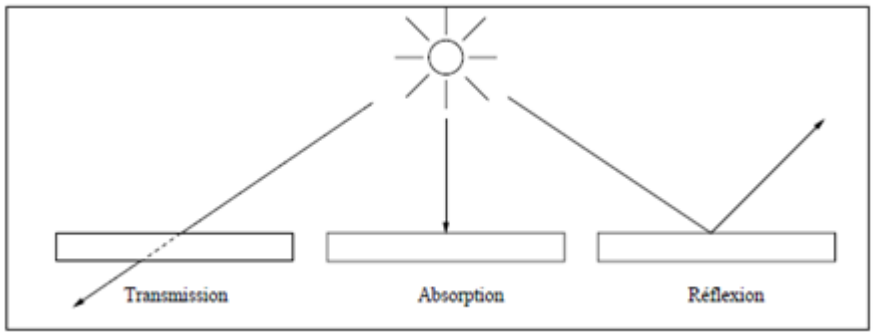


FIGURE 2.4 – Différents types d'interaction lumière/matière

— **Réflexion**

Par opposition à la transmission et à l'absorption, la réflexion est l'ensemble des rayons lumineux qui émergent de l'objet dans la zone où ils sont entrés en contact avec lui. La réflexion peut être spéculaire ou diffuse selon la nature de l'objet. Sur une surface irrégulière, la réflexion se fait dans plusieurs directions, elle est alors dite diffuse. Sur une surface plane, un rayon incident ne produit qu'un unique rayon réfléchi, la réflexion est ainsi dite spéculaire. La plupart des objets ne réfléchissent qu'une partie du rayon incident, car une partie pénètre dans le matériau. Si la lumière est réfléchie totalement, le matériau est dit opaque, si elle traverse totalement, le matériau est dit transparent, et si une partie est réfléchie tandis qu'une autre traverse, le matériau est dit translucide. Après la pénétration, la lumière peut être ensuite absorbée ou transmise par le matériau.

— **Absorption et transmission**

Lorsqu'elle pénètre dans le matériau, la lumière change de milieu de propagation, elle est déviée, c'est le phénomène de réfraction. Dans le matériau, la lumière rencontre des pigments, particules qui vont lui donner sa couleur en absorbant une

partie des radiations du rayon et en diffusant et transmettant le reste. Ainsi la couleur d'un objet est donnée par l'ensemble des radiations qui n'a pas été absorbé par cet objet.

— **Réflectance et transmittance**

Un matériau réfléchissant de la lumière est caractérisé par sa réflectance spectrale, c'est-à-dire le rapport entre l'intensité de la lumière incidente et celle de la lumière réfléchie en fonction de la longueur d'onde :

$$Rt(\lambda) = \frac{I_{réfléchie}(\lambda)}{I_{incidente}(\lambda)} \tag{2.4}$$

Les pigments colorés sont un exemple de matériau réfléchissant. Utilisés en peinture, leur mélange permet de générer de nouvelles couleurs.
De la même façon, un matériau transmettant de la lumière est caractérisé par sa transmittance spectrale, c'est-à-dire le rapport entre l'intensité de la lumière incidente et celle de la lumière transmise en fonction de la longueur d'onde :

$$T(\lambda) = \frac{I_{transmise}(\lambda)}{I_{incidente}(\lambda)} \tag{2.5}$$

Les filtres, par exemple, sont des objets transparents permettant de ne transmettre qu'une partie du spectre de la lumière incidente par absorption sélective.

2.2.2 Œil

L'œil est le capteur visuel naturel. C'est sur la base de notre compréhension de son fonctionnement qu'ont été élaborés les capteurs artificiels tels que les caméras. Il est donc utile de rappeler les caractéristiques morphologiques et physiologiques générales du système visuel humain et en particulier celles de l'œil.
L'œil constitue l'interface physique du système visuel humain. C'est un capteur qui convertit l'information électromagnétique en activité neuronale.
Sur son chemin vers la rétine, le photon rencontre premièrement la cornée (figure 2.5), fine pellicule transparente protégeant l'humeur aqueuse. En plus de sa fonc-

tion protectrice, elle a le rôle de lentille dans le système optique. Ensuite, la lumière incidente traverse le système optique principal formé par l'iris, la pupille et le cristallin. La pupille contrôle la quantité de lumière entrante dans l'œil en réduisant ou augmentant sa taille à l'aide de l'iris. Le cristallin, situé derrière la pupille, reçoit le signal et le projette sur la rétine qui se trouve au fond de l'œil. En faisant varier sa courbure, le cristallin fait converger les rayons lumineux de manière à obtenir une image nette sur la rétine.

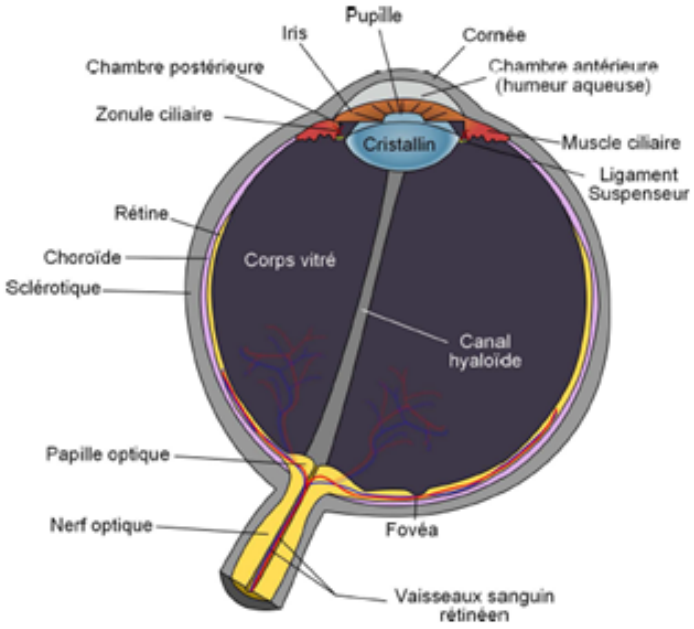


FIGURE 2.5 – Coupe de l'œil humain (source : wikipedia commons).

La rétine est l'organe photosensible du système visuel. Elle est composée de centaines de millions de cellules nerveuses : les cônes et les bâtonnets. Le rôle de ces cellules est capital. Elles permettent de voir les détails, les lumières, les couleurs et les formes. Les cônes et les bâtonnets sont les cellules photo-réceptrices. Ce sont ces cellules qui transforment le signal électromagnétique en influx nerveux et transmettent au cerveau l'information visuelle sous la forme d'un signal neuronal. Il y a beaucoup plus de bâtonnets (130 millions) que de cônes ($6-7$ millions). Le diamètre des cônes est beaucoup plus petit que celui des bâtonnets. Plus on s'éloigne de la partie centrale de la rétine, plus les cônes se font rares et plus leur diamètre augmente (figure 2.6).

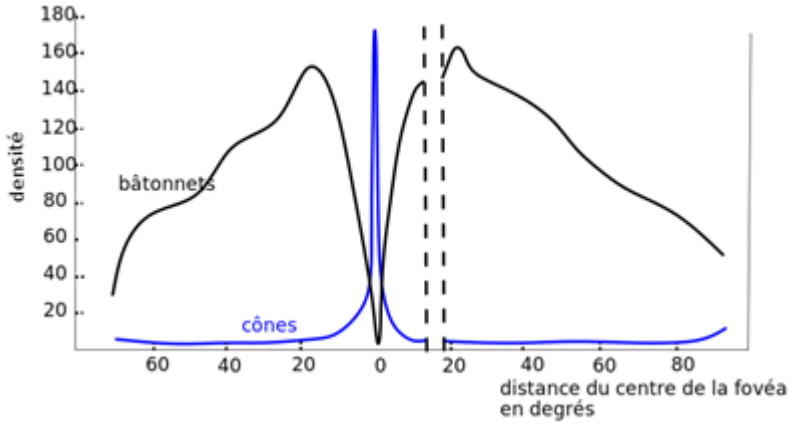


FIGURE 2.6 – Répartition des cônes et des bâtonnets dans la rétine (source : wikipedia commons).

Les cônes ont besoin de plus de lumière que les bâtonnets pour être excités. Ils réagissent plus en éclairage diurne qu'en éclairage nocturne. Les bâtonnets eux, assurent la vision nocturne.

Il existe trois sortes de cônes (figure 2.7) qui réagissent à des longueurs d'onde différentes et que l'on dénomme short, médium et long (SML).

Ces cellules sont à la base de la trichromaticité humaine [89]. Elles sont donc responsables de la vision des couleurs.

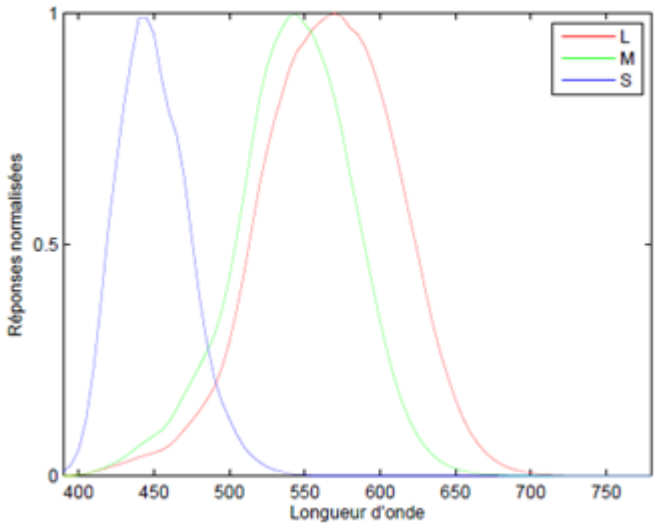


FIGURE 2.7 – Réponse spectrale des trois types de cônes estimés par Stockman et Sharpe (source : [107]).

2.2.3 Système d'interprétation de stimulus de la couleur

Le LGN (Corps Géniculé Latéral) reçoit les informations directement des cellules ganglionnaires rétiniennes via le nerf optique. Les signaux y sont codés de manière antagoniste, c'est-à-dire sous forme d'un signal achromatique noir et blanc, et de deux signaux chromatiques d'opposition rouge-vert et jaune-bleu. Les cellules du LGN vont ensuite rejoindre leur cible principale : le cortex visuel où s'effectue l'interprétation de la couleur (figure 2.8).

Le cortex visuel est divisé en deux zones : le cortex visuel primaire qui est une projection directe de la rétine et effectue un traitement de bas niveau sur les données visuelles (identification des lignes, des couleurs, des sens de déplacements) et un cortex visuel secondaire qui rassemble ces éléments pour obtenir des objets ayant une forme, une couleur et un mouvement précis.

Le cerveau est le système d'interprétation de l'homme. La perception de la couleur pour chaque être humain dépend du signal couleur parvenant à son cortex cérébral (aspects physiques et physiologiques) et d'autre part d'aspects psychologiques, c'est-à-dire la connaissance à priori de son environnement. Les mécanismes neurophysiologiques liés à cette interprétation sont relativement complexes et encore assez mal connus.

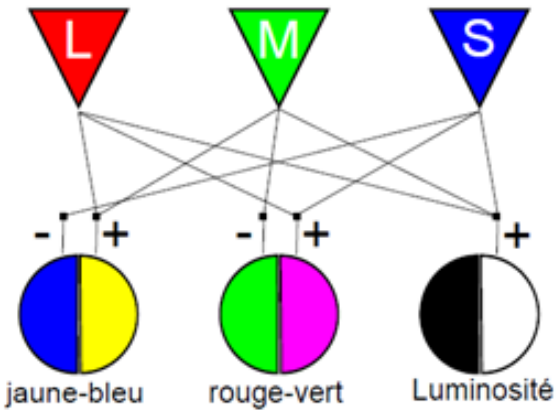


FIGURE 2.8 – Codage antagoniste des couleurs (source : wikipedia commons).

2.3 Mesure de la couleur

La Commission internationale de l'éclairage s'intéressa dès les années 1920 à l'élaboration d'un système de modélisation colorimétrique de l'œil humain.

2.3.1 Du stimulus couleur à ses composantes trichromatiques

Dès 1802, Thomas Young $(1773-1829)$ émit l'hypothèse de la trivariance visuelle. Il voyait dans le système visuel une composition de trois récepteurs sélectifs en longueur d'onde : l'un sensible au rouge, l'autre au vert, et le dernier en bleu. En 1855, James Clerk Maxwell $(1831-1879)$ effectua les premières mesures visuelles dans l'objectif de tester l'hypothèse portant sur la trivariance. Ses résultats permirent de la vérifier et d'unifier les différentes théories de l'époque sur cette question. En 1931, la CIE reprit les travaux menés par William David Wright [131, 132] et John Guild [40], et définit un observateur colorimétrique de référence dont le système visuel était composé de trois filtres rétiniens, sous forme de trois fonctions primaires.

2.3.2 Espace de couleur RGB

En 1931, en se basant sur les travaux de Wright et Guild [41, 132], la CIE définit les primaires notées ici R_P, G_P et B_P, correspondant aux stimuli rouge, vert, et bleu et de longueur d'onde respective $700.0nm$, $546.1nm$, et $435.8nm$ [87]. Ces primaires sont considérés comme des stimuli de référence dont le mélange unitaire doit reproduire l'impression visuelle du spectre éco énergétique (illuminant E). Les valeurs unitaires associées à chaque primaire sont alors ajustées pour que les composantes trichromatiques des primaires, associées au spectre éco énergétique, soient toutes égales : les ratios sont $1.0000 : 4.5907 : 0.0601$, pour $R : G : B$ respectivement.
Les trois fonctions colorimétriques (ou CMFs en anglais pour Color Matching Functions) associées à ces primaires : $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ et $\bar{b}(\lambda)$, sont les descriptions numériques normalisées de la réponse chromatique de l'observateur, lors de l'expérience

d'appariement. Ces courbes sont représentées dans la figure 2.9. Elles sont normalisées avec les mêmes ratios que les primaires : 1 : 4.5907 : 0.0601 pour $\bar{r} : \bar{g} : \bar{b}$ [3].

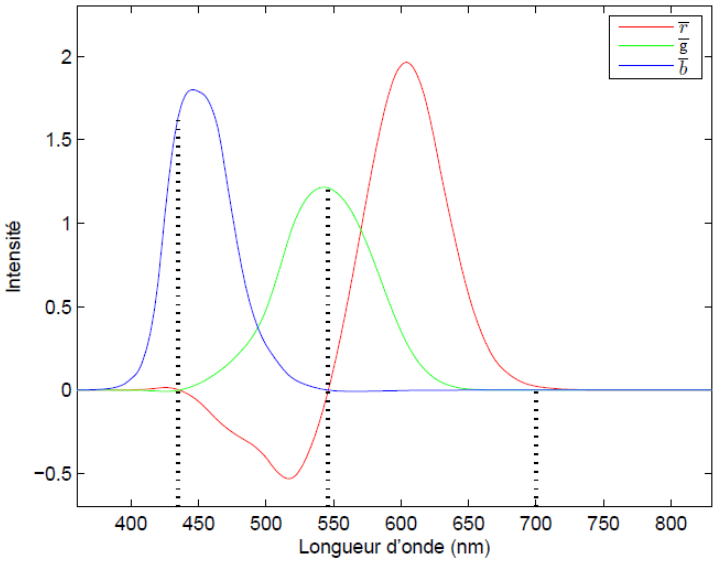


FIGURE 2.9 – Fonctions colorimétriques de la CIE (1931) $\vec{r}$,$\vec{g}$ et $\vec{b}$ (source : [3])

Les composantes chromatiques d'un stimulus de couleur C avec une distribution $C(\lambda)$ se calculent à l'aide des fonctions colorimétriques :

$$\begin{cases} R = \int_{380}^{780} C(\lambda)\,\bar{r}(\lambda)\,d\lambda \\ G = \int_{380}^{780} C(\lambda)\,\bar{g}(\lambda)\,d\lambda \\ B = \int_{380}^{780} C(\lambda)\,\bar{b}(\lambda)\,d\lambda \end{cases} \tag{2.6}$$

Dans la pratique, les fonctions colorimétriques relèvent des résultats expérimentaux et les données sont disponibles sous forme discrète et non continues, avec un pas constant de longueur d'onde, et le calcul des composantes trichromatiques se fait en remplaçant l'intégrale de l'équation 2.6 par une somme.

Les composantes trichromatiques sont liées à la luminance du stimulus. Deux stimuli de couleur peuvent ainsi posséder le même caractère chromatique ou chrominance, mais avoir des composantes trichromatiques différentes, car leur luminance est différente. Pour caractériser la chrominance, il faut utiliser les coordonnées chromatiques représentant les composantes chromatiques normalisées par leur lumi-

nance :

$$\begin{cases} r = \dfrac{R}{R+G+B} \\ g = \dfrac{G}{R+G+B} \\ b = \dfrac{B}{R+G+B} \end{cases} \tag{2.7}$$

L'espace obtenu avec ces coordonnées chromatiques est l'espace $(r(\lambda), g(\lambda), b(\lambda))$ ou l'espace (R,G,B) normalisé. $r(\lambda)$, $g(\lambda)$ et $b(\lambda)$ sont les composantes trichromatiques spectrales.

$$\begin{cases} r(\lambda) = \dfrac{\bar{r}(\lambda)}{\bar{r}(\lambda)+\bar{g}(\lambda)+\bar{b}(\lambda)} \\ g(\lambda) = \dfrac{\bar{v}(\lambda)}{\bar{r}(\lambda)+\bar{g}(\lambda)+\bar{b}(\lambda)} \\ b(\lambda) = \dfrac{\bar{b}(\lambda)}{\bar{r}(\lambda)+\bar{g}(\lambda)+\bar{b}(\lambda)} \end{cases} \tag{2.8}$$

La transformation définie par l'équation 2.7 correspond à la projection de la couleur sur le plan normal à l'axe achromatique d'équation : $r+g+b=1$, où la luminosité est constante. Les couleurs ayant des composantes positives dans ce plan forment un triangle équilatéral dont les sommets sont les trois primaires, appelé triangle de Maxwell. Tenant compte du fait que $r+g+b=1$, et que donc deux coordonnées suffisent pour caractériser une couleur, Wright et Guild ont proposé une représentation dans le plan (r, g) appelé diagramme de chromaticité (b peut être déduit avec $b=1-r-g$) [41, 132]. La courbe formée par les coordonnées trichromatiques spectrales dans ce diagramme est appelée spectre locus. Toutes les couleurs du spectre sont contenues dans la région fermée délimitée par le spectre locus et la ligne qui joint ses deux extrêmes ou droite des pourpres. Celui-ci ainsi que la projection du triangle de Maxwell dans ce plan sont représentés figure 2.10. Il faut remarquer que le triangle de Maxwell ne permet pas de représenter toutes les couleurs du spectre, car beaucoup de couleurs ont des coordonnées négatives.

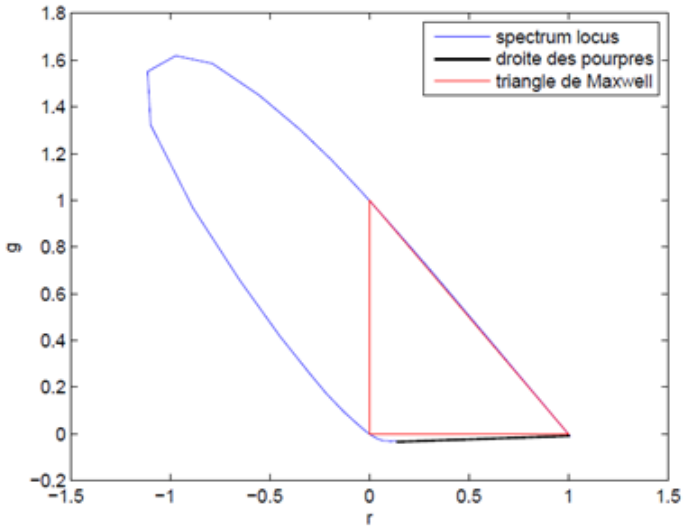


FIGURE 2.10 – Diagramme de chromaticité (r, g) lié au système RGB de la CIE (source : http ://www.cvrl.org/).

Le choix des primaires R_P, G_P et B_P n'est pas unique. Les différentes expériences d'appariement n'ont pas été réalisées avec le même système de primaires. Il est possible de réaliser un changement de primaires par des relations simples. Grâce aux lois de Grassmann, en exprimant les trois nouvelles primaires R'_P, G'_P et B'_p à l'aide des primaires R_P, G_P et B_P, le changement de primaire correspond à une transformation matricielle :

$$\begin{cases} R'_P = q_{11}R_P + q_{12}G_P + q_{13}B_P \\ G'_P = q_{21}R_P + q_{22}G_P + q_{23}B_P \\ B'_p = q_{31}R_P + q_{32}G_P + q_{33}B_P \end{cases} \tag{2.9}$$

Le système RGB basé sur les fonctions colorimétriques $\bar{r}(\lambda)$, $\bar{g}(\lambda)$ et $\bar{b}(\lambda)$ permet donc de quantifier la couleur, mais reste désavantageux. En effet, le système est dépendant du choix des primaires, il existe donc une multitude de systèmes RGB. D'autre part, les composantes du système peuvent prendre des valeurs négatives, et la luminance n'est pas une composante elle-même, mais une combinaison linéaire des trois composantes.

Les systèmes «humains» de perception de la couleur sont quant à eux issus d'une transformation non linéaire du système de primaires RGB. Il existe un grand nombre

d'espaces perceptuels se différenciant par les transformations nécessaires à leur obtention. Citons notamment le modèle cylindrique ISH (en anglais pour Intensity Saturation Hue) et le modèle triangulaire HSL (en anglais pour Hue Saturation Luminance). Nous avons choisi de ne développer qu'un des espaces les plus courants, le système HSV (en anglais pour Hue saturation value), que nous utilisons ultérieurement dans ce manuscrit.

2.3.3 Espace de couleur HSV

L'espace HSV est l'un des systèmes de couleurs cylindriques les plus courantes des points dans un modèle de couleur RGB. Cette représentation réorganise la géométrie du RGB dans le but d'être plus intuitive et pertinente sur le plan perceptuel que la représentation cartésienne (cube). Développé dans les années 1970 pour les applications d'infographie, HSV est aujourd'hui utilisé dans les sélecteurs de couleurs, dans les logiciels de retouche d'image et moins couramment dans l'analyse d'image et la vision par ordinateur.

L'espace HSV est une géométrie cylindrique, avec une teinte, sa dimension angulaire, commençant au primaire rouge à $0°$, passant par le primaire vert à $120°$ et le primaire bleu à $240°$, puis en retournant à rouge à $360°$. Les primaires additifs et soustractifs se situent sur les sommets des hexagones. La saturation S exprime l'éloignement de la couleur vis-à-vis de l'axe achromatique, tandis que la luminosité est donnée par la composante V, valant zéro pour le noir, et atteignant la valeur maximale au maximum de clarté que peut atteindre la couleur. Dans cette géométrie, l'axe vertical central comprend les couleurs neutres, achromatiques ou grises, allant du noir à la clarté 0 ou la valeur 0, le bas, au blanc à la clarté 1 ou la valeur 1, le haut. Dans la géométrie HSV, les couleurs primaires et secondaires additives — rouge, jaune, vert, cyan, bleu et magenta — et les mélanges linéaires entre des paires adjacentes, parfois appelées couleurs pures, sont disposés autour du bord extérieur du cylindre avec la saturation 1. Le mélange de ces couleurs pures avec du noir — produisant des nuances — laisse la saturation inchangée. Dans l'espace

HSV, la teinte seule réduit la saturation. Ce modèle de cône hexagonal est représenté sur la figure 2.11 b. La composante de teinte est matérialisée par un angle qui peut varier conventionnellement de $0°$ à $360°$ (figure 2.11).

$$\begin{cases} H = \arctan\left(\dfrac{\sqrt{3}\,(G-B)}{2R-G-B}\right) \\ V = \max(R,G,B) \\ S = \dfrac{\max(R,G,B)-\min(R,G,B)}{\max(R,G,B)}\, si \max(R,G,B) \neq 0, \text{ sinon } S = 0 \end{cases} \tag{2.10}$$

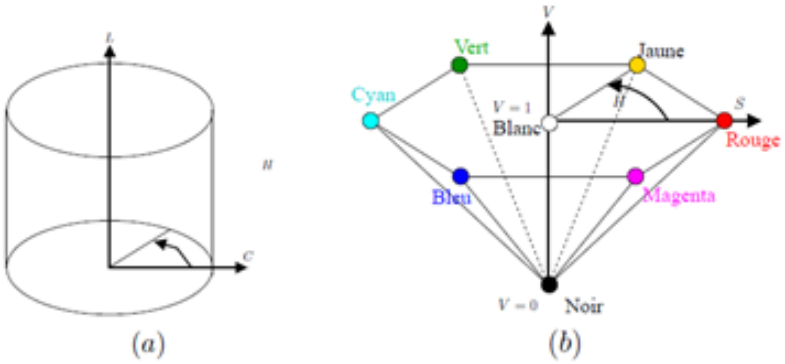


FIGURE 2.11 – Représentation du système de couleurs HSV : (a) Représentation cylindrique d'un espace perceptuel. (b) Système héxagonal HSV (source : [53])

2.3.3.1 Utilisation de système HSV

Dans [108], les auteurs ont analysé les propriétés de l'espace colorimétrique HSV en mettant l'accent sur la perception visuelle de la variation des valeurs de teinte, de saturation et d'intensité d'un pixel d'image. Ils ont extrait les caractéristiques des pixels en choisissant la teinte ou l'intensité comme propriété dominante en fonction de la valeur de saturation d'un pixel. La segmentation utilisant cette méthode permet une meilleure identification des objets dans une image par rapport à ceux générés en utilisant l'espace colorimétrique RGB.

Dans [53], les auteurs ont proposé une approche de segmentation d'image couleur où ils ont d'abord converti l'image RGB en une image HSV. Ensuite, ils ont appliqué le multi seuillage d'Otsu sur le canal V pour obtenir le meilleur seuil de

l'image. L'image résultante est ensuite segmentée avec le clustering K-Means pour fusionner les régions sursegmentées qui se sont produites en raison de l'application du multi seuillage d'Otsu. Enfin, ils ont effectué une soustraction de fond avec un traitement morphologique. Le résultat de l'approche s'avère satisfaisant selon les valeurs de MSE (en anglais pour Mean Square Error) et PSNR (en anglais pour Peak Signal to Noise Ratio) obtenues à partir de l'expérience.

Dans [20], les auteurs ont proposé une nouvelle technique de quantification pour l'espace colorimétrique HSV afin de générer un histogramme couleur et un histogramme gris pour le clustering K-Means, qui opère à travers différentes dimensions dans l'espace colorimétrique HSV. Dans cette approche, l'initialisation des centres de gravité et de nombre de grappes sont automatiquement estimés. Un filtre de post-traitement est introduit pour éliminer efficacement les petites régions spatiales. Cette méthode permet d'atteindre une vitesse de calcul élevée et les résultats sont proche aux perceptions humaines. Avec cette méthode, il devient également possible d'extraire efficacement les régions saillantes des images.

2.4 Image numérique couleur

Dans le domaine de la vision artificielle, les images couleur sont généralement acquises par une caméra vidéo couleur puis numérisées par un ordinateur via une carte d'acquisition (dans un système d'acquisition classique (figure 2.12)).

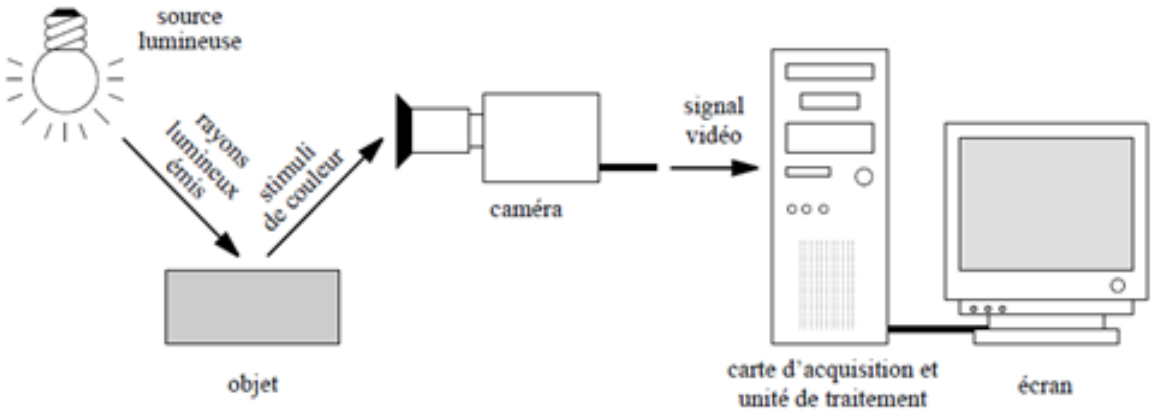


FIGURE 2.12 – Vision artificielle (source : [72]).

Après avoir défini le terme d'image numérique couleur. Nous proposons de présenter le principe d'acquisition des images par des caméras couleur classiques. Nous aborderons enfin l'influence de l'éclairage sur les couleurs caractérisant les pixels.

2.4.1 Définition d'une image numérique couleur

L'image numérique est un code qui est acquis, stocké et interprété. Cette abstraction de l'image en un code apporte un grand nombre d'avantages. Tout d'abord, cela permet sa reproductibilité à l'infini sans dégradation, puisqu'un fichier numérique est une série d'octets que l'on peut dupliquer exactement, quelle que soit la nature de ce fichier. Le codage permet aussi de faciliter la transmission de l'image, par exemple en la scindant en plus petites parties dont on conserve simplement les coordonnées dans le plan. Comme elle permet d'abstraire l'image de son support, il est également possible de la visualiser simultanément de manière rigoureusement identique. On peut par ailleurs l'afficher en fonction d'un environnement déterminé, ce qui permet l'interactivité (c'est le cas par exemple lorsqu'on navigue dans les grossissements d'une lame virtuelle ou que l'on insère ou affiche des annotations). En revanche, une image numérique est contrainte par l'espace de stockage que l'on souhaite lui assigner. Elle dépend également du système de codage de l'image et de son décodage (algorithme de compression et de décompression qui constitue le format de l'image), et enfin du matériel d'affichage (écran, projecteur, etc.).

2.4.2 Caractéristiques d'une image numérique couleur

L'image numérique la plus simple à stocker et à représenter se décompose en partie élémentaires : "PICTure ELementS" ou "PIXELS". Chaque pixel représente un point de l'image (dans un espace colorimétrique prédéfini). Dans une image en noir et blanc par exemple, un pixel sera représenté par 1 bit (0 = noir, 1 = blanc). Une image en niveaux de gris est codée avec 8 bits = 1 octet = $2^8 = 256$ valeurs du noir au blanc. Dans une image en couleurs, codée par exemple dans l'espace RGB (256

teints de rouge, 256 teints de vert, 256 teints de bleu), chaque pixel est représenté par 3 octets = $3 * 8$ bits permettant de représenter $\left(2^8\right)^3 = 16,8$ millions de couleurs.

2.4.3 Couleur des pixels et éclairage

La lumière réfléchie tombe sur le capteur de l'observateur (œil humain ou puce CCD (en anglais pour Charged Coupled Device) de la caméra) et conduit finalement à une perception ou une mesure de la couleur. Parce que nous traiterons des images numériques, nous concentrerons notre discussion sur la création d'images numériques et non sur l'exploration de la perception humaine de la couleur. La figure 2.13 montre le processus de formation d'image. La lumière d'une source lumineuse tombe sur une surface et est réfléchie. En fonction de l'angle entre la normale de la surface et la lumière incidente, une quantité différente de lumière est réfléchie. La lumière réfléchie tombe sur le capteur d'un observateur, ce qui conduit finalement à une perception ou à une mesure de la couleur.

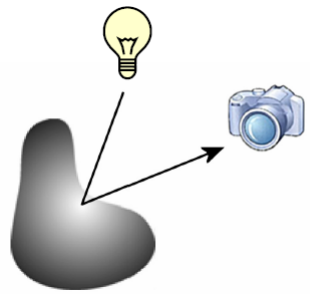


FIGURE 2.13 – Processus de formation d'image.

Les caméras numériques mesurent la lumière qui tombe sur leur puce CCD. Cette puce contient de nombreux petits capteurs. Ces capteurs ont une sensibilité qui varie en fonction de la partie du spectre lumineux qu'ils mesurent. Nous désignerons la sensibilité d'un capteur à la longueur d'onde λ par $SC(\lambda)$. Comme il n'existe pas de capteur unique capable de mesurer avec précision l'ensemble du spectre de la lumière visible, les dispositifs de capture d'image typiques échantillonnent la lumière entrante à l'aide de trois capteurs. En général, ces capteurs sont sensibles à la longueur d'onde de la lumière pour la couleur rouge, verte et bleue. Les réponses de

ces capteurs sont notées CR, CG et CB. Ensemble, ils forment un triplé de nombres. Mathématiquement, les réponses sont liées à la lumière, à la surface et au capteur du processus de formation d'image et sont définies comme suit :

$$\begin{pmatrix} CR \\ CG \\ CB \end{pmatrix} = \begin{pmatrix} (\vec{e}\cdot\vec{n}) \int E(\lambda) Rs(\lambda) SC_R(\lambda) \\ (\vec{e}\cdot\vec{n}) \int E(\lambda) Rs(\lambda) SC_G(\lambda) \\ (\vec{e}\cdot\vec{n}) \int E(\lambda) Rs(\lambda) SC_B(\lambda) \end{pmatrix} \tag{2.11}$$

Avec $E(\lambda)$ le spectre de la source lumineuse, $Rs(\lambda)$ la réflectance de surface et $SC_S(\lambda)$ la sensibilité du capteur S a différentes parties du spectre (figure 2.14).

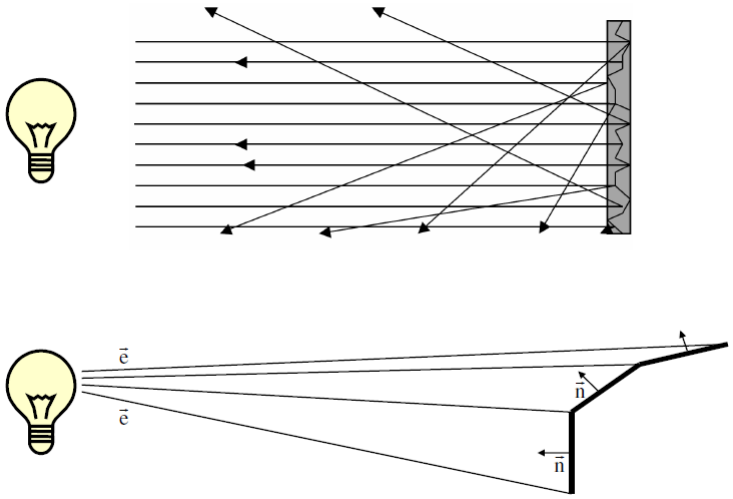


FIGURE 2.14 – La réflectance lambertienne d'une surface.

Dans la figure 2.14, en haut, la réflectance lambertienne d'une surface est illustrée. La lumière pénètre dans une surface et quitte la surface dans toutes les directions, réfléchie par les micro surfaces. Il s'agit du signal de couleur , qui dépend de la lumière incidente et de la réflectance de surface . En bas, on montre comment une surface lambertienne, qui réfléchit la lumière avec une intensité égale dans toutes les directions, peut encore avoir une intensité dépendant de l'angle entre la surface normale et l'angle d'incidence de la lumière : la lumière par unité de surface est différente.

Le signal de couleur $C(\lambda)$ qui est perçu ou mesuré par un capteur est créé par la lumière $E(\lambda)$ pénétrant la surface et réfléchie par les micro surfaces, en fonction de la réflectance de la surface $Rs(\lambda)$.

Nous supposons que les surfaces sont lambertiennes, c'est-à-dire qu'elles reflètent la lumière avec une intensité égale dans toutes les directions. Cela ne signifie pas que toutes les surfaces semblent également lumineuses. La quantité de lumière arrivant dépend de l'angle de la surface par rapport à la lumière. Ceci est illustré sur la figure 2.14 en bas. L'intensité de la lumière réfléchie dépend de l'angle entre la normale $\vec{n}$ de la surface et l'angle d'incidence de la lumière $\vec{e}$. On peut écrire le facteur d'échelle $(\vec{e} \cdot \vec{n})$ en dehors de l'intégrale, car il ne dépend pas de la longueur d'onde de la lumière.

Les images numériques sont créées en prenant des échantillons de couleur d'une scène à de nombreux endroits adjacents. Un pixel est l'élément fondamental d'une image numérique. Avec un pixel, une seule couleur est associée. Un seul pixel dans cette grille correspond à une zone finie d'une puce d'appareil photo numérique. Sur cette zone finie, les capteurs R, G et B sont échantillonnés. Ces échantillons sont combinés en un triplet RGB, qui est associé à un pixel. Les échantillons RGB doivent être quantifiés dans une plage et une précision limitées pour permettre le stockage numérique. Nous supposerons une plage de $[0.1]$ pour le triplet RGB. Les triplets RGB comprennent : $(1,0,0)$ est rouge, $(0,1,0)$ est vert, $(0,0,1)$ est bleu, $(1,1,1)$ est blanc et $(0,0,0)$ est noir, $(1,1,0)$ est jaune et $\left(1,\frac{3}{4},\frac{3}{4}\right)$ est rose.

2.5 Conclusion

La couleur est une sensation résultant de l'interaction des neurones dans le cerveau humain qui est projetée sur le monde extérieur, permettant ainsi d'améliorer la perception et le traitement de l'information visuelle. Bien que la couleur dépende du monde physique qui nous entoure, nous avons vu dans ce chapitre que la couleur n'a pas de réalité physique. C'est le résultat d'une interprétation subjective des

signaux visuels conçus dans le système visuel composé de l'œil et du cerveau. C'est une perception sensorielle. Nous avons vu également que la couleur comme la brillance, la forme, la texture est un attribut d'apparence qui caractérise et identifie un objet. Elle dépend à la fois de l'éclairage, de la nature du matériau et de l'observateur. Une couleur isolée est identifiée suivant trois entités : sa teinte, sa saturation et sa clarté.

La colorimétrie s'intéresse à l'étude de la sensation colorée ainsi qu'à sa mesure objective, par l'utilisation des propriétés de la trichromie. La quantité d'espaces couleur existant est si importante, du fait de la diversité des méthodes d'analyse couleur, qu'il paraît vain de vouloir en dresser une liste exhaustive. Néanmoins, nous en avons décrit quelques-uns, suffisamment représentatifs des différents types d'espaces rencontrés dans la littérature. Du fait de cette diversité, le choix de l'espace couleur le mieux approprié à une application n'est pas aisé et il n'existe malheureusement pas de loi universelle.

Quand nous parlons de la couleur d'un objet, nous ne prenons généralement pas la peine de préciser la couleur de la lumière qui l'éclaire. Nous avons pourtant tous déjà pu constater dans ce chapitre que la couleur d'éclairage peut donner lieu pour l'objet à des changements de couleur spectaculaires, au point qu'il n'est parfois plus possible de deviner ses « vraies » couleurs (c'est-à-dire ses couleurs en lumière blanche).

CHAPITRE 3

Catégorisation des images

«Objectivité ne signifie pas impartialité mais universalité.»

Raymond Aron

Résumé

Dans ce chapitre, nous allons présenter les principales étapes et outils nécessaires pour développer un système de reconnaissance d'objets ainsi que les différentes améliorations qui ont été proposées jusqu'à présent pour traiter les différentes limitations.

Contents

Dans l'Allégorie de la Caverne [93], Platon présente sa théorie des idées, et introduit les notions de forme et de chose. Les formes sont des modèles, des concepts, et les Choses sont leurs réalisations. Par exemple, le cercle mathématique (le concept) est une Forme, et un cercle dessiné sur une feuille est une Chose, ce que l'on peut appeler une réalisation ou une instance du concept. Dans le cadre plus moderne de la reconnaissance d'objets en vision par ordinateur, on différencie de la même manière les catégories d'objets (associées aux Formes) et les instances d'objets (associées aux choses). Les bicyclettes, les ordinateurs, les visages, sont des exemples de catégories d'objets. La bicyclette rouge et verte de M. Mohamed, un ordinateur $ToshibaA-610$ et le visage de M. Mohamed sont des instances d'objets. Par conséquent, ce domaine est depuis longtemps un objet d'intérêt pour la communauté scientifique. Différents objectifs et méthodes ont été proposés, depuis plus de 50 ans. En général, on peut les catégoriser en trois tâches en fonction de leur objectif [62] :

— La catégorisation ou classification d'images qui consiste à donner un label à une image en fonction de la présence ou non d'un objet appartenant à une catégorie donnée.

— La détection d'objets qui désigne la tâche de localisation des objets d'une catégorie donnée.

— La segmentation de classes d'objets qui consiste à déterminer quels sont les pixels de l'image qui appartiennent à un objet d'une des classes d'intérêt.

À vrai dire, ces trois tâches sont étroitement liées. Toutes les trois suivent un paradigme assez ancien proposé par David Marr [77]. Ce paradigme suggère une analyse uniquement ascendante, centrée sur les données. De ce fait, les mêmes outils peuvent être mis en œuvre pour les résoudre. Ce paradigme constitue le cœur de la plupart des méthodes de reconnaissance d'objets et surtout des systèmes de catégorisation d'images. L'objectif de ces systèmes est de prédire la nature de l'objet dans une image au sein d'une liste exhaustive de possibilités. Dans cette thèse, nous allons nous concentrer sur la tâche de «catégorisation d'images» puisqu'elle

peut être considérée comme une généralisation des autres tâches citées ci-dessus. La classification d'images selon la catégorie d'objet reste un vrai challenge, étant donné que l'apparence des objets au sein d'une catégorie varie grandement, suite aux modifications de position, orientation et échelle, aux modifications d'illumination, occultations et aux grandes variabilités de formes au sein de cette classe.

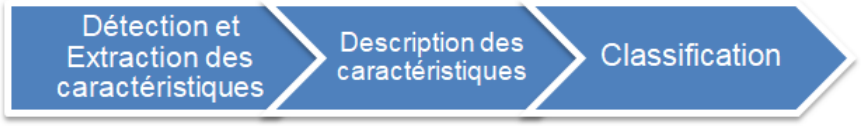


FIGURE 3.1 – Architecture simple d'un système de reconnaissance d'objets

Pour ces raisons, les méthodes ont eu recours à des approches plus locales. Le schéma classique est représenté dans la figure (3.1). Dans ces approches, l'image est considérée comme une collection de régions d'intérêts (ou points d'intérêts), généralement de taille faible par rapport à la taille de l'image. Ces régions sont détectées dans l'étape «Détection et Extraction des caractéristiques», et ensuite transformées en vecteurs représentant les caractéristiques de l'image, par exemple des contours et/ou orientations (Description des caractéristiques). À partir de ces vecteurs, chaque image est représentée par un histogramme servant comme base pour catégoriser l'image selon l'objet qu'elle contient (Classification). Ces approches seront abordées en détail dans la suite de ce chapitre. Nous allons présenter les principales étapes et outils nécessaires pour développer un système de reconnaissance d'objets ainsi que les différentes améliorations qui ont été proposées jusqu'à présent pour traiter les différentes limitations citées ci-dessus.

3.1 Détection et Extraction des caractéristiques

La détection des caractéristiques est l'étape requise pour obtenir des descriptions locales. Il localise des points et des régions et est généralement capable de reproduire des niveaux de performances similaires aux observateurs humains en localisant les caractéristiques élémentaires dans un large éventail de types d'images.

Solon G.GALES [33], la plupart des détecteurs existants peuvent être classés en deux types :

— **Détecteurs à échelle fixe** : la réponse est calculée avec une taille de fenêtre et une force de lissage gaussien prédéfini.

— **Détecteurs à multiéchelle** : la réponse est calculée avec des tailles de fenêtre (ou des résolutions d'image) et des forces de lissage gaussien différentes.

Dans la partie suivante, nous décrirons quelques détecteurs à échelle fixe et certains détecteurs multi-échelles.

3.1.1 Détecteurs à échelle fixe

Les détecteurs à échelle fixe consistent à traiter tous les pixels dans l'image. Cette approche à l'avantage d'être la plus informative que les autres méthodes de détections, cependant, elle nécessite des ressources en temps et en mémoire très importantes : la plupart du temps de calcul étant passé a traité les régions peu informatives. Dans cette partie, nous décrirons quelques détecteurs à échelle fixe.

3.1.1.1 Détecteur de Harris et Stephens

La contribution de Harris et Stephens [45] intervient pour améliorer le détecteur proposé par Moravec. Ils ont corrigé plusieurs défauts parmi lesquels la prise en compte des variations d'intensité dans toutes les directions et l'utilisation de fenêtres circulaires au lieu de fenêtres rectangulaires autour des pixels. Ce détecteur recherche les coins où le gradient change dans deux directions. La direction du gradient n'affecte pas la détection, rendant ainsi la rotation de détection invariante. Le détecteur est quelque peu robuste aux variations d'éclairage, car le gradient est insensible aux changements de luminosité. La matrice de second moment basée sur

une dérivée de premier ordre (voir l'équation (3.1)) est utilisée pour la détection.

$$M = \mu(x, \sigma_I, \sigma_D) = \sigma_D^2 g(\sigma_I) * \begin{bmatrix} I_x^2(x, \sigma_D) & I_x(x, \sigma_D) I_y(x, \sigma_D) \\ I_x(x, \sigma_D) I_y(x, \sigma_D) & I_y^2(x, \sigma_D) \end{bmatrix} \tag{3.1}$$

Où $I_x(x, \sigma_D)$ et $I_y(x, \sigma_D)$ sont les dérivées de premier ordre de l'image. σ_D est l'échelle gaussienne à laquelle les premières dérivées partielles de l'image sont calculées. $g(\sigma_I)$ est la fonction d'intégration gaussienne utilisée pour la moyenne des dérivées. La matrice M (équation (3.1)) est également appelée matrice d'auto-corrélation. Il décrit la distribution du gradient dans un voisinage local d'un point. Les valeurs propres de cette matrice représentent deux changements de signal principaux dans le voisinage. Cette propriété permet l'extraction de points d'angle pour lesquels les deux valeurs propres sont significatives. Quand une valeur propre est grande et l'autre est petite, la fenêtre d'image transmet un bord. Si les deux valeurs propres sont petites, il s'agit d'une zone homogène. La mesure de Harris (équation (3.2)) évite de calculer directement les valeurs propres. Il mesure les bords (corneness) en utilisant le déterminant et la trace de la matrice. Les points dont la valeur de mesure de Harris est supérieure à un seuil sont sélectionnés comme points détectés.

$$\text{cornerness} = \det(M) - \alpha \cdot trace^2(M) \tag{3.2}$$

Avec α une constante empirique avec des valeurs comprises entre $0,04$ et $0,06$.

3.1.1.2 Détecteur SUSAN (Smallest Univalue Segment Assimilating Nucleus)

Smith et Brady [106] ont considéré que les pixels dans une région relativement petite sont uniformes en termes de luminosité si ces pixels appartiennent au même objet. Sur la base de ce point de vue, le SUSAN est implémentée en comparant la luminosité au sein d'un masque circulaire MC. En détail, la luminosité de chaque pixel du masque, $\overrightarrow{p} \in MC$, est comparée à celle du noyau $\overrightarrow{p}_0$ (le centre de MC) par

une fonction de comparaison :

$$m(\overrightarrow{p}) = \exp\left(-\left(t^{-1}\left(I(\overrightarrow{p}) - I(\overrightarrow{p}_0)\right)\right)^6\right) \tag{3.3}$$

Où t est un seuil. Un noyau $\overrightarrow{p}_0$ est un coin, si le nombre de pixels, qui sont similaires à $\overrightarrow{p}_0$ en termes de luminosité, est inférieur à un seuil donné. Le détecteur SUSAN a une bonne fiabilité. Il localise les coins avec précision et est extrêmement insensible au bruit. C'est relativement rapide. Cependant, il fonctionne mal pour les images floues.

3.1.1.3 Détecteur de Trajkovic

Miroslav Trajkovic et Mark Hedley [113] ont développé cet opérateur afin d'obtenir des taux de répétabilité et des performances de localisation comparables à ceux des détecteurs de coin les plus populaires, tout en nécessitant un minimum de calcul. Ils ont comparé leur opérateur à l'opérateur de Plessey (et à d'autres) et affirment que le taux de répétabilité est légèrement inférieur, mais que la localisation est comparable sur les jonctions en L et améliorées sur d'autres types de jonctions. Ils montrent empiriquement que leur opérateur est cinq fois plus rapide que l'opérateur de Plessey et au moins trois fois plus rapide que tous les opérateurs considérés, y compris l'opérateur SUSAN qui est déjà considéré comme efficace sur le plan informatique.

Trajkovic et Hedley adoptent la même définition d'un coin que les opérateurs Moravec et Harris : les coins sont des points où le changement d'intensité d'image est élevé dans toutes les directions. Comme Moravec, ils définissent la mesure de la corneness comme le changement minimal d'intensité dans toutes les directions possible. Les performances sont améliorées par rapport à l'opérateur de Moravec en effectuant une approximation interpixel afin d'estimer le changement d'intensité dans toutes les directions (contrairement au nombre fini de directions envisagées par Moravec). Les performances et la demande de calcul sont toutes deux amélio-

rées en utilisant une approche multigrille où les emplacements probables des coins sont d'abord trouvés dans une version basse résolution de l'image d'origine.
Les exigences minimales de calcul de cet opérateur le rendent bien adapté aux applications en temps réel. Cependant, cet opérateur montre n'est pas invariant en rotation, est sensible au bruit et elle réagit trop facilement aux bords diagonaux. Les applications nécessitant un taux de répétabilité élevé favoriseront probablement encore l'opérateur Harris, notamment avec les extensions récentes pour le rendre invariant en rotation.

3.1.1.4 Détecteur de Rosten et Drummond

Rosten et Drummond [99] ont utilisé l'apprentissage automatique pour accélérer la détection des coins. Le processus comprend les trois étapes suivantes :

— Test de segment sur un cercle de Bresenham d'un pixel central p de rayon 3 : cette étape est efficace sur le plan informatique pour éliminer la plupart des candidats non-coin. Il est basé sur un test logistique, c'est-à-dire que p n'est pas un coin si un pixel avec la position x et un autre pixel avec la position $x+8$ sont similaires à p en termes d'intensité. Le test sera effectué sur 12 positions consécutives.

— Détection de coin basé sur la classification : appliquer le classificateur d'arbre $ID3$ [96] pour déterminer si p est un coin basé sur 16 caractéristiques. Chaque fonction est 0, 1 ou -1. Si un pixel avec la position x sur le cercle de Bresenham de p est plus grand (plus petit) que p, la caractéristique correspondante est $1(-1)$. Sinon, la fonction est 0.

— Vérification des coins : la suppression non maximale est utilisée pour la vérification.

3.1.1.5 Autres détecteurs

Le détecteur de coin de Beaudet [8] est un opérateur invariant en rotation basée sur une mesure de coin. Cet opérateur est sensible au bruit du fait du calcul des

dérivées secondes. Kitchen et Rosenfeld [30] détectent les coins au maximum local d'une mesure de coin en fonction du changement de direction du gradient le long d'un bord pondéré par la magnitude du gradient [56]. Semblable au détecteur de coin de Harris, l'opérateur Forstner utilise également une mesure de coin définie par la seconde matrice de moments. Le seuil est déterminé par les statistiques locales.

3.1.2 Détecteurs à multiéchelle

La majorité des méthodes proposées souffrent du problème de changement d'échelle. Ce qui a fait émerger la notion de multiéchelles selon laquelle les points d'intérêt sont calculés sur plusieurs échelles. La totalité de ces points est ensuite utilisée pour représenter l'image. Dans l'approche proposée par [103], les points sont extraits par l'opérateur Harris sur plusieurs échelles parce que ce détecteur a prouvé ses excellents résultats face aux rotations, bruit et condition de luminance, mais a échoué en présence de changement d'échelle. Dans cette partie, nous décrirons quelques détecteurs multi-échelles.

3.1.2.1 Détecteurs Harris-affine et Hessian-affine

Soit p un pixel d'une image intensité I de coordonnées (x,y), la matrice Hessienne H_σ en p et à l'échelle σest donné par l'expression suivante :

$$H_\sigma = \begin{bmatrix} L_{xx}(p,\sigma) & L_{xy}(p,\sigma) \\ L_{yx}(p,\sigma) & L_{yy}(p,\sigma) \end{bmatrix} \tag{3.4}$$

Où $L_{xx}(p,\sigma)$ dénote le produit de convolution de l'image intensité I par la dérivée de deuxième ordre d'une Gaussien $\partial^2 g(\sigma)/\partial x^2$, similairement à L_{yy} et L_{yx}.
En 2004, Mikolajczyk et Schmid proposent de coupler un détecteur multiéchelles avec la méthode d'adaptation (ou de normalisation) affine décrite précédemment. Ils proposent ainsi deux nouveaux détecteurs [79] : le Harris—affine et le Hessian—affine. Pour chacun d'entre eux, la méthode repose tout d'abord sur l'extrac-

tion multiéchelles des points d'intérêt, puis sur la détermination itérative d'une région locale circulaire.

La partie analyse multiéchelles de la scène est assurée par l'utilisation du détecteur Harris—Laplace (équation (3.1)) pour le Harris—affine, et celle du détecteur Hessian—Laplace (équation (3.4))) pour le Hessian—affine. La détermination de la région elliptique centrée sur le point d'intérêt se divise en plusieurs étapes. Dans un premier temps, l'initialisation de la matrice de transformation U^0 (correspondant à l'identité) et la récupération des données extraites $\left(x^0, \sigma_D^0, \sigma_I^0\right)$ est nécessaire. L'application de la matrice U^k sur l'image permettra de la déformer an de faire converger la région locale vers une forme circulaire. L'étape suivante consiste à déterminer les nouvelles valeurs d'échelle d'intégration σ_I^k et de différenciationσ_D^k en se basant sur les équations suivantes :

$$\sigma_I^k = \underset{\sigma_I = t\sigma_I^{(k-1)} \, pour \, t \in [0,7;1,4]}{\arg\max} \left(\sigma_I^2 \times \det\left(L_{xx}\left(x, \sigma_I\right) + L_{yy}\left(x, \sigma_I\right)\right)\right) \tag{3.5}$$

et :

$$\sigma_D^k = \underset{\sigma_D = s\sigma_I^k \, pour \, s \in [0,5;0,75]}{\arg\max} = \frac{\lambda_{\min}\left(\mu\left(x^k, \sigma_I^k, \sigma_D\right)\right)}{\lambda_{\max}\left(\mu\left(x^k, \sigma_I^k, \sigma_D\right)\right)} \tag{3.6}$$

En utilisant ces deux valeurs dans l'équation (3.1), Mikolajczyk et Schmid proposent un recalage du point d'intérêt $x^0 \rightarrow x^k$ et le définissent par :

$$x^k = \underset{x \in W\left(x^{k-1}\right)}{\arg\max} \left(\det\left(\mu\left(x, \sigma_I^k, \sigma_D^k\right)\right) - \alpha trace^2\left(\mu\left(x, \sigma_I^k, \sigma_D^k\right)\right)\right) \tag{3.7}$$

Les étapes suivantes consistent à mettre à jour la matrice $U^k = \mu^k U^{k-1}$ et à réitérer ce processus jusqu'à l'obtention d'une région locale circulaire. La normalisation affine, caractérisée par la matrice U^k entraîne la transformation de l'ellipse en cercle. Les deux régions obtenues sont liées par une simple rotation R, permettant ainsi la suppression des problèmes dus aux transformations affines et projectives. En définitive, ces deux méthodes ne diffèrent que par leur partie analyse multiéchelles, et donnent un nombre de points et des résultats similaires.

3.1.2.2 Détecteurs EBR (Edge Based Region detector)

Tuytelaars et Van Gool ont proposé [119] en 1999 un détecteur de régions invariantes aux transformations affines, désigné par l'acronyme EBR pour Edge Based Region. Cette méthode est basée sur la détection de coins de Harris [45] et sur les contours de l'image. Ces deux types d'information possèdent la particularité d'être détectables, quels que soient les changements d'échelle, de point de vue et d'éclairage. Soit p un point à l'intersection d'un point d'intérêt de Harris et d'un contour obtenu à l'aide du détecteur de contours de Canny-Deriche [17]. Soit p_1 et p_2 deux points se déplaçant en sens opposé de part et d'autre de p. On définit l_i la vitesse des points par l'équation suivante :

$$l_i = \int abs\left(\left|p_i^{(1)}(s_i)\, p - p_i(s_i)\right|\right) ds_i \tag{3.8}$$

Avec s_i un paramètre de courbure arbitraire, $p_i^{(1)}$la dérivée première de$p_i(s_i)$, $abs()$ représente la valeur absolue et $||$le déterminant. Cette égalité stipule que l'aire contenue entre le segment $[p_1, p]$et le contour d'une part, et l'aire entre le segment $[p_2, p]$et le contour d'autre part, reste identique. Pour plus de simplicité, nous utiliserons l pour se référer à$l_1 = l_2$. Pour toutes valeurs de l, on définit à partir des points $p_1(l)$, $p_2(l)$ et $p(l)$, un parallélogramme $\Omega(l)$, construit à partir des vecteurs $p_1(l)\bar{}p(l)$ et $p_2(l)\bar{}p(l)$. Enfin ne seront sélectionnés comme régions d'intérêt que les parallélogrammes donnant des extrêmes à une fonction monodimensionnel contenant des attributs de textures. Ces attributs sont définis de la manière suivante :

$$Inv_1 = abs\left(\frac{|p_1 - p_g\ p_2 - p_g|}{|p - p_1\ p - p_2|}\right) \frac{\xi_{00}^1}{\sqrt{\xi_{00}^2 \xi_{00}^0 - \left(\xi_{00}^1\right)^2}} \tag{3.9}$$

$$Inv_2 = abs\left(\frac{|p - p_g\ q - p_g|}{|p - p_1\ p - p_2|}\right) \frac{M_{00}^1}{\sqrt{M_{00}^2 M_{00}^0 - \left(M_{00}^1\right)^2}} \tag{3.10}$$

Avec :

$$\xi_{pq}^n = \int_{\Omega} I^n(x,y)\, x^p y^q dxdy \tag{3.11}$$

Et :

$$p_g = \left(\frac{\xi^1_{10}}{\xi^1_{00}}, \frac{\xi^1_{01}}{\xi^1_{00}} \right) \tag{3.12}$$

Où ξ^n_{pq} est le moment d'ordre n et de degré $p+q$ calculé dans la région $\Omega(l)$. p_g est le centre de gravité pondéré par l'intensité moyenne de $I(x,y)$. q est un sommet du parallélogramme situé à l'opposé de p. La principale limite de cette méthode de détection de régions d'intérêts est qu'elle est dépendante des performances du détecteur de Harris et du détecteur de contours utilisé. EBR a été essentiellement utilisé dans les applications de reconnaissance de formes [119].

3.1.2.3 Détecteurs IBR (Intensity-based region detector)

D'une manière générale, on peut dire qu'IBR (Intensity Based Region) est une méthode de détection de régions d'intérêts qui utilise comme point d'ancrage un extremum de l'intensité de l'image. À partir de ce point d'ancrage, on parcourt l'image à l'aide de rayons tracés autour de ce point.

Le principe d'exploration est le suivant : partant d'un extremum d'intensité de l'image, on construit la fonction d'intensité définie par les valeurs contenues sur les rayons du cercle centré sur l'extremum local détecté. Outre cette fonction d'intensité, on calcule en tous points du rayon la fonction suivante :

$$f_I(t) = \frac{abs(I(t) - I_0)}{\max(K, d)} \tag{3.13}$$

Avec :

$$K = \frac{\int_0^t abs(I(t) - I_0)}{t} \tag{3.14}$$

Avec t un paramètre arbitraire le long du rayon, $I(t)$ l'intensité de l'image à la position t, I_0 l'intensité au point d'ancrage et $d > 0$. Les points pour lesquels cette fonction atteint un extremum sont invariants aux transformations affines et aux transformations photométriques. Enfin pour construire la région finale, on joint les extrêmes de toutes les fonctions dans l'espace de l'image. Comme pour EBR, IBR est dé-

pendante des points d'ancrage donc de la robustesse de la méthode de détection de ces points. Les régions d'intérêts localisées par IBR ont notamment été utilisées dans les approches de type correspondance de ligne de base large (wide basiline matching) [120].

3.1.2.4 Le SIFT (Scale Invariant Features Transform)

Le détecteur SIFT de David Lowe [71] est une contribution majeure dans le domaine, présentant une invariance à l'échelle et approchant le fonctionnement du cortex visuel humain. Son fonctionnement n'est pas très compliqué : il commence par générer des images progressivement floutées par application du filtre gaussien en augmentant la variance σ^2 à chaque étape. Le pixel à la position x, y dans l'image I a un équivalent flouté défini par :

$$L(x,y,\sigma) = G(x,y,\sigma) * I(x,y) \tag{3.15}$$

Avec :

$$G(x,y,\sigma) = \frac{1}{2\pi\sigma^2} e^{-(x^2+y^2)/2\sigma^2} \tag{3.16}$$

Lowe recommande respectivement 4 et 5 images, une fois ces images générées il est alors procédé aux calculs des différences des échelles successives dans chaque octave (différences de gaussiennes). Ces calculs sont en réalité une approximation du Laplacien de convolutions gaussiennes (LOG(en anglais pour Laplacian Of Gaussian)), pratique pour détecter des zones de fortes variations dans l'image. Pour chacune de ces nouvelles images, on parcourt les pixels un à un en inspectant les 8 voisins dans l'échelle courante ainsi que dans les échelles adjacentes (pour un total de 26 pixels voisine) : si le pixel étudié à l'intensité la plus petite ou la plus grande de cet ensemble de voisinage, il est marqué comme un point d'intérêt. Lowe propose ensuite de faire une recherche subpixelique du maximum local (puisque le pixel est en réalité une approximation de ce maximum) par le biais d'une expansion de Taylor de l'image autour du pixel.

Enfin, les points d'intérêt candidat sont sélectionnés selon un critère de saillance [52] qui est l'intensité du contraste local : si celle-ci est en dessous d'un certain seuil, le point d'intérêt n'est pas conservé. Il en va de même pour les pixels situés sur un bord. Une mesure similaire à celle de Harris par l'étude des gradients verticaux et horizontaux permet de sélectionner ceux qui sont proches d'un coin. Notons en défaut qu'il n'y a aucune garantie sur une fréquence spatiale équitablement répartie des extremums dans l'espace des échelles, c'est même souvent le contraire qui se produit si des zones regroupent de nombreux points dans un espace géographique réduit.

3.1.2.5 Le SURF (Speeded Up Robust Features)

SIFT a été un véritable succès dans la communauté de vision, mais son coût calculatoire élevé était un défaut contraignant. La contribution de SURF [7] visait à répondre à cette problématique en présentant un détecteur et un descripteur quasiment aussi performant, mais beaucoup plus rapide à calculer. Dans les grandes lignes, cet algorithme est très similaire à SIFT : il s'agit d'approximer le Laplacien de convolutions de Gaussiennes pour détecter les extremums dans l'espace-échelle. Mais les auteurs font remarquer qu'il est possible d'aller encore plus loin dans l'approximation en remplaçant les convolutions de filtres gaussiens par des filtres rectangulaires. L'intérêt d'utiliser des filtres rectangulaires est qu'ils peuvent être calculés très rapidement grâce au principe des images intégrales [22]. Dans les faits, une telle approximation se révèle suffisante.

3.2 Description des caractéristiques

Lorsque des zones de l'image ont été définies et sélectionnées, il faut analyser l'information qu'elles contiennent. Deux types de traitement sont envisageables. Le premier type consiste en une amélioration de la qualité de l'image, par exemple pour atténuer les défauts du capteur et les bruits associés aux conditions d'acqui-

sition (floue, mauvaise illumination...). Ces traitements sont parfois réalisés avant la phase de Détection des caractéristiques. Le deuxième type est directement lié à la tâche de reconnaissance d'objets. Âpres avoir extrairez des zones d'intérêt (les caractéristiques) des informations relatives à ce qui est contenu dans l'image, on passe à regrouper ces caractéristiques dans des vecteurs s'appelant descripteurs. Selon Leyrit [65] il existe des descripteurs d'images qui permettent de caractériser l'information disponible. Trois façons de procéder sont envisageables :

— Soit tous les pixels de l'image correspondants à la zone d'intérêt sont pris en compte dans la description : c'est un descripteur global.

— Soit l'image est décrite en des points particuliers et significatifs : c'est un descripteur local et ces points intéressants sont des points d'intérêt.

Dans la partie suivant, nous décrirons quelque exemple de descripteurs globaux et d'autre de descripteurs locaux ainsi que descripteurs multiples.

3.2.1 Descripteurs globaux

Pour le dire simplement, les techniques utilisant des descripteurs globaux visent à reconnaître l'objet dans son ensemble. Pour atteindre ce résultat, il faut généralement apprendre, à partir d'un ensemble d'images, l'objet à reconnaître.

Certaines premières applications de vision par ordinateur utilisaient directement des valeurs d'échelle de gris de l'image. La similitude entre deux images est calculée comme la distance entre les valeurs de pixels des deux images. Par exemple, la corrélation d'images calcule la distance euclidienne entre deux images. Ce type d'application nécessite que les images soient prises sur un fond clair et aussi que les objets soient bien alignés sans occlusion ni déformation. Ces conditions sont difficiles à satisfaire dans la pratique, en particulier dans la reconnaissance des catégories d'objets, où les objets ont de nombreuses formes et apparences. La méthode de correspondance globale ne peut être appliquée qu'à la mise en correspondance du même objet avec des conditions de visualisation similaires.

Pour améliorer la robustesse à la variation de l'image, des méthodes statistiques

peuvent être utilisées. Les techniques de visages propres (Eigenfaces) [118] sont un ensemble «d'ingrédients normalisés pour le visage», dérivé de l'analyse statistique de nombreuses images de visages. Tout visage humain peut être considéré comme une combinaison de ces visages standards. Pour générer un ensemble de visages propres, les images de dimension $m * n$ sont traitées comme des vecteurs de mn-dimension dont les composantes sont les valeurs de leurs pixels. Les vecteurs propres de la matrice de covariance de la distribution statistique de ces vecteurs de mn-dimension sont ensuite extraits. Ces vecteurs propres sont des visages propres. La méthode des visages propres génère des représentations compressées d'images et rend la représentation moins sensible au bruit. Pour que la méthode des visages propres fonctionne, les images doivent être alignées avant le traitement. Cependant, cette méthode n'est pas robuste à la déformation, à l'occlusion et au fouillis d'arrière-plans.

Afin d'éliminer l'influence de l'occlusion partielle et du fouillis d'arrière-plans, A.Leonardis et H.Bischoff [68] extraient les visages propres par un paradigme d'hypothèse et de test utilisant des sous-ensembles de points d'image. Les hypothèses concurrentes sont ensuite soumises à une procédure de sélection basée sur le principe de la longueur minimale de description. Leur méthode réduit l'influence de l'occlusion et l'encombrement des arrière-plans. Cependant, il présente toujours une sensibilité à la déformation de l'image.

Cependant, l'utilisation des fonctionnalités globales présente plusieurs inconvénients : tout d'abord, l'objet doit remplir toute l'image de test pour correspondre au modèle. Pour surmonter ce problème, les techniques de fenêtres glissantes sont généralement utilisées pour permettre l'invariance de la translation, de la mise à l'échelle et de la rotation. Cette solution a néanmoins un coût de calcul important (des milliers de fenêtres doivent être examinées [39] alors que la reconnaissance d'objets spécifiques implique généralement des contraintes en temps réel. La récupération précise de la pose du modèle 3D à l'aide de descripteur global semble également très difficile. Troisièmement, la quantité de données nécessaires à l'en-

traînement est généralement énorme, ainsi que le temps d'entraînement. Un dernier problème est que ces approches ont des difficultés à traiter les occlusions partielles. Ces problèmes sont admissibles pour la reconnaissance classe d'objets, car de nombreuses images de modèle sont nécessaires de toute façon pour apprendre précisément les variations intraclasse, et la tâche est suffisamment difficile pour permettre d'éviter le problème d'occlusion. Au contraire, nous attendons davantage d'un système plus simple traitant d'objets spécifiques : c'est-à-dire la formation du modèle à partir de seulement quelques images et portant des occlusions.

3.2.2 Descripteurs locaux

Nous classons dans cette partie les descripteurs locaux en fonction des primitives utilisées pour décrire les objets tels que sa forme, son contour, sa texture, sa couleur, sa région, ces points d'intérêt, etc.

3.2.2.1 Descripteurs de formes

Les descripteurs de formes permettent, comme leur nom l'indique, de présenter une information pertinente sur le contenu de l'image et précisément sur la forme. Il existe différents types de descripteurs de formes qui se différencient par leur simplicité/complexité. Ces descripteurs sont généralement basés sur les moments et l'utilisation des moments est répandue dans le domaine de la reconnaissance des formes. Depuis leur introduction par Hu [48] en 1961, qui proposait un ensemble de moments invariants à la translation, à l'échelle et à la rotation en utilisant la théorie de l'invariant algébrique, Fulsser et Suk [29] l'ont étendu à l'invariant des moments affines (AMI) qui est invariants à la transformation affine, et Van Gool [126] suggère un ensemble qui est en outre invariant à la condition photométrique. Ces moments basés sur la forme sont plus sensibles au bruit que le moment basé sur une base orthogonale comme les moments de Zernike [55] qui sont invariants à la rotation et à l'échelle. Wang et al [130] ont étendu ces moments pour qu'ils soient invariants à l'illumination, avec une bonne expérience, mais la méthode implique une grande

complexité de calcul. Les travaux de Adam [2] ont montré que la transformée de Fourier Mellin donne de meilleurs résultats que d'autres signatures généralement utilisées dans la littérature pour la reconnaissance de caractères avec des rotations multiorientées multiéchelles sur leurs images jusqu'à 180 et robuste contre le bruit.

3.2.2.2 Descripteurs de contour

La deuxième approche généralement mentionnée est le descripteur de Fourier [100]. Elle implique une caractérisation des contours de la forme. Les descripteurs d'espace à l'échelle de courbure (CSSDs (en anglais pour Curvature Scale Space Descriptors)) [1] sont également largement utilisés, comme les descripteurs de forme qui détectent les points de courbure des contours à différentes échelles à l'aide d'un noyau gaussien pour convoler des silhouettes successives. Les résultats expérimentaux de Zhang [139] montrent que les descripteurs de Fourier sont plus robustes au bruit que les CSSD. Le principal inconvénient est la nécessité d'obtenir un objet avec un contour clairement segmenté. Il est difficile d'obtenir un contour complet proche de l'objet. De plus, la détection de tous les bords contenus dans une image peut être perturbée par les contours internes ou externes de l'objet. Nous n'allons pas détaille ce type de descripteurs parce que nous ne les avons pas utilisé dans notre travail.

3.2.2.3 Descripteurs de texture

De nombreux algorithmes d'extraction de descripteurs de texture ont été proposés dans la littérature du traitement d'images. Reed et DuBuf [97] les catégorisent en des approches à base de descripteurs, de modèles et de structures. Feddaoui et Hamrouni [83], quant à eux, les classifient en trois approches : structurelles, statistiques et spatiofréquentielles. Toyoda et Hasegawa [112] les classent en deux types : locales (par exemple, le LBP (en anglais pour Local Binary Patterns) [88]) et fréquentielles (les transformées en rondelettes [74], filtres de Gabor [75]. Les méthodes d'extraction et d'analyse à base de texture peuvent être divisées en quatre

classes [19, 117] :

— Les méthodes statistiques analysent la distribution spatiale des valeurs de niveaux de gris par le calcul des indices locaux dans l'image et déduisent par la suite un ensemble de statistiques. La Matrice de cooccurrence du niveau de gris (GLCM (en anglais pour Grey Level Co-occurrence Matrix)) [44] est l'une des méthodes statistiques de segmentation à base de texture fréquemment citée.

— Les méthodes géométriques sont utilisées pour décrire les motifs complexes et déduire les propriétés des textures. Les descripteurs de texture peuvent être extraits, par exemple, par une différence de gaussiennes [116]. Ces méthodes permettent de caractériser les propriétés géométriques des textures et trouver les règles qui régissent leur organisation spatiale.

— Les méthodes à base de modèle estiment un modèle paramétrique en fonction de la distribution d'intensité des descripteurs de texture calculés. Les méthodes à base de modèles probabilistes sont largement utilisées telles que les champs aléatoires conditionnels (CRF (en anglais pour Conditional Random Fields)) [59], les champs aléatoires Markoviens (MRF (en anglais pour Markoviens Random Fields)) [85], les champs aléatoires Markoviens gaussiens (GMRF(en anglais pour Gaussiens Markoviens Random Fields)) [18], les fractales [26], LBP,etc.

— Les méthodes fréquentielles analysent les fréquences de l'image. Les méthodes fréquentielles les plus utilisées sont les filtres de Gabor [51], la transformée de Fourier [101], les transformées en rondelettes [74] et les méthodes de segmentation à base de calcul de moments [115]. Certaines approches étudient les propriétés locales de l'image analysée (GMRF, LBP, etc.). D'autres méthodes sont basées sur des représentations statistiques et/ou spatiales et/ou fréquentielles (les transformées en rondelettes, les filtres de Gabor, etc.

Un aperçu complet des dernières techniques d'extraction d'indices de texture

pour la segmentation d'images est présenté dans [97] incluant les filtres de Gabor, GLCM, fractales,etc. Qiao et al [95] combinent les filtres de Gabor, les ondelettes et les méthodes à base de noyau pour la segmentation d'images de documents. Nourbakhsh et al [86] évaluent deux catégories de descripteurs de texture extraits des filtres de Gabor et des ondelettes pour séparer les zones textuelles des régions non textuelles dans des images de documents. Dans cet article, trois catégories de descripteurs de texture sont extraites : la fonction d'autocorrélation, les matrices de co-occurrence des niveaux de gris et les filtres de Gabor.

3.2.2.4 Descripteurs basés sur les distributions

Les méthodes proposées dans cette catégorie utilisent des histogrammes pour représenter les différentes caractéristiques d'apparence ou de forme d'une région locale. Le descripteur le plus simple à développer est l'histogramme qui décrit la distribution des intensités de pixels. Cependant, et contrairement à ce qu'on recherche d'un descripteur pour la reconnaissance d'objet, un tel descripteur n'est pas du tout invariant aux petites transformations géométriques (translation par exemple), ni aux changements de conditions d'illumination ou aux autres variations couramment rencontrées dans les images (bruit, occultation, etc.). Par conséquent, la plupart des descripteurs dans cette famille se reposent sur des histogrammes de type orientation de gradients ou les ondelettes de Haar [90, 128]. Le descripteur le plus connu est le « Scale Invariant Feature Transform » ou SIFT. Il a été proposé par Lowe en 2004 et il est le descripteur local le plus utilisé dans les systèmes de reconnaissance d'objets [78]. Dans cette partie on va présenter brièvement les descripteurs basés sur les distributions les plus utilisées, SIFT, HOG et SURFS.

— **SIFT**

Le descripteur SIFT [69, 70] propose de représenter la géométrie locale autour des points-clés par des histogrammes d'orientation du gradient. La Figure (3.2) illustre le calcul du descripteur. Pour un point-clé (x, y, σ, θ), on sélectionne l'image de la pyramide de différence des gaussiens (DoG (en angais pour Difference of

Gaussians)) correspondant à l'échelle σ et on calcule la norme et l'orientation du gradient sur un voisinage autour de (x,y). Afin d'assurer l'invariance à la rotation, les coordonnées des pixels et les valeurs d'orientations du gradient subissent une rotation de θ autour du point-clé. Un patch centré autour de (x,y)est ensuite extrait, de la même façon que les régions d'intérêt normalisées présentées précédemment. La taille des patchs est fixe ici $(16 \times 16\, pixel)$, mais comme les images de la pyramide DOG sont sous-échantillonnée en fonction de σ, la taille de la zone d'intérêt dépend effectivement de l'échelle caractéristique. Les valeurs de la norme sont pondérées à l'aide d'une fenêtre gaussienne centrée en (x,y) et de paramètre égal à la moitié de la largeur du patch. Ceci permet de donner plus de poids aux pixels proches du point-clé. Le patch est ensuite divisé en 16 secteurs de taille 4×4 pixels. Sur chacun de ces secteurs, un histogramme d'orientation du gradient pondéré par la norme est construit avec 8 directions. Un schéma du calcul du descripteur SIFT est représenté Figure 3.2. Ces 16 histogrammes sont concaténés en un vecteur, par la suite normalisé à l'unité, que l'on nomme descripteur SIFT.

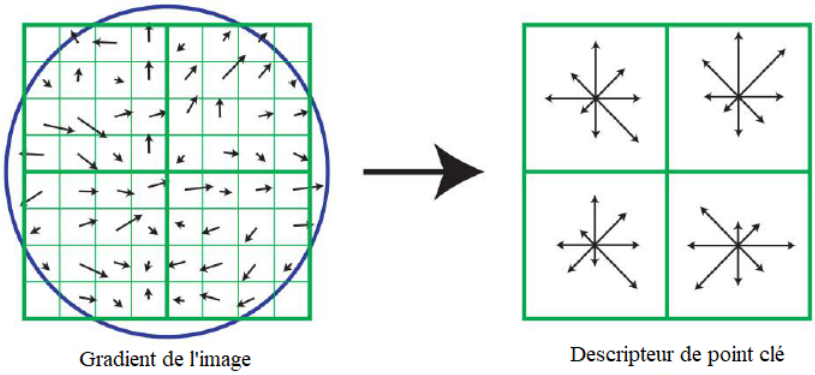


FIGURE 3.2 – principe de calcul des descripteurs SIFT (source : [70])

Cette configuration permet de gérer les effets de bords liés à de faibles erreurs pour la détection de point-clé. En effet si deux points-clés présentent un faible décalage de quelques pixels et/ou une faible erreur d'orientation, l'utilisation d'histogrammes calculés sur un bloc de pixels permet d'atténuer ces imprécisions, contrairement à l'utilisation directe des patchs. La normalisation du vecteur est nécessaire

pour obtenir une invariance aux changements affins de contraste sur la région d'intérêt. Ce descripteur est donc bien invariant par translation, changement d'échelle, rotation et changement d'illumination localement affine.

— **SURF**

Le descripteur SURF (Speeded Up Robust Features) [7] est inspiré du descripteur SIFT, mais présente un temps de calcul optimisé. Il repose sur un calcul d'ondelettes de Haar plutôt que sur un calcul du gradient. Pour chaque point-clé, une région d'intérêt normalisé de taille 10σ est extrait sur I. Les réponses en ondelettes de Haar $2D$ pour la direction horizontale, D_x, et la direction verticale, D_y, sont calculées avec une taille d'ondelette de 2σ. Les réponses D_x et D_y sont pondérées par une gaussienne centrée en (x,y) et de paramètre 3.3σ. La région d'intérêt est ensuite divisée en 4×4 sous-régions. Sur chacune d'elles est extrait un vecteur de 4 dimensions : $v = \left(\sum D_x, \sum D_y, \sum |D_x|, \sum \left|D_y\right|\right)$. Tous ces vecteurs sont rassemblés dans un même vecteur de dimensionnalité 64. Ce vecteur est ensuite normalisé à l'unité pour obtenir l'invariance au contraste. Le vecteur obtenu est appelé descripteur SURF.

— **HOG (Histogramme du Gradient Orienté)**

Le descripteur HOG a été introduit par Dalal et Triggs [23, 24]. L'idée principale de ce descripteur est que l'apparence locale et la forme de l'objet dans une image peuvent être décrites par la distribution d'intensité des gradients ou la direction des contours.

Dans l'article original, le descripteur HOG est proposé pour la détection des piétons (humains) et plus tard, de nombreux chercheurs les ont utilisées pour détecter d'autres objets tels que des voitures, des chiens, des chats, etc. Dans [63], les auteurs montrent comment le temps de prédiction peut être réduit pour la détection de voiture. Dans [67], une approche HOG plus rapide pour la détection des voitures en détectant la zone d'ombre sous les voitures est proposée. Dans [47], les auteurs présentent en détail l'implémentation de HOG et SVM, pour la détection de personne. Dans [5, 6], et toujours pour la détection de personnes, en utilisant un FPGA

(en anglais pour Field Programmable Gate Arrays), un CPU (en anglais pour Central Processing Unit) et un GPU (en anglais pour Graphics Processing Unit) dans une architecture de pipeline, les auteurs présentent une autre implémentation efficace de la même suite d'algorithmes. L'implémentation détaillée de ce descripteur va être présentée dans le chapitre 05.

3.2.2.5 Descripteurs basés sur les Filtres

Les filtres renvoient une image en sortie : chaque pixel ou région de pixels est donc traité. Les descripteurs d'images de type filtre sont très variés.

— **Les filtres binaires et les filtres de rang**

Les filtres binaires et les filtres de rang ont pour point commun de ne pas se focaliser sur les valeurs des pixels, mais plutôt sur les relations existent entre elles (relation logique ou relation d'ordre), sur parle de descripteur non paramétrique. Cette approche est donc originale et différente des autres descripteurs de type filtre. Les descripteurs binaires utilisent un test logique pour créer des valeurs ou des vecteurs de description. Certains comparent par exemple des couples de pixels sur un patch donné (par example, BRIEF (en anglais pour binary robust independent elementary features), LBP, Census, FAST(en anglais pour Features from Accelerated Segment Test)). Le vecteur de description est alors composé des valeurs binaires résultant de ces tests. Les filtres de rang remplacent quant à eux les valeurs des pixels par leur rang dans une fenêtre donnée (par example, Rank Transform).

— **Les filtres de contours**

Les filtres de contours sont les plus employés. Parmi ces derniers, on retrouve des filtres appliqués sur la dérivée première (par example, GradX, GradY, Sobel, Canny, et d'autres variantes [24]), des filtres appliqués sur la dérive seconde (par exemple, Laplacien), des filtres appliqués sur des dérivées multiples ou d'autres méthodes moins courantes. Alors que le gradient classique (GradX, GradY) consiste simplement à déterminer la dérivée de l'image par différence de pixels de proche en proche, les deux autres filtres (Sobel, Canny), par effectuer un lissage de l'image

avant d'effectuer l'opération de dérivation. Ce filtrage a pour but d'éliminer les faux contours. Le filtre de Canny, datant de 1986 et encore plus développé que le filtre de Sobel de 1968, effectué ensuite une opération de sélection des maximas et de seuillage. Cette étape supplémentaire permet une fois encore d'éliminer les faux contours au bruit de la caméra lors de la capture.

— **Les filtres morphologiques**

Les filtres morphologiques sont basés sur la morphologie mathématique. Ils comprennent ainsi parmi d'autres : érosion, dilatation, ouverture, fermeture, haut de forme (top hat), bas de forme (bottom hat), chapeau haut de forme (hit or miss). Toutes ces transformations ne seront pas décrites car elles sont peu utilisées dans le contexte routier.

— **Les filtres de gabarit**

Les filtres de gabarit correspondant à une mesure de corrélation avec une forme particulière appliquée à chaque point de l'image [15, 64]. Ainsi, cette méthode peut s'apparenter à un filtrage avec un filtre représentant la forme recherchée et une mesure de corrélation à la place d'une convolution. Ces méthodes peuvent s'employer sur n'importe quelle forme. C'est le cas sur des exemples de détecteurs de piétons ou de véhicules. Les formes sont alors souvent générées à partir d'un modèle projeté sur l'image (correspondance de modèle) ou apprennent à partir des temples tirés de bases de données (correspondance de forme).

— **Les filtres de pièces**

Les filtres de pièces sont également très utilisés. Ces descripteurs sont identifiés sur la dérivée première (par exemple, Harris, KLT(Kanade Lucas Tomasi)), sur la dérivée seconde ou encore sur l'utilisation directe de l'intensité avec un filtre de pièces spécifiques. La plupart de ces descripteurs reprennent des idées du descripteur de Harris. Ce descripteur classique est en général utilisé à la base d'un algorithme de détection des points d'intérêt. L'avantage du descripteur de Harris est invariant aux transformations euclidiennes [38, 66]. C'est donc un descripteur robuste dans de nombreux cas rencontrés en contexte routier ou en traitement d'images en général.

Le descripteur KLT peut aussi être mis en avant car il est souvent employé. Il se trouve que le descripteur utilisé dans le module de détection de l'algorithme de suivi KLT, décrit dans cette partie, est similaire à celle de Harris même si sa justification est différente [113]. Les descripteurs de pièces enregistrées sur la dérivée seconde sont aussi nombreux, mais ils sont en général moins employés. Enfin, certains descripteurs de pièces utilisent des filtres particuliers ou d'autres méthodes. Moravec [82] propose ainsi une fonction directement sur l'intensité pour détecter les pièces. C'est cette idée qui été d'ailleurs, reprendre par Harris plus tard [38]. Le descripteur SUSAN et ces variantes [113] consistent à calculer le nombre de pixels plus clairs ou plus sombres lors du passage d'un disque sur le pixel d'intérêt afin de déterminer si ce dernier est une pièce ou non. Il faut remarquer que le descripteur SUSAN, par sa méthode, se rapproche aussi des filtres binaires.

— **Les filtres d'apprentissage**

Les descripteurs d'images enregistrées sur un apprentissage sont fondés sur un principe complètement différent des autres. L'idée est de créer un filtre par un apprentissage sur une base de données. Le retour attendu des descripteurs est connu, ce qui permet de créer un filtre sur mesure et optimisé pour l'application choisie. La première version du filtre par apprentissage est basée sur une étude du fonctionnement des mécanismes de détection humaine. L'objectif était de reprendre les mécanismes de détection visuelle mis en œuvre chez l'homme, ce qui a permis de concevoir les LRF (en anglais pour Local Receptive Fields). Dans les LRF, les valeurs d'une fenêtre de pixels sont entrées dans un réseau de neurones. Chaque pixel est ainsi relié aux autres selon plusieurs canapés successifs de neurones. Un apprentissage sur une base de données permet de pondérer le réseau de neurones qui se spécialise alors pour la tâche à laquelle il est entraîné. De nombreux autres filtres par apprentissage ont ensuite été développés (par exempl, TDNN (en anglais pour Time-Delay Neural Network)). Il faut remarquer que les méthodes d'apprentissage profondes peuvent s'apparenter à l'utilisation des filtres par apprentissage. En effet, les premières couches des réseaux de neurones convolutionels – ou CNN (en

anglais pour Convolutional Neural Network) correspondant bien à des filtres optimisés pour une application donnée. Une structure inversée des couches supérieures reprend la structure d'un algorithme de classification.

— **Les filtres de Gabor**

Les filtres de Gabor sont utilisés en tant que descripteurs dans de nombreuses variantes, mais les plus courantes restent celles de la norme MPEG-7 (en anglais pour Moving Picture Experts Group) [76, 98, 133]. Les filtres de Gabor sont des filtres linéaires, c'est-à-dire appliqués par une convolution, qui sont composés d'une composante sinusoïdale et d'une composante gaussienne. Ces filtres correspondent à une pondération par une fonction gaussienne dans le domaine fréquentiel.

3.2.2.6 Descripteurs couleur

La couleur fournit des informations puissantes pour la reconnaissance d'objets. Un schéma de reconnaissance simple et efficace consiste à représenter et à faire correspondre les images sur la base d'histogrammes de couleurs comme proposés par Swain et Ballard [109]. Le travail apporte une contribution significative à l'introduction de la couleur pour la reconnaissance d'objets. Cependant, il présente l'inconvénient que lorsque les circonstances d'éclairage ne sont pas égales, la précision de reconnaissance d'objet se dégrade de manière significative. Cette méthode est étendue par Funt et Finlayson [32], basée sur la théorie de Retinex de Land [61], pour rendre la méthode d'éclairage indépendante en indexant sur l'éclairage des descripteurs de surface invariants (rapports de couleurs) calculés à partir de points voisins. Cependant, on suppose que les points voisins ont la même normale de surface. Par conséquent, les descripteurs de surface invariants à l'éclairage dérivé sont négativement affectés par des changements rapides d'orientation de la surface de l'objet (c'est-à-dire la géométrie de l'objet). Healey et Slater [46] et Finlayson et al. [27] utilisent des moments invariants d'éclairage de distributions de couleurs pour la reconnaissance d'objets. Ces méthodes sont sensibles à l'occlusion et à l'encom-

brement des objets, car les moments sont définis comme une propriété intégrale sur l'objet comme un seul. Dans les méthodes globales, en général, les parties occluses perturberont la reconnaissance. Slater et Healey [105] contournent ce problème en calculant les caractéristiques de couleur à partir de petites régions d'objets au lieu de l'objet entier.

À partir des observations ci-dessus, le choix des modèles de couleur à utiliser ne dépend pas seulement de leur robustesse face à un éclairage variable à travers la scène (par exemple, plusieurs sources lumineuses avec différentes distributions spectrales de puissance spectrale), mais aussi de leur robustesse contre les changements de l'orientation de la surface de l'objet (c'est-à-dire la géométrie de l'objet) et leur robustesse contre l'occlusion et l'encombrement de l'objet. De plus, les modèles de couleurs doivent être concis, discriminatoires et résistants au bruit. Par conséquent, dans cette partie, notre objectif est de présenter différents descripteurs de couleurs à utiliser à des fins de reconnaissance d'objets multicolores. Tout d'abord, les descripteurs de couleur basés sur des histogrammes sont présentés. Ensuite, les moments de couleur et les moments de couleur invariants sont présentés.

— **Histogrammes de couleur**

L'histogramme couleur reste la référence des descripteurs couleur utilisés aujourd'hui grâce à sa quantification systématique de l'information couleur. L'histogramme de l'image couleur fait référence à la probabilité conjointe des trois canaux de couleur. Il est défini par :

$$h_{A,B,C}(a,b,c) = N.Prob(A=a,B=b,C=c) \tag{3.17}$$

Où A, B et C représentent les trois canaux couleur (R, G et B ou H, S et V, etc.) et N est le nombre de pixels dans l'image.

— **Histogramme RGB**

L'histogramme RGB est une combinaison de trois histogrammes 1-D basés sur les canaux R, G et B de l'espace colorimétrique RGB [57].

— **Histogramme de couleur-structure CS**

Le descripteur de couleur-structure CS étend et enrichit la notion d'histogramme en introduisant dans la représentation un minimum d'information spatiale locale. Un élément structurant, définis un masque binaire, est translaté en chaque pixel de l'image. Tous les intervalles de l'histogramme correspondant aux couleurs présentent à l'intérieur du masque sont alors incrémentés. Ainsi, l'histogramme CS représente-t-il la fréquence relative des éléments structurants contenant une couleur donnée. Les mesures de similarités adoptées pour l'histogramme CS sont la distance Minkowski et la distance euclidienne [92].

Pour assurer une certaine invariance par rapport aux homothéties, les images sont normalisées à une taille fixe avant l'extraction du descripteur.

— **Histogramme rg**

Dans le modèle de couleur RGB normalisé, les composantes de chromaticité r et g décrivent les informations de couleur dans l'image (b est redondant comme $r+g+b=1$) :

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \\ \frac{B}{R+G+B} \end{pmatrix} \tag{3.18}$$

En raison de la normalisation, r et g sont invariables à l'échelle et donc invariants aux changements d'intensité lumineuse, aux ombres et à l'ombrage [36].

— **Distribution des couleurs transformée**

L'histogramme RGB n'est pas invariant aux changements des conditions d'éclairage. Cependant, en normalisant les distributions de valeurs de pixels, l'invariance d'échelle et l'invariance de décalage sont obtenues par rapport à l'intensité lumineuse. Parce que chaque canal est normalisé indépendamment, le descripteur est également normalisé contre les changements de couleur de lumière et les déca-

lages arbitraires :

$$\begin{aligned} R' &= \frac{R-\mu_R}{\sigma_R} \\ G' &= \frac{G-\mu_G}{\sigma_G} \\ B' &= \frac{B-\mu_B}{\sigma_B} \end{aligned} \tag{3.19}$$

Avec μ_C la moyenne et σ_C l'écart-type de la distribution dans le canal C. Cela donne pour chaque canal une distribution où $\mu = 0$ et $\sigma = 1$ [57].

— **Moments statistiques**

La méthode d'histogramme utilise la distribution complète de la couleur. On doit stocker de nombreuses données. Au lieu de calculer la distribution complète, dans les systèmes de recherche d'images, on calcule seulement des caractéristiques dominantes de couleur telle que l'espérance, la variance et d'autres moments.
En mathématiques, nous calculons des moments sur des distributions ou des fonctions jusqu'à un ordre arbitraire. Le premier moment est égal à la moyenne d'une distribution. Le deuxième moment est égal à la variance de la distribution. Le troisième moment est appelé asymétrie après l'inclinaison d'une distribution.
Une image couleur correspond à une fonction mathématique définissant des triplets RGB pour les positions d'image (x,y) : $I : (x,y) \rightarrow (R(x,y), G(x,y), B(x,y))$. En considérant les triplets RGB comme des points de données provenant d'une distribution, il est possible de définir des moments. Mindru et al [81] ont défini des moments de couleur généralisés M_{pq}^{abc} :

$$M_{pq}^{abc} = \iint x^p y^q [I_R(x,y)]^a [I_G(x,y)]^b [I_B(x,y)]^c dxdy \tag{3.20}$$

M_{pq}^{abc} est appelé un moment d'ordre de couleur généralisé $p+q$ et degré $a+b+c$. Notez que les moments d'ordre 0 ne contiennent aucune information spatiale, tandis que les moments de degré 0 ne contiennent aucune information photométrique. Ainsi, les descriptions des moments d'ordre 0 sont invariantes en rotation, tandis que les ordres supérieurs ne le sont pas. Un grand nombre de moments peuvent

être créés avec de petites valeurs pour l'ordre et le degré. Cependant, pour des valeurs plus grandes, les moments sont moins stables. Des moments de couleur généralement généralisés jusqu'au premier ordre et au deuxième degré sont utilisés.

En utilisant la bonne combinaison de moments, il est possible de se normaliser contre les changements photométriques. Ces combinaisons sont appelées des moments de couleur invariants. Les invariants n'impliquant qu'un seul canal de couleur (par exemple sur a, b et c deux sont 0) sont appelés invariants à 1 bande. De même, il existe des invariants à 2 bandes impliquant seulement deux bandes de couleurs sur trois. Les invariants à 3 bande impliquent tous les canaux de couleur, mais ceux-ci peuvent toujours être créés en utilisant des invariants à 2 bandes pour différentes combinaisons de canaux.

— **Moments de couleur**

Le descripteur moment de couleur utilise tous les moments de couleur généralisés jusqu'au deuxième degré et au premier ordre. Cela a conduit à neuf combinaisons possibles pour le degré : M_{pq}^{000}, M_{pq}^{100}, M_{pq}^{010}, M_{pq}^{001}, M_{pq}^{200}, M_{pq}^{110}, M_{pq}^{020}, M_{pq}^{011}, M_{pq}^{002} et M_{pq}^{101}. Combiné avec trois combinaisons possibles pour la commande : M_{00}^{abc}, M_{10}^{abc} et M_{01}^{abc}, le descripteur de moment de couleur à 27 dimensions. Ces moments de couleur n'ont qu'une invariance de décalage. Ceci est obtenu en soustrayant la moyenne de tous les canaux d'entrée avant de calculer les moments.

— **Moments de couleur Invariants**

Les moments de couleur invariants peuvent être construits à partir de moments de couleur généralisés. Tous les invariants à 3 bandes sont calculés à partir de Mindru et al. [81]. Pour être comparables, les invariants $\tilde{C}_{02}$ sont considérés. Cela donne un total de 24 invariants de moment de couleur.

— **Descripteurs basés sur les moments**

Il existe plusieurs descripteurs basés sur les moments en traitement d'images. Ils étaient originellement utilisés pour de la reconnaissance d'objets. L'intérêt d'uti-

liser les moments est qu'ils sont invariants aux translations, aux rotations et aux changements d'échelle isotropes [38]. Au départ, les moments ont été utilisés pour décrire des histogrammes (par exemple, moyenne, variance). Cela permettait de réduire la taille des descripteurs employés et de les rendre plus robustes. En effet, un histogramme de taille importante peut être résumé par quelques moments. C'est le cas de la matrice de co-occurrence [43, 44]. Bien que des variantes de cette matrice aient ensuite été utilisées directement comme descripteur, la matrice de co-occurrence n'était au départ qu'une étape de la description. Le descripteur proposé à l'origine repose en effet sur 14 moments calculés sur la matrice de co-occurrence. Il existe aussi des descripteurs d'images pour lesquels les moments sont calculés directement sur les images. Les plus connus sont les moments de Zernike dont plusieurs implémentations existent [21, 110] et les moments de Hu [48].

— **Vecteur de Cohérence de couleurs**

Le vecteur de cohérence de couleurs CCV représente une autre variante, plus détaillée, de l'histogramme de couleurs. Il a été proposé par Pass [91]. Dans cette technique, chaque rang de l'histogramme peut être partitionné en deux catégories :

— Cohérent, s'il appartient à une région de couleur uniforme,

— Incohérent, sinon.

On note α_i le nombre de pixels cohérents dans le $i^{éme}$ rang de couleur et β_i le nombre de pixels incohérents. Le CCV d'une image est défini par le vecteur $[(\alpha_1,\beta_1),(\alpha_2,\beta_2),...,(\alpha_N,\beta_N)]$, tel que la somme : $(\alpha_1+\beta_1,\alpha_1+\beta_1,...,\alpha_N+\beta_N)$ donnera l'histogramme de couleurs de l'image. L'avantage de cette approche réside dans l'ajout de l'information spatiale à l'histogramme et cela à partir de leur raffinement, mais elle présente l'inconvénient d'amplifier la sensibilité aux conditions d'illumination.

— **Distribution spatiale de couleurs**

Ce descripteur vise à capturer la disposition spatiale (color layout) des couleurs dans une image. L'image est tout d'abord divisée en 64 $(8*8)$ blocs rectangulaires. Chaque rectangle est représenté par sa couleur dominante, qui est par définition la

moyenne des couleurs des pixels constituant le rectangle considéré. On construit ainsi trois matrices $(8*8)$, une par composante de couleur. Les coefficients quantifiés des transformées en cosinus discret $2D$ (DCT (en anglais pour Discrete Cosine Transform)) de chacune de ces matrices sont ensuite calculés. Ces coefficients sont parcourus en zigzag, à partir des fréquences les plus basses [140].
La mesure de similarité recommandée pour ce descripteur est une distance euclidienne pondérée entre les coefficients DCT ainsi déterminés. Notons que ce descripteur n'est pas invariant aux rotations. Son utilisation est donc pertinente uniquement pour des applications concernant des requêtes globales, où le positionnement des différentes couleurs dans l'image doit être pris en compte.

3.2.3 Descripteurs multiples

Parmi les différents descripteurs que nous avons déjà présentés, on peut conclure que chaque descripteur permet de décrire un objet selon un caractère particulier. D'un autre côté, les objets cibles peuvent posséder des caractéristiques similaires telles que les habits vestimentaires, dans ce cas par exemple, l'histogramme de couleurs ne peut pas être un modèle discriminant, donc le choix du descripteur dépend de plusieurs facteurs à savoir la nature de l'objet cible, la relation entre les objets à suivre, le facteur de luminosité et d'échelle, etc.

Afin d'obtenir la modélisation la plus robuste possible, certains travaux ont utilisé plusieurs descripteurs combinés pour décrire la région de l'objet cible, prenons comme exemple, les auteurs dans [73], qui ont proposé un modèle d'apparence basé sur la combinaison de l'histogramme pondéré de couleurs et l'histogramme d'orientations du gradient. Un objet cible est approximé par une ellipse. Pour l'histogramme de couleurs, un noyau elliptique est appliqué afin de favoriser les pixels qui sont proches du centre. Par la suite, l'histogramme de couleurs est normalisé. Un vecteur de probabilité est ainsi obtenu pour chaque histogramme. Il est intégré dans un processus de suivi basé sur l'algorithme du Filtre de Particules.

Dans un autre article Possegger et al. [94], la modélisation de l'objet cible se fait

en exploitant certaines caractéristiques géométriques. Pour chaque objet i à l'emplacement x, l'information sur l'occultation $c_{o,i}(x)$ (l'évolution spatio-temporelle des régions où il y a des occultations), la fiabilité du détecteur d'objet $c_{d,i}$ et la prédiction du mouvement de l'objet cible $c_{p,i}$ doivent être estimées. Le score de confiance est le produit de ces trois facteurs :

$$\varphi(x) = c_{o,i}(x)\, c_{d,i}(x)\, c_{p,i}(x) \tag{3.21}$$

Cette approche de suivi a prouvé que les propriétés géométriques peuvent aider à gérer l'occultation entre les objets cibles.

Les auteurs dans Yao et al. [136] proposent une méthode du suivi pour des vidéos de sports. Vu la complexité du modèle de joueur, ce dernier est représenté par une combinaison de deux modèles d'apparence : un modèle statique et un modèle dynamique. Le modèle statique permet de coder l'information de couleur et de texture. Les auteurs ont utilisé un histogramme de couleur HSV et un descripteur LBP. Pour chaque particule pf_t^i, la mesure de similarité statique correspondant au descripteur f est le suivant :

$$V\left(pf_t^i, f\right) = 1 - BC\left(h_T^f, h^f\left(pf_t^i\right)\right) \tag{3.22}$$

Où : BC est le coefficient de Bhattacharyya, h_T^f est le descripteur de modèle de l'objet cible et h^f est le modèle de la particule (objet candidat). Le modèle dynamique est constitué de la position géométrique, la vélocité et le flux optique. Vu le mouvement des athlètes, le déplacement de la cible suit une transition gaussienne. La vélocité est une mixture pondérée entre le flux optique et la vélocité calculée dans la trame précédente. Ces descripteurs sont combinés à l'aide d'un vecteur de coefficient de poids.

Dans Possegger et al. [94], l'objet cible est décrit en utilisant : l'histogramme de couleur, les matrices de covariance et l'histogramme de gradients HOG. La combinaison de ces descripteurs est faite en utilisant l'algorithme Adaboost.

Les auteurs dans Breitenstein et al. [14] exploitent aussi l'idée de combiner plusieurs descripteurs dans le cadre d'un traqueur Filtre de Particules. En fait, l'objet cible est représenté par deux caractéristiques : la caractéristique géométrique (la taille et la position) et la caractéristique de mouvement. Pour la taille, un objet cible possède une taille moyenne sur un nombre de trames (quatre dernières trames). Le modèle de mouvement est basé sur l'hypothèse que les objets cibles ont une vitesse de déplacement constante durant toute la séquence vidéo. De ce fait, la position courante de l'objet cible est égale à la position de l'objet plus la vitesse de déplacement estimée à la trame précédente. Après le calcul de mesure de similarité (en fonction des descripteurs mentionnés ci-dessous), il faut définir le poids de chaque particule. Ce dernier est estimé en fonction de trois termes de confiance (la somme) qui définissent eux-mêmes un modèle d'apparence pour la particule pf associé à l'objet cible tr. Premièrement, le terme de détection qui consiste à calculer la distance entre la particule et l'objet candidat associé. Ce terme est pris en considération seulement dans le cas de la mise en correspondance entre l'objet candidat et l'objet cible. Deuxièmement, le terme de confiance du détecteur d'objets (estimé en fonction de la densité de confiance du détecteur à la position de la particule en question). Troisièmement, le terme de classificateur (estimé en utilisant les caractéristiques de couleur et de texture associées à la particule). Ces caractéristiques sont ainsi combinées en calculant la somme pondérée par des poids. Ces poids sont déterminés expérimentalement.

Une autre approche similaire a été développée par Yan et al. [134], les auteurs ont proposé un modèle d'apparence en combinant plusieurs caractéristiques à savoir : l'histogramme de couleur, les caractéristiques de mouvement (incluant la vélocité de mouvement, l'échelle de l'objet et l'angle de déplacement de l'objet) et le flux optique (l'histogramme de mouvement basé sur la magnitude et l'angle du vecteur flux optique). Chaque score de similarité obtenu par une de ces caractéristiques est pondéré par un poids afin de calculer le score de similarité globale. Le poids est déterminé via un processus d'apprentissage discriminatif.

Dans Yang et al. [135], les auteurs utilisent un modèle d'apparence basé sur la combinaison de plusieurs caractéristiques dans le cadre du suivi multi objets, mais d'une manière différente. En fait, des modèles d'apparence différents sont utilisés pour représenter des parties particulières du corps humain. Un histogramme de couleur pondéré est calculé pour la partie supérieure du torse (incluant la tête). L'historique de chaque histogramme calculé est utilisé pour calculer le score de similarité (deux modèles d'histogramme moyen sont sauvegardés : à court terme et à long terme). La partie de la tête est représentée par une forme elliptique qui est modélisée par le vecteur d'intensité du gradient. Finalement un ensemble de caractéristiques locales (descripteur SIFT) est estimé pour la partie inférieure du torse. Les points du type SIFT sont estimés sur une grille carrée de pixels de taille 4. Un histogramme SPM (en anglais pour Spatial Pyramid Matching) est calculé pour les points. En gros, chaque partie est représentée par un modèle différent dépendamment de sa localisation et de son importance dans la modélisation globale de l'objet cible. Les détections sont obtenues en utilisant un détecteur de tête humaine basé sur le réseau de neurones CNN sur des séquences vidéo multi vues. Les scores de similarité sont linéairement combinés pour obtenir le score de similarité global.

Dans Kuo and Nevatia [58], l'objet cible est représenté en utilisant un descripteur de la couleur (histogramme de couleur RGB), un descripteur de la forme (histogramme HOG) et un descripteur de la texture (matrice de covariance). Contrairement aux autres travaux qui utilisent plusieurs descripteurs, un seul descripteur sera sélectionné pour chaque trame. En fait, ces descripteurs sont appris en utilisant l'algorithme Adaboost afin de sélectionner séquentiellement le meilleur descripteur (c'est le descripteur qui donne la meilleure valeur de similarité). Selon Kuo and Nevatia [58], l'histogramme de couleur est le plus souvent sélectionné comme le meilleur descripteur.

Une autre approche récente, Milan et al. [80], est basée sur une fonction de coût obtenue à l'aide d'une représentation complète de l'objet cible. Outre les descripteurs usuels de la modélisation d'un objet, d'autres descripteurs sont estimés. La

fonction de coût (la probabilité de similarité) est estimée en fonction : du terme de donnée qui sert à garder les trajectoires proches des objets candidats, du terme dynamique qui consiste à estimer la contrainte de mouvement basée sur une vitesse de déplacement constant, du terme de l'occultation mutuelle qui est une pénalité de continuité afin de gérer les cas d'occultation entre deux objets cibles, du terme de persistance de trajectoire qui permet d'éviter toutes les fragmentations ou les terminaisons abruptes de trajectoires et finalement du terme de régularisation qui sert à contrôler le nombre de trajectoires crées. En plus des termes définis ci-dessus, l'histogramme gaussien pondéré a été intégré pour créer le modèle d'apparence. Cette pondération a pour but de favoriser les pixels les plus importants (qui appartiennent à l'objet cible) et de défavoriser les pixels de bruit (qui appartiennent à l'arrière-plan). Pour le détecteur d'objets, un histogramme de gradients et un histogramme de flux optiques sont utilisés afin de construire le modèle d'apparence pour le processus de détection. Le modèle d'apparence basé sur le mouvement a été largement utilisé.

Dans Yoon et al. [137], le modèle d'apparence de mouvement est basé sur la relation dynamique entre les objets. Ainsi, un réseau de mouvements relative RMN (en anglais pour Relative Motion Network) est construit. Ce dernier permet de sauvegarder les relations spatiales et dynamiques entre les objets en se basant sur la différence entre la vélocité et la position géométrique. Par la suite, pour chaque objet cible, un vecteur de mouvement est obtenu qui contient la relation du mouvement par rapport à chaque autre objet cible dans la séquence vidéo. Le score de similarité est ainsi estimé en comparant la similarité entre les modèles de mouvement relatif des objets cibles. Cette modélisation est intégrée au sein d'un processus de suivi basé sur le filtre bayésien. En plus du modèle de mouvement, l'association des données est réalisée en utilisant la taille de l'objet comme modèle d'apparence ainsi que l'apparence de couleur (histogramme de couleurs).

Bosch et al. [13] calculer les descripteurs SIFT sur les trois canaux du modèle de couleur HSV, plutôt que sur le canal d'intensité uniquement. Cela donne dimensions par descripteur, 128 par canal. L'inconvénient de cette approche est que l'instabilité

de la teinte pour une faible saturation est ignorée. Les propriétés des canaux H et S s'appliquent également à ce descripteur : il est invariant à l'échelle et invariant au décalage. Cependant, les descripteurs H et S SIFT ne sont pas invariants aux changements de couleur de la lumière, seul le descripteur SIFT d'intensité (canal V) est invariant à cela. Par conséquent, le descripteur n'est que partiellement invariant aux changements de couleur de la lumière.

Van de Weijer et al. [122] introduisent une concaténation de l'histogramme de teinte avec le descripteur SIFT. Par rapport à HSV-SIFT, l'utilisation de l'histogramme de teinte pesée corrige l'instabilité de la teinte autour de l'axe gris. Parce que les cases de l'histogramme de teinte sont indépendantes, il n'y a aucun problème avec la périodicité du canal de teinte pour HueSIFT. Semblable à l'histogramme de teinte, le descripteur HueSIFT est invariant à l'échelle et invariant au décalage. Cependant, seul le composant SIFT de ce descripteur est invariant aux changements ou décalages de couleur d'éclairage, l'histogramme de teinte ne l'est pas.

[57] décrit tous les canaux dans l'espace colorimétrique de l'adversaire à l'aide de descripteurs SIFT(OpponentSIFT). Les informations dans le canal $O3$ sont égales aux informations d'intensité, tandis que les autres canaux décrivent les informations de couleur dans l'image. Cependant, ces autres canaux contiennent des informations d'intensité : ils ne sont donc pas invariants aux changements d'intensité lumineuse.

Dans l'espace colorimétrique de l'adversaire, les canaux $O1$ et $O2$ contiennent encore des informations d'intensité. Pour ajouter l'invariance aux changements d'intensité, [35] propose l'invariant W qui élimine les informations d'intensité de ces canaux. Le descripteur W-SIFT utilise l'invariant W, qui peut être défini pour l'espace colorimétrique de l'adversaire comme O_1/O_3 et O_2/O_3 . En raison de la division par intensité, la mise à l'échelle dans le modèle diagonal s'annulera, ce qui rend l'échelle W-SIFT invariante par rapport à l'intensité lumineuse. Comme pour les autres descripteurs couleur-SIFT, la composante couleur du descripteur n'est pas invariante aux changements de couleur claire.

Pour le descripteur rg-SIFT, des descripteurs sont ajoutés pour les composantes de chromaticité r et v du modèle de couleur RGB normalisé, qui est déjà invariant à l'échelle. Étant donné que le descripteur SIFT utilise des dérivés des canaux d'entrée, le descripteur rv-SIFT devient également invariant par rapport au décalage. Cependant, la partie couleur du descripteur n'est pas invariante aux changements de couleur d'éclairage.

Pour la couleur transformée SIFT, la même normalisation est appliquée aux canaux RGB que pour l'histogramme de couleurs transformées. Pour chaque canal normalisé, le descripteur SIFT est calculé. Le descripteur est invariant à l'échelle, invariant au décalage et invariant aux changements de couleur et au décalage.

3.3 Classification

La classification est le problème de déterminer dans quel ensemble de catégories, une nouvelle observation appartient, sur la base d'un ensemble de données d'apprentissage contenant des observations (ou instances) dont la catégorie est déjà connue. Un exemple serait l'affectation d'un e-mail donné en classe "spam" ou "non-spam", ou l'attribution d'un diagnostic à un patient donné, comme décrit par les caractéristiques observées du patient (sexe, la pression artérielle, la présence ou l'absence de certains symptômes, etc.).

Dans notre thèse, nous avons choisi d'utiliser deux célèbres classificateurs largement utilisés dans le domaine de la reconnaissance d'objets, les classificateurs de machine à vecteur de support et les classificateurs K-voisin le plus proche.

3.3.1 Machine à vecteurs de support

Supposons un couple (x_k, y_k) de variables aléatoires de valeurs dans $R^n \times \{-1, 1\}$, où x_k sont les les descripteurs de teinte pour les images d'entraînement et y_k sont les étiquettes des classes. Le SVM [127] nécessite la résolution du pro-

blème d'optimisation suivant :

$$\min_{w,b,\xi} \frac{1}{2} w^T w + C \sum_{k=1}^{n} \xi_k \tag{3.23}$$

Sous la contrainte :

$$\begin{aligned} y_k &\leq \left(w^T + b\right) \geq 1 - \xi_k \\ \xi_k &\geq 0 \end{aligned} \tag{3.24}$$

où w est le vecteur orthogonal à l'hyper plan, b est le déplacement par rapport à l'origine, ξ_k variables de libération de contrainte et C c'est la variable d'équilibrage.

Dans notre thése nous avons utilisé un SVM multi-classes un contre tous avec une fonction noyau gaussien.

3.3.2 K-voisins les plus proches

L'algorithme k-plus proche voisin [111] est une méthode non paramétrique, lorsqu'une nouvelle instance à classer arrive, elle est comparée aux instances d'apprentissage à l'aide d'une mesure de similarité en calculant les distances entre ces instances. Cette nouvelle instance est affectée à la classe la plus représentée parmi les k sorties associées aux k entrées les plus proches de la nouvelle entrée x.

Il existe plusieurs distances utilisées par l'algorithme KNN pour comparer deux instances. Dans notre thèse, nous avons utilisé la distance euclidienne :

$$D\left(X_p, X_q\right) = \sqrt{\sum_{i=1}^{n} \left(x_P i - x_q i\right)^2} \tag{3.25}$$

$X_i = (x_i 1, x_i 2, ..., x_i n)$ le vecteur caractéristique de l'instance i, n La dimension de X_i, p et q deux instances à comparer.

3.4 Conclusion

Dans ce chapitre, nous avons évoqué les étapes nécessaires pour construire un système de reconnaissance d'objets (figure 3.1). Nous avons présenté les différents travaux proposés pour la détection et l'extraction des caractéristiques la première

étape d'un système de reconnaissance d'objets, dans la section 3.1. Nous avons pu conclure qu'il n'y a pas jusqu'à présent un type de détecteur des caractéristiques universel. Par ailleurs, nous avons présenté les différentes techniques proposées pour réaliser la deuxième étape du système de reconnaissance d'objets : la description des caractéristiques dans la section 3.2.. Dans la section 3.3. nous avons présenté les méthodes de classification utilisée dans notre thèse. Dans le chapitre suivant, nous allons présenter notre méthode proposée pour la catégorisation des objets ainsi que les exceptionnelles performances de cette méthode.

CHAPITRE 4

Méthode proposée et résultats

« Il faut toujours connaître les limites du possible. Pas pour s'arrêter, mais pour tenter l'impossible dans les meilleures conditions.»

Romain Gary

Résumé

Dans ce chapitre, nous décrivons la méthode de construction du descripteur proposé pour la reconnaissance d'objets. Nous présentons ensuite les différentes performances de ce descripteur.

Contents

Reconnaître des objets à partir d'une image est une tâche difficile en vision par ordinateur. De toute évidence, les humains reconnaissent les objets par vision avec une grande précision et peu d'effort, et on ne sait toujours pas comment cette performance parfaite est obtenue. Lors du développement d'un modèle de vision par ordinateur pour reconnaître un objet, des problèmes théoriques difficiles se posent, comme comment modéliser l'apparence visuelle et comment reconnaître des objets. En général, le modèle de reconnaissance d'objets est défini comme l'acquisition d'image, l'extraction des caractéristiques, la description des caractéristiques et la classification.

4.1 Modélisation des changements géométriques

L'équation 4.1 montre que lorsque la surface ou la géométrie de l'éclairage change, les réponses des capteurs changent le facteur d'échelle unique $(\vec{e}\cdot\vec{n})$. Cela signifie que les réponses des capteurs à une surface vue sous deux géométries ou intensités lumineuses différentes sont liées :

$$\frac{R_2}{R_2+G_2+B_2}=\frac{sR_2}{s(R_2+G_2+B_2)} \tag{4.1}$$

Il est simple de rendre les réponses des capteurs indépendantes de l'intensité lumineuse en les divisant par la somme des réponses R, G et B. Cela donne l'espace colorimétrique normalisé RGB :

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} \frac{R}{R+G+B} \\ \frac{G}{R+G+B} \\ \frac{B}{R+G+B} \end{pmatrix} \tag{4.2}$$

Le triple (r,g,b) est invariant aux changements de géométrie et d'intensité d'éclairage, car la somme du triple est toujours égale à 1.

Les transformations géométriques incluent la translation, la rotation, l'isométrie [1],

1. Le changement d'angle et de distance de vue

l'homothétie [2] et la similitude [3].

4.2 Modélisation des changements de condition d'éclairage

Lorsque les conditions d'éclairage d'une scène changent, il devient très difficile d'obtenir des mesures stables. Dans cette partie, nous modéliserons le changement de couleur de la lumière, le changement d'intensité lumineuse, le décalage d'intensité lumineuse et le changement-décalage d'intensité lumineuse.

4.2.1 Changements de couleur d'éclairage

Selon le modèle de Kries [129], les changements de couleur d'éclairage peuvent être modélisés par le mappage diagonal de Von. Cette cartographie diagonale est donnée comme suit :

$$I^c = D^{u,c}.I^u \tag{4.3}$$

où I^u est l'image prise sous une source de lumière inconnue, I^cest la même image transformée, il semble donc qu'elle a été prise sous la lumière de référence (appelée illuminant canonique), et $D^{u,c}$ est une matrice diagonale qui cartographie les couleurs qui sont prises sous une source de lumière inconnue u à leurs couleurs correspondantes sous l'illuminant canonique c :

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \tag{4.4}$$

R est le composant rouge des images dans l'espace colorimétrique RGB, G est le composant vert et B représente le composant bleu. Le changement de couleur

2. Agrandir ou réduire une figure selon un rapport d'homothétie et un centre d'homothétie
3. Comme le parallélisme par exemple

d'éclairage correspond à un changement de couleur de l'éclairage et de la diffusion de la lumière.

4.2.2 Changement d'intensité lumineuse

Pour l'équation 4.4, lorsque les valeurs d'image changent d'un facteur constant dans tous les canaux (c'est-à-dire $a = b = c$), cela équivaut à un changement d'intensité lumineuse :

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} \tag{4.5}$$

Les changements d'intensité lumineuse incluent les ombres et les changements de géométrie d'éclairage tels que l'ombrage. Par conséquent, lorsqu'un descripteur est invariant aux changements d'intensité lumineuse, il est invariant à l'échelle par rapport à l'intensité (lumineuse) [121].

4.2.3 Décalage d'intensité lumineuse

Un décalage égal des valeurs d'intensité des images dans tous les canaux, c'est-à-dire un décalage d'intensité lumineuse, où $(o_1 = o_2 = o_3)$ et $(a = b = c = 1)$ produiront :

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix} \tag{4.6}$$

Les décalage d'intensité lumineuse correspondent aux reflets des objets sous une source de lumière blanche et à la diffusion d'une source blanche. Lorsqu'un descripteur est invariant à un changement d'intensité lumineuse, il est invariant par rapport au décalage d'intensité lumineuse [121].

4.2.4 Changement et décalage d'intensité lumineuse

Les deux classes de changements ci-dessus peuvent être combinées pour modéliser à la fois les changements et les décalages d'intensité lumineuse :

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & a & 0 \\ 0 & 0 & a \end{pmatrix} \begin{pmatrix} R^u \\ G^u \\ B^u \end{pmatrix} + \begin{pmatrix} o_1 \\ o_1 \\ o_1 \end{pmatrix} \tag{4.7}$$

Autrement dit, un descripteur d'image robuste à ces changements est invariant d'échelle et invariant de décalage par rapport à l'intensité lumineuse [121].

4.3 Méthodes de comparaison

Dans cette section, nous présenterons quelques méthodes classiques que sont les histogrammes de l'adversaire, les histogrammes de teinte de Gever, les descripteur HOG et un descripteur d'image local à partir des réponses de filtre d'Even Gabor.

4.3.1 HOG (Histogram of Oriented Gradient)

Le descripteur HOG, largement inspiré du SIFT, a été proposé par Dalal et Triggs en 2005 pour répondre aux limites du SIFT dans le cas des grilles denses [24]. L'idée principale de ce descripteur est que la structure locale de l'objet se caractérise par le calcul de la distribution des gradients des intensités locales ou des directions des contours, sans avoir une connaissance préalable de l'emplacement du gradient ou de la position du contours de l'objet de l'image.

Les bords inférieur et droit de l'image sont découpés pour obtenir une image (I_{bords}). Pour chaque bloc (I_{bords}), une correction gamma est appliquée à l'ensemble de

l'image en utilisant l'équation 4.8 :

$$I_{gamma}(p) = \sqrt{I_{bords}(p)} \tag{4.8}$$

où I_{gamma} est l'image corrigée et p un pixel de l'image.
La dérivée horizontale (respectivement verticale) de l'image, I_x (respectivement I_y) est ensuite calculée par convolution en utilisant le filtre dx (respectivement dy).

$$dx = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \tag{4.9}$$

$$dy = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \tag{4.10}$$

Pour chaque pixel p, la magnitude du gradient $g(p)$ et son orientation $\theta(p)$ sont ensuite calculées en suivant les équations (4.11,4.12).

$$g(p) = \sqrt{I_x(p)^2 + I_y(p)^2} \tag{4.11}$$

$$\theta(p) = \arctan\left(\frac{I_y}{I_x}\right) \tag{4.12}$$

À partir d'ici, l'image est découpée en $N \times M$ cellules de dimension 8×8. Pour chaque cellule, la magnitude du gradient g est pondérée en fonction de la distance du pixel (p) par rapport au pixel du centre de la cellule (c). Un masque de pondération gaussienne d'écart-type $\sigma = 4$ est alors utilisé.

$$g_\sigma(p) = g(p).e^{-\frac{(x_p - x_c)^2 + (y_p - y_c)^2}{2\sigma^2}} \tag{4.13}$$

où le pixel p (respectivement c) a pour coordonnees (x_p, y_p) (respectivement (x_c,y_c)). Un histogramme des orientations des gradients est ensuite déterminé pour chaque cellule. Pour cela, un histogramme contenant 9 orientations est considéré ($[0 \cup 19]$, $[20 \cup 39]$, $[40 \cup 59]$, $[60 \cup 79]$, $[80 \cup 99]$, $[100 \cup 119]$, $[120 \cup 139]$, $[140 \cup 159]$ et

$[160 \cup 179]^o$). Ensuite, chacun des 64 pixels de la cellule vote pour l'orientation du gradient qu'il détient. Il vote alors proportionnellement à sa magnitude dans les deux compartiments encadrant son orientation (au prorata de la distance de chaque compartiment).

Un histogramme contenant 9 compartiments est donc obtenu pour chaque cellule de l'image. Le descripteur HOG utilisé dans le cadre de la thèse correspond finalement à la concaténation des $N \times M$ histogrammes obtenus avec $N = M = 3$. La dimension de la sortie du descripteur HOG est donc de $9 \times 3 \times 3 = 81$ valeurs.

Finalement pour que ce descripteur soit robuste aux changements d'illuminations et de contraste, Dalal et Trigs [24] propose l'usage del'un des quatre types de normalisation. Soit v le vecteur non normalisé contenant tous les descripteurs HOG donné, $\|v\|_k$ sa k-norme pour $k = 1, 2, ...$ et e une petite constante (la valeur exacte, n'a pas d'importance). Les facteurs de normalisation sont alors définis par :

— La norme $L2$:

$$f = \frac{v}{\sqrt{\|v\|_2^2 + e^2}} \tag{4.14}$$

— La norme $L1$:

$$f = \frac{v}{(\|v\|_1 + e^2)} \tag{4.15}$$

— La racine $L1$:

$$f = \sqrt{\frac{v}{(\|v\|_1 + e^2)}} \tag{4.16}$$

Une quatrième norme $L2 - hys$, consistant à calculer v d'abord par la norme $L2$, puis à limiter le maximum de valeurs à 0.2 et enfin effectuer une renormalisation.

Depuis la proposition du descripteur HOG, plusieurs systèmes de reconnaissance y ont eu recours et ils ont montré de très bonnes performances, ce qui a ravivé l'intérêt pour les descripteurs denses non quantifiés.

4.3.1.1 Limitations du HOG

Comme nous l'avons mentionné précédemment, l'histogramme de gradient orienté est calculé dans de petites régions d'images appelées cellules, nous obtenons ces cellules en divisant l'image en fonction d'un certain nombre de pixels dans cette image, par exemple une cellule de $8 * 8$ pixels ou $5 * 5$ pixels, à notre avis, pour une grande image de test (par rapport aux images d'entraînement), le HOG est moins symétrique, en raison de l'impact du changement du nombre de pixels dans l'image.
En outre, cette méthode est sensible aux conditions d'éclairage changeantes, y compris le changement de la couleur d'éclairage, la diffusion de la lumière, les ombres, les changements de géométrie d'éclairage tels que l'ombrage, les reflets sous une source de lumière blanche et de la diffusion d'une source blanche, tout cela peut affecter considérablement le taux de reconnaissance de cette méthode.

4.3.1.2 Solutions proposées pour les limitations du HOG

La nouveauté de notre descripteur par rapport au descripteur HOG est d'utiliser la stabilité de la teinte contre le changement d'éclairage dans l'espace colorimétrique HSV et la constance de la couleur Gray-Edge pour rendre les couleurs des objets indépendantes de la couleur de la source d'éclairage afin de rendre le descripteur proposé robuste contre les changements des conditions d'éclairage. Nous avons également changé la méthode de choix de la taille des cellules pour garder les descripteurs symétriques afin de faciliter la tâche de classification et de rendre notre descripteur plus invariant aux transformations géométriques et photométriques.

4.3.2 Histogramme d'adversaire

L'histogramme de l'adversaire est une combinaison de trois histogrammes $1D$ basés sur les canaux de l'espace colorimétrique de l'adversaire (opponent color

space) :

$$O_1 = \frac{R-G}{\sqrt{2}} \tag{4.17}$$

$$O_2 = \frac{R+G-2B}{\sqrt{6}} \tag{4.18}$$

$$O_3 = \frac{R+G+B}{\sqrt{3}} \tag{4.19}$$

L'intensité lumineuse est représentée dans le canal $O3$ et les informations de couleur sont dans les canaux $O1$ et $O2$. En raison de la soustraction dans $O1$ et $O2$, les décalages seront annulés s'ils sont égaux pour tous les canaux (par exemple, une source de lumière blanche). Par conséquent, ces motifs de couleur sont invariants par rapport à l'intensité lumineuse. Le canal d'intensité $O3$ n'a pas de propriétés d'invariance. Les intervalles d'histogramme pour l'espace colorimétrique de l'adversaire ont des plages différentes de celles du modèle RGB.

D'après Van de Weijer [125], l'angle de l'adversaire ang_x^o dans l'espace colorimétrique des adversaires est supposé être invariant spéculaire. L'angle de l'adversaire ang_x^o est défini comme :

$$ang_x^O = arctag\left(\frac{O_{1x}}{O_{2x}}\right) \tag{4.20}$$

Où O_{1x} désigne le dérivé de premier ordre de O_1 et O_{2x} désigne le dérivé de premier ordre de O_2, etc. Les auteurs dans cite van2009evaluating ont appliqué une analyse d'erreur à l'angle de l'adversaire∂ang_x^O, cette erreur est défini comme le poids de l'angle de l'adversaire :

$$\partial ang_x^O = \frac{1}{\sqrt{O_{1x}^2 + O_{2x}^2}} \tag{4.21}$$

Les canaux $O1$ et $O2$ ont des bacs régulièrement espacés sur toute la plage $\left[-\frac{1}{2}, \frac{1}{2}\right]$. Tous les échantillons en dehors de cette plage sont placés dans les bacs extérieurs. Le canal $O3$ a ses bacs régulièrement espacés sur la plage $\left[0, \sqrt{3}\right]$. Les échantillons ne peuvent pas sortir de cette plage. Notre implémentation correspond au descripteur d'histogramme de teinte introduit par Van de Weijer [124]. L'histogramme de l'adversaire est quantifié à 36 bacs.

4.3.2.1 Limitations de l'histogramme d'adversaire

Selon Van de Sande et al [57], les histogrammes de l'adversaire ne sont pas invariants au changement d'intensité lumineuse, au décalage d'intensité lumineuse, au changement-décalage d'intensité lumineuse et au changement de couleur de la lumière. Ce descripteur n'est donc pas invariant à l'existence d'ombres, aux changements de géométrie d'éclairage comme l'ombrage, aux réflexions d'objets sous une source de lumière blanche, à la diffusion d'une source blanche, aux changements de couleur d'éclairage et aux changment de la diffusion de la lumière. De plus, ce descripteur n'est pas invariant à la transformation géométrique et photométrique.

4.3.2.2 Solutions proposées pour les limitations de l'histogramme d'adversaire

Le descripteur de teinte proposé résout ces inconvénients en utilisant la teinte qui est invariante au changement d'intensité lumineuse (changement des ombres et changements de géométrie d'éclairage comme l'ombrage), au décalaged'intensité lumineuse (les objets se détachent sous une source de lumière blanche et le diffusion d'une source blanche) et au changement-décalage de l'intensité lumineuse (combinaisons des deux conditions ci-dessus). De plus, et par l'utilisation de la constance de couleur Gray-Edge nous résolvons le problème du manque d'invariance au changement de couleur de la lumière (changement de la couleur de l'éclairage et de la diffusion de la lumière). Enfin, les idées de cellules et de bins utilisées dans le descripteur de teinte proposé résolvent l'absence d'invariance à la transformation géométrique et photométrique.

4.3.3 Histogramme de teinte

Dans l'espace colorimétrique HSV, on sait que la teinte devient instable autour de l'axe gris. À cette fin, Van de Weijer et al. [122] applique une analyse d'erreur à la teinte. L'analyse montre que la certitude de la teinte est inversement proportionnelle

à la saturation. Par conséquent, l'histogramme de teinte est rendu plus robuste en pesant chaque échantillon de la teinte par sa saturation. Les modèles de couleur H et S sont invariables en échelle et invariables en ce qui concerne l'intensité lumineuse.

La teinte et la saturation de l'espace colorimétrique HSV peuvent être calculées à partir des couleurs de l'histogramme de l'adversaire de Van de Weijer [121] :

$$teinte = \arctan\left(\frac{O_1}{O_2}\right) = \arctan\left(\frac{\sqrt{3}\,(R-G)}{R+G-2B}\right) \tag{4.22}$$

$$saturation = \sqrt{O_1^2 + O_2^2} = \sqrt{\frac{2}{3}\,(R^2+G^2+B^2-RG-RB+GB)} \tag{4.23}$$

Où O_1 et O_2 sont les deux composantes de l'espace colorimétrique des adversaires respectivement cités dans l'équation 4.17 et l'équation 4.18.

L'histogramme de teinte est un histogramme $1D$ basé sur le canal de teinte de l'espace colorimétrique HSV, et comme nous l'avons déjà mentionné, pour remédier à l'instabilité de la teinte dans les zones sombres et /ou grises, les échantillons de teinte sont pesés par saturation. Le canal de teinte est divisé en intervalles de 36 régulièrement espacés. Notre implémentation correspond au descripteur d'histogramme de teinte introduit par Van de Weijer [121].

Ce descripteur est spécifiquement conçu pour les régions d'intérêt. Lorsqu'il est appliqué à des régions d'intérêt, chaque échantillon de teinte est également pesé par sa distance du centre de la région d'intérêt. Les échantillons près du centre reçoivent un poids plus lourd que ceux près de la limite de la région. La pesée est réalisée à l'aide d'un masque gaussien. Puisque le gaussien est symétrique en rotation, il garantit l'invariance rotationnelle pour ce descripteur.

4.3.3.1 Limitations de l'histogramme de teinte

D'après Van de Sande et al [121], l'histogramme de teinte de Gevers n'est pas invariant aux changements de couleur de la lumière, donc le taux de reconnaissance

sera affecté par un changement de couleur d'éclairage et de diffusion de la lumière. lumière. De plus, la méthode de construction de ce descripteur n'est pas efficace à grande échelle, de sorte que le système de reconnaissance ne sera pas invariant à la transformation géométrique et photométrique.

4.3.3.2 Solutions proposées pour les limitations de l'histogramme de teinte

La nouveauté du descripteur de teinte proposé par rapport à ces histogrammes de teinte de Gevers est l'utilisation de la constance de couleur Gray-Edge pour rendre le descripteur invariant aux changements de couleur de la lumière, ainsi que la modification de la méthodologie de construction du descripteur. En utilisant les idées de cellules et de bins dans la construction du descripteur proposé, ce dernier devient invariant à la transformation géométrique et photométrique.

4.3.4 Descripteur local d'image à partir des réponses du filtre Gabor

Bien que les filtres Gabor soient bien connus et souvent utilisés, mais dans cite inproceedings, les auteurs utilisent ces filtres pour la première fois pour présenter un descripteur basé sur des filtres Gabor multi-échelle et multi-diffusion. Selon les auteurs, ce descripteur robuste aux changements d'éclairage est un histogramme commun 3D $H\left(\theta_i, w_j, l\right)$ des valeurs de $\tilde{F}$:

$$H\left(\theta_i, w_j, l\right) = \sum_{p \in \tilde{F}} C_l\left(p\right).\tilde{F}\left(p, \theta_i, w_j\right) \tag{4.24}$$

$\tilde{F}$: La carte de descripteur normalisée, L Cellules, $C_l\left(p\right), l = 1...L$ est définie comme représentante de la pondération de l'emplacement spatial pour le sous-histogramme local de la cellule. Pour plus de détails sur ce descripteur d'image local, même à partir des réponses du filtre Gabor, le lecteur pourra se référer à [138].

4.3.4.1 Limitations du descripteur local du filtre Gabor

Dans [138], les auteurs présentent le descripteur d'image local des réponses du filtre Even Gabor comme un descripteur robuste aux changements de conditions d'éclairage uniquement. Ce descripteur n'est pas invariant à la transformation géométrique et photométrique, de sorte que le taux de reconnaissance de cette méthode peut être réduit en cas d'existence de la translation, la rotation et le changement de direction, etc.

4.3.4.2 Solutions proposées pour les limitations du descripteur local du filtre Gabor

Le descripteur de teinte proposé corrige les inconvénients du descripteur d'image local à partir des réponses du filtre de Gabor en utilisant des idées de cellules et de bins.

4.4 Méthode proposée : descripteur de teinte

Selon Van de Sande et al. [102], les taches de l'image sont représentées par un histogramme sur la teinte calculée à partir des valeurs RGB correspondantes de chaque pixel par :

$$\text{teinte} = \arctan\left(\frac{\sqrt{3}\,(R-G)}{R+G-2B}\right) \tag{4.25}$$

Pour contrer les instabilités de teinte, son impact sur l'histogramme est pondéré par la saturation du pixel correspondant. Le descripteur de teinte est invariant par rapport à la géométrie d'éclairage lors de l'hypothèse d'un éclairage blanc.

Pour construire notre descripteur de teinte [42], nous commençons par mesurer les couleurs des objets indépendamment de la couleur de la source de lumière, pour rendre cette indépendance nous utilisons la constance de couleur Gray-Edge. L'étape suivante consiste à diviser l'image en cellules de 25 (sous-image) superposées de 50 %. Pour chaque cellule, nous calculons la valeur de teinte de chaque

pixel et nous utilisons l'idée de bins pour former le descripteur de cellule. De cette façon chaque cellule sera codée avec un vecteur de 12 valeurs chaque valeur correspond à la grandeure d'un bin (le choix des paramètres 25 cellules et 12 bins sera justifié dans la partie expérimentale). Ce travail sera répété pour les cellules de 25. La dernière étape consiste à regrouper les vecteurs cellulaires en un seul descripteur de caractérisation de l'image d'objet appelé descripteur de teinte (figure 4.1).

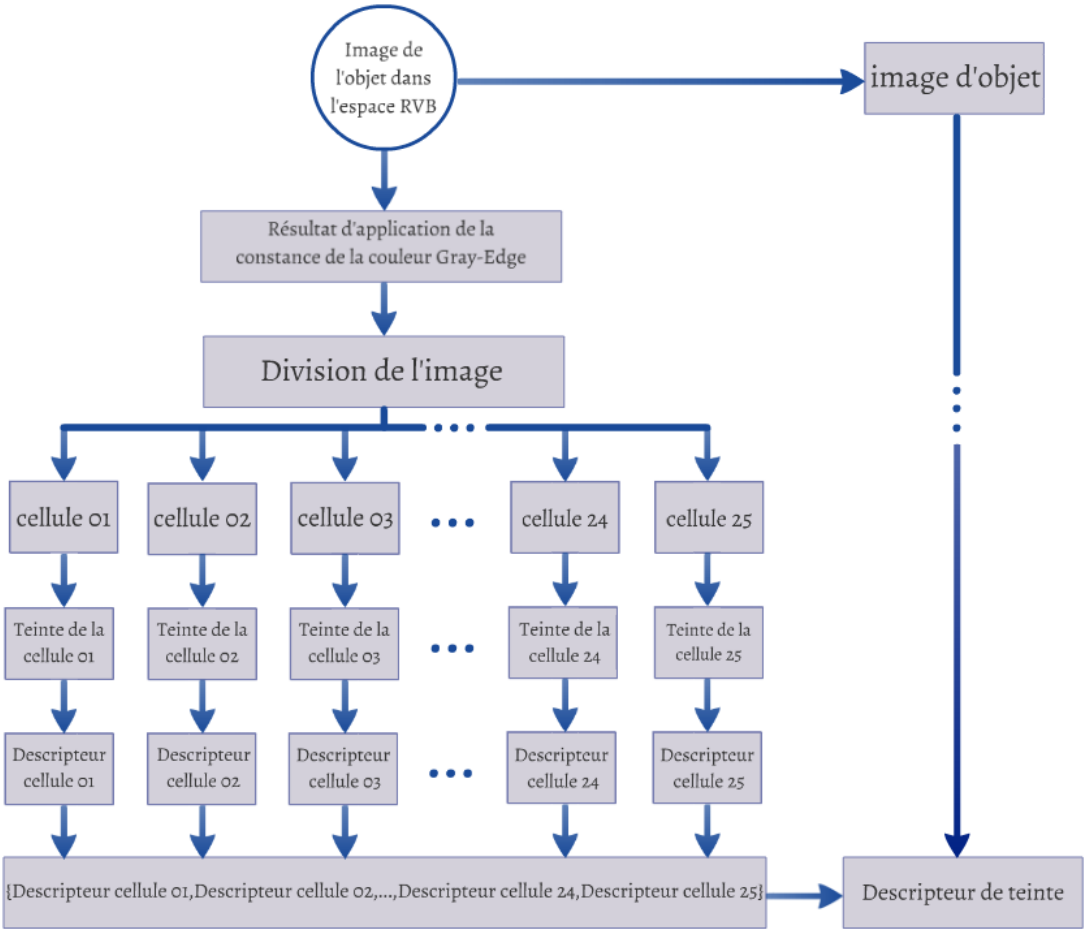


FIGURE 4.1 – Étapes de construction du descripteur de teinte (source : [42]).

4.4.1 Étape 01 : Constance de la couleur

La constance des couleurs est la capacité de reconnaître les couleurs des objets indépendamment de la couleur de la source de lumière [31]. La constance des couleurs est importante pour de nombreuses applications de vision par ordinateur, telles que la récupération d'images, la classification d'images, la reconnaissance

d'objets couleur et le suivi d'objets.

Dans notre descripteur, la première étape consiste à appliquer cette constance de couleur Gray-Edge [4] pour rendre les couleurs de l'image de l'objet invariantes aux changements de couleur de la lumière. Un exemple de l'application de cette constance de couleur Gray-Edge est illustré à la figure 4.2.

(a) Image originale 1_l5c3.png

(b) Résultat d'application de la constance de la couleur Gray-Edge sur l'image 1_l5c3.png

FIGURE 4.2 – Effets d'application de la constance de la couleur Gray-Edge

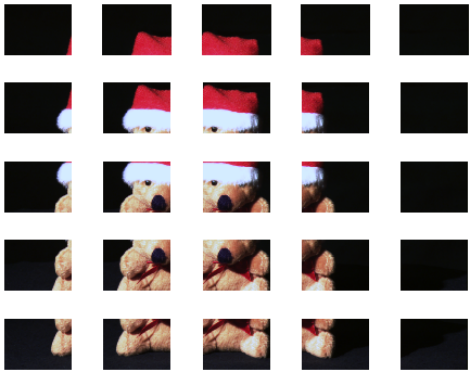
4.4.2 Étape 02 : Division de l'image

L'un des principaux avantages du descripteur HOG est son invariance aux transformations géométriques et photométriques puisqu'il opère sur des cellules locales cite ilas2018new. Afin de profiter de cette invariance face aux transformations géo-

4. Voir l'annexe C

métriques et photométriques dans notre descripteur proposé, nous avons utilisé cette idée de cellules pour faire fonctionner le descripteur de teinte proposé également sur les cellules locales. Pour cette raison, et après plusieurs études, nous avons choisi de diviser l'image de l'objet en cellules superposées de 25 par 50 % figure 4.3.

(a) Résultat d'application de la constance de la couleur Gray-Edge sur l'image 1_l5c3.png.



(b) Division de l'image en 25 cellules

FIGURE 4.3 – Résultat de la division de l'image (cellules de descripteur).

4.4.3 Étape 03 : Création de descripteurs des cellules

Dans cette partie, nous commençons par diviser l'espace de teinte en parties de 12 appelées bins (bins).

$$\text{Teinte} = \left\{ \begin{array}{l} bin1 \cup bin2 \cup bin3 \cup bin4 \cup bin5 \cup bin6 \cup \\ bin7 \cup bin8 \cup bin9 \cup bin10 \cup bin11 \cup bin12 \end{array} \right\} \tag{4.26}$$

Avec :

$$\begin{aligned} &bin01 = [000^o, 029^o], bin02 = [030^o, 059^o], bin03 = [060^o, 089^o], \\ &bin04 = [090^o, 119^o], bin05 = [120^o, 149^o], bin06 = [150^o, 179^o], \\ &bin07 = [180^o, 209^o], bin08 = [210^o, 239^o], bin09 = [240^o, 269^o], \\ &bin10 = [270^o, 299^o], bin11 = [300^o, 329^o], bin12 = [330^o, 359^o] \end{aligned} \tag{4.27}$$

Après avoir créé les 12bins, nous calculons la valeur de teinte de chaque pixel de cette cellule en fonction de l'équation 4.25. Ensuite, le vote dans les bins se fera en fonction de la valeur de teinte de chaque pixel, la grandeur de chaque bin est calculée en additionnant les magnitudes de teinte de tous les pixels correspondants. De cette façon, nous construisons un vecteur descripteur (cellule D) de valeurs de 12 (chaque valeur de ces valeurs de 12 correspond à la magnitude d'un bin).

$$\text{D cellule } 01 = \left\{ \begin{array}{ll} \text{Magnitude bin}01, & \text{Magnitude bin}02, \\ \text{Magnitude bin}03, & \text{Magnitude bin}04, \\ \text{Magnitude bin}05, & \text{Magnitude bin}06, \\ \text{Magnitude bin}07, & \text{Magnitude bin}08, \\ \text{Magnitude bin}09, & \text{Magnitude bin}10, \\ \text{Magnitude bin}11, & \text{Magnitude bin}12 \end{array} \right\} \tag{4.28}$$

4.4.4 Étape 04 : Normalisation des histogrammes

Dans notre thèse, nous utilisons la normalisation $L2$. Pour obtenir le descripteur final de cellule 1, nous normalisons le descripteur de cellule trouver dans l'équation 4.28 avec la normalisation $L2$ (équation 4.14).

4.4.5 Étape 05 : Descripteur de teinte final

Pour obtenir le descripteur final, les étapes 3 et 4 sont répétées pour toutes les cellules (25 cellules). Et puis, tous les descripteurs cellulaires sont regroupés en un

seul vecteur (figure 4.1) appelé «descripteur de teinte».

$$\text{Descripteur de teinte} = \left\{ \begin{array}{l} \text{D cellule}01, \text{D cellule}02, \text{D cellule}03, \text{D cellule}04, \text{D cellule}05, \\ \text{D cellule}06, \text{D cellule}07, \text{D cellule}08, \text{D cellule}09, \text{D cellule}10, \\ \text{D cellule}11, \text{D cellule}12, \text{D cellule}13, \text{D cellule}14, \text{D cellule}15, \\ \text{D cellule}16, \text{D cellule}17, \text{D cellule}18, \text{D cellule}19, \text{D cellule}20, \\ \text{D cellule}21, \text{D cellule}22, \text{D cellule}23, \text{D cellule}24, \text{D cellule}25 \end{array} \right\} \tag{4.29}$$

La taille de la sortie de ce descripteur de teinte correspond à la multiplication du nombre de cellules (25 cellules) par le nombre de bins choisis (12 bins) soit donc $25 * 12 = 300$ valeurs.

4.5 Base de données d'images

Cette partie est consacrée à présenter brièvement la bibliothèque d'images d'objets Columbia (COIL-100) et la bibliothèque d'images d'objets d'Amsterdam (ALOI).

4.5.1 Bibliothèque d'images d'objets Columbia (COIL-100)

La bibliothèque d'images d'objets Columbia (COIL-100) [84] est une base de données d'images en couleur contenant 7200 images (128×128 pixels) de 100 objets, 72 images différentes par objet. Les images des objets ont été prises à des intervalles d'exposition de 5 degrés. La grande variation des angles rend cette base de données idéale pour tester la robustesse face aux transformations géométriques et photométriques. C'est pourquoi cette base de données est largement utilisée dans les expériences de reconnaissance d'objets.

La figure 4.4 montre certains objets de la base de données coil-100, tandis que la figure 4.5 montre le même objet avec des vues différentes.

FIGURE 4.4 – Échantillons d'images de la base de données COIL-100.

FIGURE 4.5 – Exemple d'objet de COIL-100 avec différentes orientations.

4.5.2 Bibliothèque d'images d'objets d'amsterdam (The Amsterdam Library of object images (ALOI))

Geusebroek et al. [34], présentent la collection ALOI (Amsterdam Library of Object Images) de 1000 objets (figure 4.6) enregistrés dans diverses circonstances d'imagerie. Afin de capturer la variation sensorielle des enregistrements d'objets,

ils ont systématiquement varié les angles de vision (les images des objets ont été prises à des intervalles de pose de 5 degrés (figure 4.7), l'angle d'éclairage (figure 4.8) et la couleur d'éclairage pour chaque objet (figure 4.9), et ont également capturé des images stéréo à large ligne de base. Ils ont enregistré plus d'une centaine d'images de chaque objet, ce qui a donné un total de 110250 images pour la collection. Nous utilisons dans cette thèse les images de la version (384×288 pixels).

FIGURE 4.6 – Échantillons d'images de la bibliothèque ALOI.

Ces images sont mises à la disposition du public à des fins de recherche scientifique. La couleur de la lumière est modifiée en modifiant la température de couleur de l'éclairage, ce qui entraîne des objets éclairés d'une lumière rougeâtre à blanche. Pour être complet, le jeu de données ALOI comprend également des objets éclairés par un nombre différent de lumières blanches à des angles de plus en plus obliques

(entre une et trois lumières blanches autour de l'objet, introduisant un ombrage automatique pour jusqu'à la moitié de l'objet).

En bref, nous pouvons classer les images de cette base de données en 3 sous-ensembles, ALOI-angle de vue, ALOI-angle d'éclairage et ALOI-couleur d'éclairage :

— **ALOI-angle de vue :** Dans cette partie, les images ont été photographiées en rotation. Chaque photo est prise après que l'objet a été tourné de 5 degrés, ce qui donne 72 photos pour chaque objet (figure 4.7).

FIGURE 4.7 – Exemple d'objet d'ALOI avec différentes orientations.

— **ALOI-angle d'éclairage :** Les images ont été prises dans 24 directions d'éclairage différentes, ce qui donne 24 photos pour chaque objet (figure 4.8).

FIGURE 4.8 – Exemple d'objet d'ALOI vu sous 24 différentes directions d'éclairage.

— **ALOI-couleur d'éclairage :** Les images ont été prises sous 12 températures de couleur d'éclairage différente, ce qui donne 12 photos pour chaque objet (figure 4.9).

FIGURE 4.9 – Exemple d'objet d'ALOI-COL vu sous 12 températures de couleur d'éclairage différentes.

Tout comme COIL-100, le jeu de données ALOI-angle de vue est largement utilisé dans les expériences qui étudient l'invariance aux changements des conditions géométriques. Cependant, la grande variation des conditions d'éclairage dans ALOI-angle d'éclairage et ALOI-couleur d'éclairage rend cette base de données optimale pour tester l'invariance face aux changements des conditions d'éclairage. Dans nos expériences, nous collecterons ALOI-angle d'éclairage et ALOI-couleur d'éclairage dans un seul groupe appelé : (**ALOI-éclairage**), ce qui fait 36 images pour chaque objet. 11 images seront utilisées pour l'ensemble d'entraînement et 25 images pour l'ensemble tests. Pour ALOI-angle de vue, et comme pour COIL-100, 22 images seront utilisées pour l'ensemble d'entraînement et 50 images pour l'ensemble tests.

4.6 Critères d'évaluation

La mesure F est l'indicateur récapitulatif couramment utilisé depuis 25 ans pour évaluer les algorithmes de classification des données, basés sur la précision et le rappel. Dans cet thèse, nous utilisons des critères similaires à ceux proposés dans [37, 54].

4.6.1 Taux de reconnaissance

La précision (P), est la proportion d'objets reconnus parmi tous les d'objets proposés.

$$\mathrm{P} = \frac{\text{Image reconnues}}{\text{Image reconnues + Image non reconnue}} \tag{4.30}$$

Dans les statistiques, la précision est appelée valeur prédictive positive. Le taux de reconnaissance peut être calculé en utilisant le critère de précision cité dans l'équation 4.30.

$$\text{Taux de reconnaissance}\,(\%) = \mathrm{P} * 100 \tag{4.31}$$

4.6.2 Temps de calcule

En technologie, le temps de réponse est une mesure des performances d'une application interactive. Il peut être défini comme le décalage entre une entrée électronique et le signal de sortie. Dans notre système, nous définissons le temps de calcule comme le temps écoulé entre le début de la tâche et la fin de catégorisation.

4.6.3 Taille du descripteur

Dans notre système, nous définissons la taille du descripteur comme la taille de la mémoire occupée par le descripteur. Tous ces 3 critères d'évaluation seront utilisés pour évaluer les performances de notre descripteur proposé.

4.7 Évaluation du descripteur de teinte

Afin d'évaluer notre descripteur, nous avons testé notre approche de catégorisation des images d'objets couleur à l'aide de deux ensembles de données accessibles au public : la bibliothèque d'images d'objets Columbia (COIL-100) [84] et la bibliothèque d'images d'objets d'Amsterdam (ALOI) [34]. Pour une comparaison détaillée entre le descripteur proposé et certaines autres méthodes (classiques) existantes de catégorisation des images d'objets, nous avons utilisé : l'histogramme à gradient orienté (HOG), l'histogramme de l'adversaire, l'histogramme de teinte de Gevers et le descripteur des réponses du filtre Gabor. Ce dernier descripteur n'est utilisé que dans les tests de la bibliothèque d'images d'objets d'Amsterdam, car le code source fourni par les auteurs ne prend pas en charge la taille des images

de la bibliothèque d'images d'objets Columbia (128×128 pixels). L'opération de test contient deux parties principales : la première est le test paramétrique, nous nous concentrons sur les différents paramètres du descripteur proposé, tandis que la seconde partie est une étude comparative entre les résultats du descripteur proposé et les résultats des méthodes classiques de catégorisation des images d'objets.

4.7.1 Étude paramétrique

Dans cette partie, nous avons pour objectif de faire des tests sur les différents paramètres du descripteur proposé, notamment : le nombre de cellules, le nombre bins (bacs), la taille de la base d'apprentissage et la taille de la base des tests.

4.7.1.1 Configuration des tests

Chaque tâche dans cette section contient 3 parties : test sur la base de données COIL-100, tests sur la base de données ALOI-angle de vue et tests sur la base de données ALOI-éclairage. Pour les 3 parties, nous avons réalisé un test sur 50 objets, comme mentionné précédemment, pour les première et deuxième parties des tests (tests COIL-100 et ALOI-angle de vues), chaque objet a été représenté avec 72 images, 22 images ont été utilisées dans l'ensemble d'apprentissage et 50 images dans l'ensemble de tests, tandis que dans les tests de ALOI-éclairage, 11 images ont été utilisées pour l'apprentissage et 25 images pour les tests.
Les tests ont été réalisés à l'aide de Matlab $2018b$ 64 bits dans un ordinateur portable Toshiba $P50$ avec processeur $i7-4700MQ$ $2,40$ GZH, 8 Go de RAM et double carte graphique, Intel 4400 HD et NVIDIA GeForce $GT745M$.

4.7.1.2 Influence de nombre des cellules sur les performances du descripteur de teinte

Le but de ces tests est de justifier l'utilisation de l'idée de cellules dans notre descripteur. Pour cela, nous montrons dans cette partie l'effet de la variation du

nombre de cellules sur les performances de notre descripteur. La figure 4.10 montre les différentes méthodes de division d'image testées dans cette section, une seule cellule c'est-à-dire l'image entière (figure 4.10a), neuf cellules (figure 4.10b) et enfin vingt-cinq cellules (figure 4.10c).

(a) Image en une seule cellule.

(b) Image en neuf cellules.

(c) Image en vingt-cinq cellules.

FIGURE 4.10 – Les différentes méthodes de division d'image.

Pour tous les tests de cette partie, nous utilisons des classificateurs SVM et des classificateurs KNN. Nous commencerons par étudier le taux de reconnaissance sur les 3 bases de données, puis nous présenterons le temps de calcules et la taille de

mémoire occupée par notre descripteur de teinte à l'aide des différentes méthodes de division d'image. Notre première priorité est de trouver la méthode de division d'image qui donne le taux de reconnaissance le plus élevé. Les figures 4.11, 4.12 et 4.13 montrent les résultats de la reconnaissance du descripteur de teinte sur les ensembles de données COIL-100, ALOI-angle de vue et ALOI-éclairage respectivement.

Le tableau 4.1 récapitule les différents temps totaux (T.T[S]) de catégorisation et le temps moyen (T.M[S]) de catégorisation de chaque méthode, ce dernier et calculé par devise le temps total de l'expérience sur le nombre total d'images catégorisées (2500 images pour COIL-100 et ALOI-angle de vue, 1250 images pour ALOI-éclairage). Le changement de méthode de division de l'image (changement du nombre de cellules) implique un changement de la taille du descripteur, ces changements sont indiqués dans le tableau 4.2.

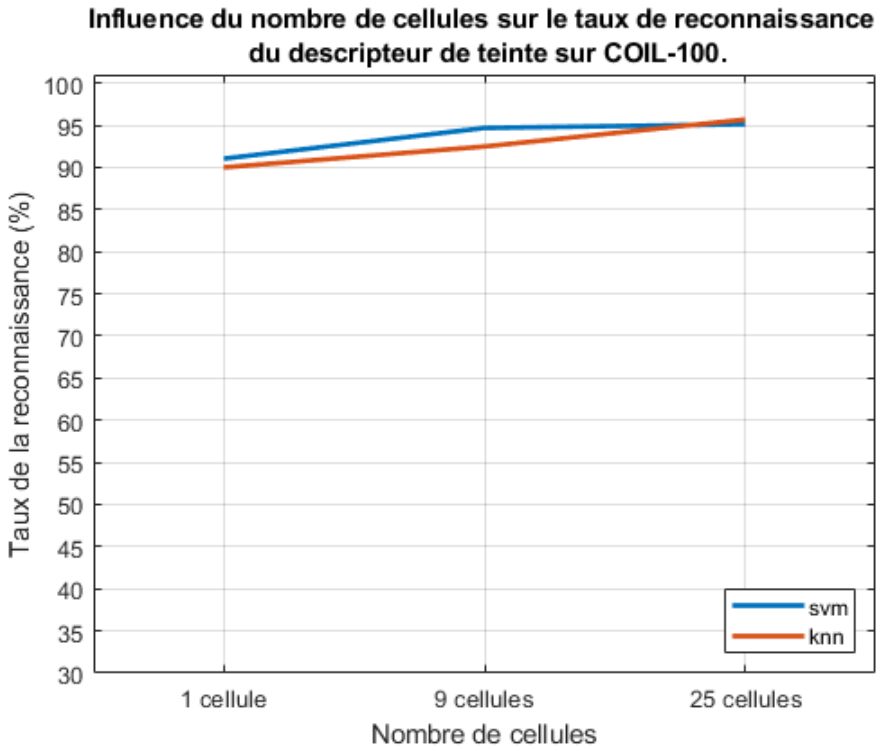


FIGURE 4.11 – Influence du nombre de cellules sur le taux de reconnaissance du descripteur de teinte sur COIL-100.

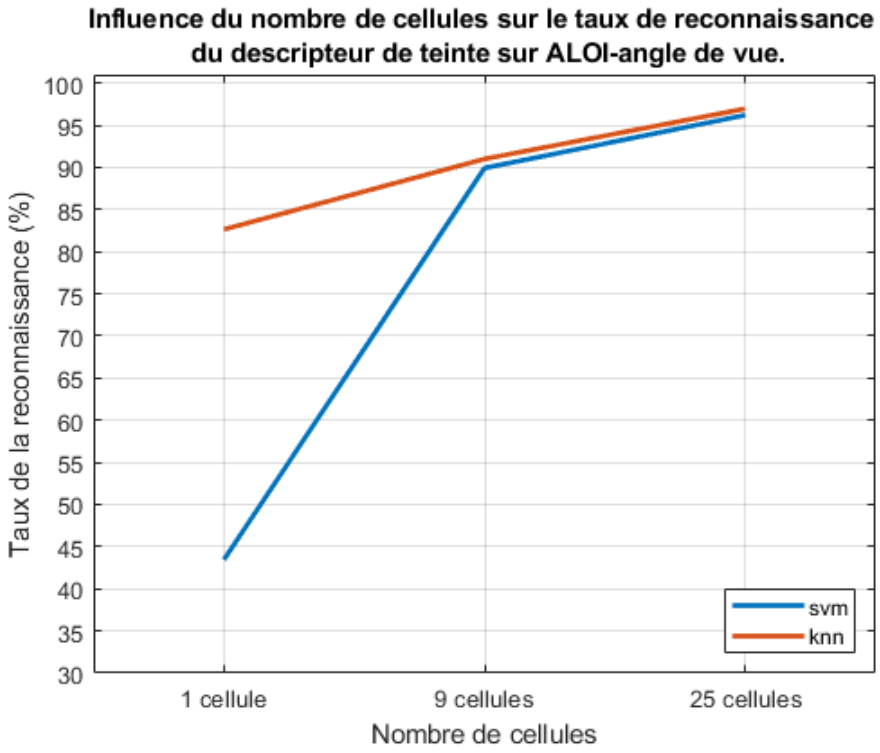


FIGURE 4.12 – Influence du nombre de cellules sur le taux de reconnaissance du descripteur de teinte sur ALOI-angle de vue.

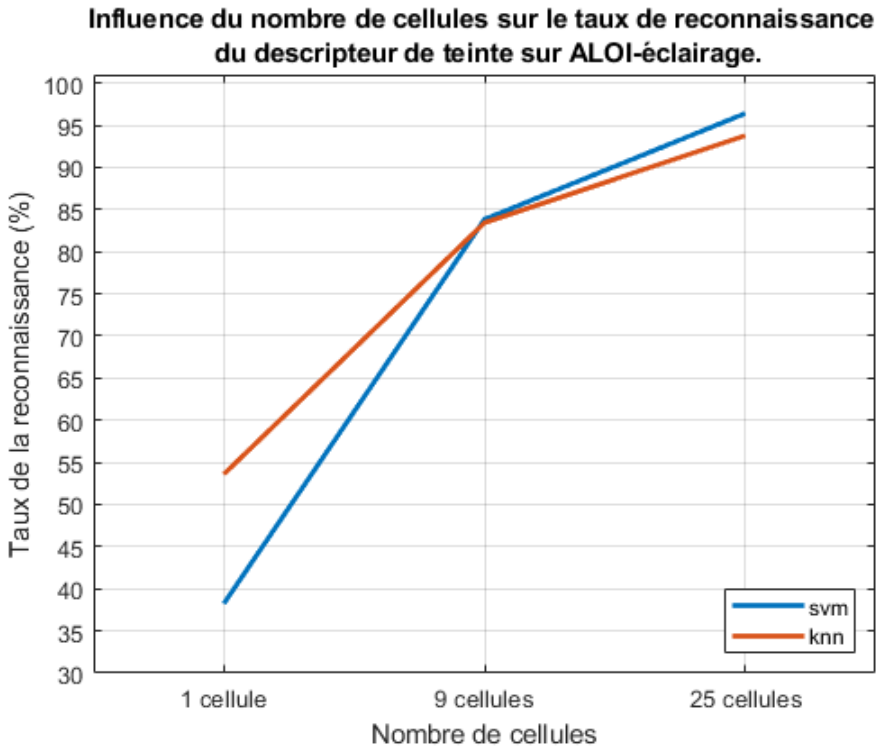


FIGURE 4.13 – Influence du nombre de cellules sur le taux de reconnaissance du descripteur de teinte sur ALOI-éclairage.

TABLE 4.1 – Influence du nombre de cellules sur le temps de réponse du descripteur de teinte.

		SVM		KNN	
		T.T[S]	T.M[S]	T.T[S]	T.M[S]
COIL-100	1 cellule	2048	0.82	7	0.01
	9 cellules	2544	1.02	13	0.01
	25 cellules	2576	1.03	92	0.04
ALOI-angle de vue	1 cellul	121278	48.51	105	0.04
	9 cellules	732	0.29	167	0.07
	25 cellules	2103	0.84	399	0.16
ALOI-éclairage	1 cellul	77626	62.1	34	0.03
	9 cellules	599	0.48	35	0.035
	25 cellules	907	0.73	238	0.19

TABLE 4.2 – Influence du nombre de cellules sur la taille du descripteur de teinte.

Nombre de cellules	Taille du descripteur
1 cellule	12 valeurs
9 cellules	108 valeurs
25 cellules	300 valeurs

Les figures 4.11, 4.12 et 4.13 confirment l'efficacité de la méthode cellulaire proposée dans notre descripteur de teinte. Dans les expériences des 3 ensembles de données, les performances des descripteurs à 9 et 25 cellules sont nettement meilleures que celles du descripteur qui utilise l'image entière à la fois (une seule cellule) dans la catégorisation des images d'objets. Ces figures montrent également qu'il existe une relation de corrélation positive entre l'augmentation du nombre de cellules et l'amélioration du taux de reconnaissance.

Les résultats de ces trois figures 4.11, 4.12 et 4.13 montrent également l'importance de faire fonctionner le descripteur localement dans zones de taille limitée (un nombre limité de pixels). En comparant les résultats du descripteur qui utilise une seule cellule sur les trois bases de données, on constate que les résultats de ce descripteur sur la base de données COIL-100 sont remarquablement meilleurs en raison de la taille des images de cette base de données (128×128 pixels) par rapport à la taille des images de la base ALOI (384×288 pixels).

D'après les résultats affichés dans les tableaux 4.1 et 4.2, on peut conclure que

malgré la corrélation positive entre le nombre de cellules et l'amélioration du taux de reconnaissance, l'augmentation de ce nombre de cellules ralentit la tâche de catégorisation et agrandit la taille du descripteur, chose qui n'est pas souhaitable.

À partir de ces deux tableaux (4.1 et 4.2) aussi, nous notons également le temps de réponse remarquablement important dépensé par les classificateurs SVM dans la tâche de catégorisation des images de la base de données ALOI en utilisant le descripteur de teinte avec une seule cellule (l'image entière à un temps). Cette difficulté de classification est due à la petite taille du descripteur (12 valeurs) qui rend l'opération de recherche d'une marge de séparation très compliquée.

4.7.1.3 Influence de nombre des bins sur les performances du descripteur de teinte

Le but de cette partie est de justifier l'utilisation de la méthode des bins dans notre descripteur de teinte. Les tests ont été réalisés en faisant varier le nombre de bins dans le descripteur, nous avons utilisé 6 bins, 9 bins, 12 bins et 15 bins. La figure 4.14 montre le cercle de teinte dans l'espace colorimétrique HSV.

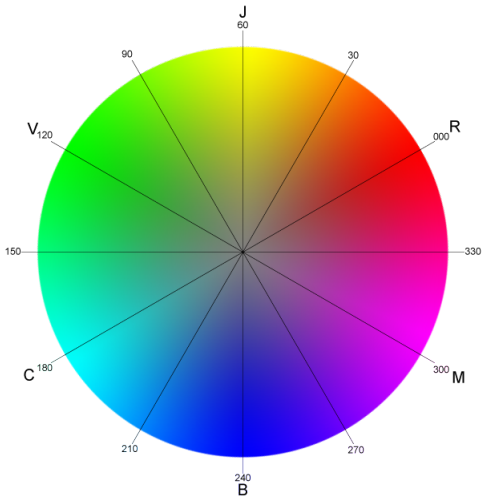


FIGURE 4.14 – Cercle de teinte dans l'espace de couleur HSV (source : [42]).

Pour avoir le nombre de bins souhaité, ce cercle (figure 4.14) doit être devisé comme suit :

1. **Descripteur de teinte avec 6 bins (DT_6)** :

$$\text{Teinte} = [\text{bin01} \cup \text{bin02} \cup \text{bin03} \cup \text{bin04} \cup \text{bin05} \cup \text{bin06}] \tag{4.32}$$

Avec :

$$\text{bin01} = [000^o, 059^o],\ \text{bin02} = [060^o, 119^o],\ \text{bin03} = [120^o 179^o]$$
$$\text{bin04} = [180^o, 239^o],\ \text{bin05} = [240^o, 299^o],\ \text{bin06} = [300^o 359^o]$$

2. **Descripteur de teinte avec 9 bins (DT_9)** :

$$\text{Teinte} = [\text{bin01} \cup \text{bin02} \cup \text{bin03} \cup \text{bin04} \cup \text{bin05} \cup \text{bin06} \cup \text{bin07} \cup \text{bin08} \cup \text{bin09}] \tag{4.33}$$

Avec :

$$\text{bin01} = [000^o, 039^o],\ \text{bin02} = [040^o, 079^o],\ \text{bin03} = [080^o 119^o]$$
$$\text{bin04} = [120^o, 159^o],\ \text{bin05} = [160^o, 199^o],\ \text{bin06} = [200^o 239^o]$$
$$\text{bin04} = [240^o, 279^o],\ \text{bin05} = [280^o, 319^o],\ \text{bin06} = [320^o 359^o]$$

3. **Descripteur de teinte avec 12 bins (DT_{12})** :

Ce descripteur est a été présenter déjà dans les équations 4.26 et 4.27.

4. **Descripteur de teinte avec 15 bins (DT_{15})** :

$$\text{Teinte} = \begin{bmatrix} \text{bin01} \cup \text{bin02} \cup \text{bin03} \cup \text{bin04} \cup \text{bin05} \cup \text{bin06} \cup \text{bin07} \cup \text{bin08} \cup \\ \text{bin09} \cup \text{bin10} \cup \text{bin11} \cup \text{bin12} \cup \text{bin13} \cup \text{bin14} \cup \text{bin15} \end{bmatrix} \tag{4.34}$$

Avec :

$$\text{bin01} = [000^o, 023^o],\ \text{bin02} = [023^o, 047^o],\ \text{bin03} = [048^o, 071^o]$$
$$\text{bin04} = [072^o, 095^o],\ \text{bin05} = [096^o, 119^o],\ \text{bin06} = [120^o, 143^o]$$
$$\text{bin07} = [144^o, 167^o],\ \text{bin08} = [168^o, 191^o],\ \text{bin09} = [192^o, 215^o]$$
$$\text{bin10} = [216^o, 239^o],\ \text{bin11} = [240^o, 263^o],\ \text{bin12} = [264^o, 287^o]$$
$$\text{bin13} = [288^o, 311^o],\ \text{bin14} = [312^o, 335^o],\ \text{bin15} = [336^o, 359^o]$$

Les figures 4.15, 4.16 et 4.17 représentent les résultats de ces descripteurs DT_{06}, DT_{09}, DT_{12} et DT_{15} dans la catégorisation d'objet sur bases de données COIL-100, ALOI-angle de vue et ALOI-éclairage respectivement.

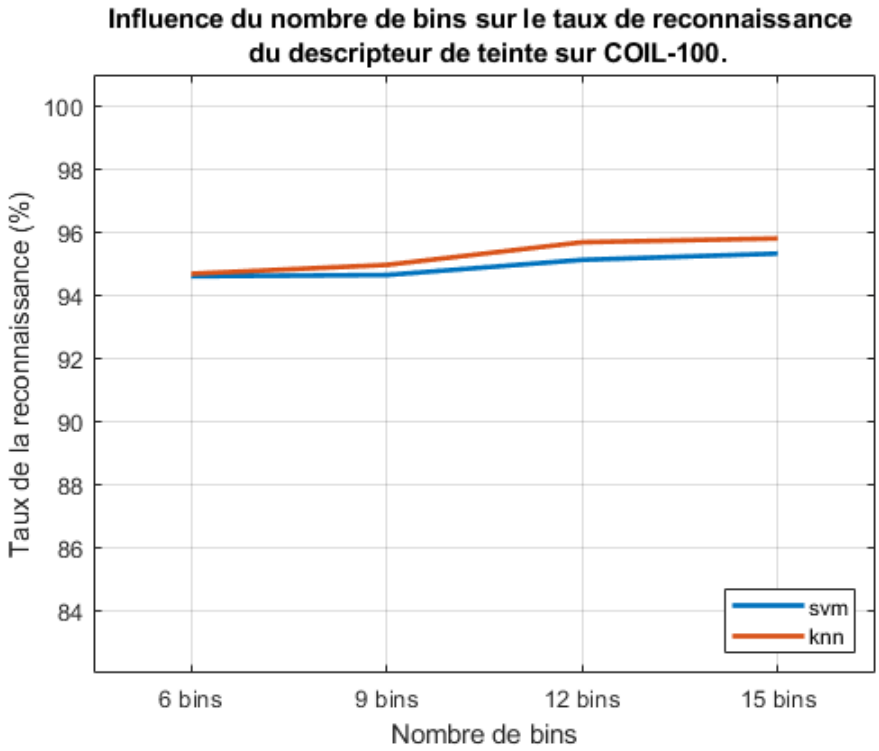


FIGURE 4.15 – Influence du nombre de bins sur le taux de reconnaissance du descripteur de teinte sur COIL-100.

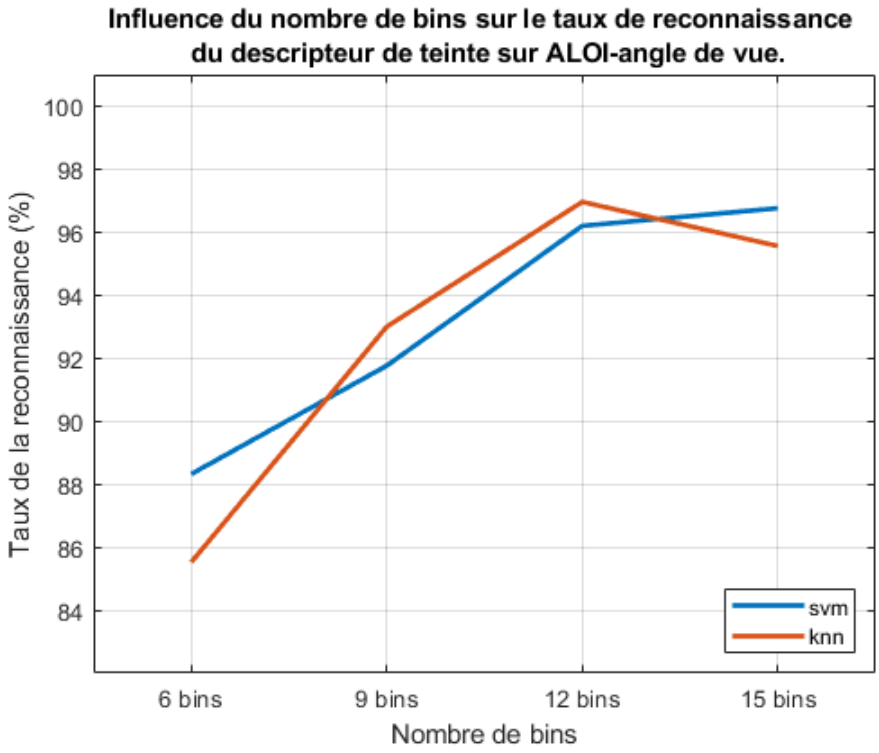


FIGURE 4.16 – Influence du nombre de bins sur le taux de reconnaissance du descripteur de teinte sur ALOI-angle de vue.

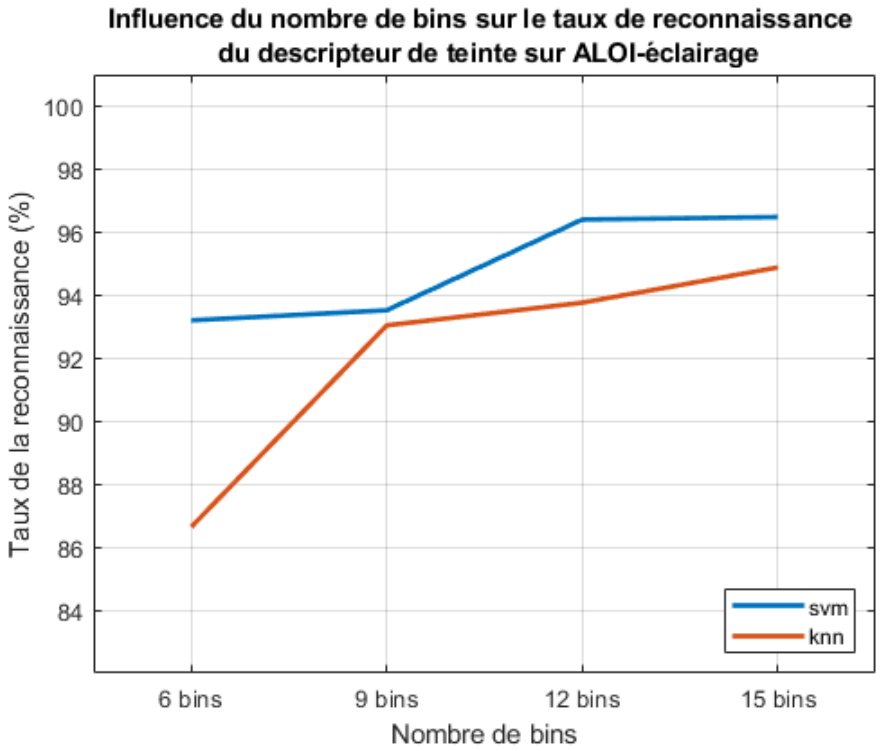


FIGURE 4.17 – Influence du nombre de bins sur le taux de reconnaissance du descripteur de teinte sur ALOI-éclairage.

La figure 4.15, la figure 4.16 et la figure 4.17 montrent qu'il existe une relation de corrélation positive entre l'augmentation du nombre de bins et l'amélioration du taux de reconnaissance. Cette méthode de bins a permis d'ajouter une certaine invariance contre les changements géométriques et photométriques au descripteur de teinte, ce qui a amélioré le taux de reconnaissance d'une manière remarquable. Cependant, et similaire au changement du nombre de cellules, l'influence de cette méthode de bins n'a pas été grande sur le taux de reconnaissance de la base de données COIL-100 en raison de la petite taille des images dans cette base de données. Une petite taille d'image signifie qu'il n'y a pas une grande quantité d'informations de couleur à extraire de l'image par rapport aux images de la base de données ALOI par exemple, dans ce cas l'influence de la méthode des bins, tout comme l'influence de la cellule méthode, ne peut pas être significative.

Le tableau 4.3 montre l'influence de la modification du nombre de bins sur la durée totale de la tâche de catégorisation ainsi que le temps moyen de cette dernière, tandis que le tableau 4.4 résume les modifications apportées par le nombre de bins

sur la taille du descripteur de teinte.

TABLE 4.3 – Influence du nombre de bins sur le temps de réponse du descripteur de teinte.

		SVM		KNN	
		T.T[S]	T.M[S]	T.T[S]	T.M[S]
COIL-100	6 bins	1297	0.52	19	0.01
	9 bins	2420	0.97	25	0.01
	12 bins	2576	1.03	92	0.04
	15 bins	3771	1.51	108	0.04
ALOI-angle de vue	6 bins	78575	31.41	116	0.05
	9 bins	910	0.36	175	0.07
	12 bins	2103	0.84	399	0.16
	15 bins	7799	3.11	479	0.19
ALOI-éclairage	6 bins	15350	12.28	61	0.05
	9 bins	365	0.29	82	0.07
	12 bins	907	0.73	238	0.19
	15 bins	3331	2.66	448	0.36

TABLE 4.4 – Influence du nombre de bins sur la taille du descripteur de teinte.

Nombre de bins	Taille du descripteur
6 bins	150 valeurs
9 bins	225 valeurs
12 bins	300 valeurs
15 bins	375 valeurs

De ces deux tableaux 4.3 et 4.4 aussi, on note également le temps de réponse remarquablement important qui a été dépensé par les classificateurs SVM dans la tâche de catégorisation des images de la base ALOI à l'aide du descripteur DT_6. Ce long temps dépensé confirme la difficulté rencontrée par les classificateurs SVM dans la tâche de trouver une marge de séparation entre les différents objets, car la taille du descripteur DT_6 n'était pas suffisante par rapport à la taille des images de cette base de données ALOI, contrairement à la classification des images de COIL-100 en utilisant ce descripteur DT_6. Donc pour conclure, nous devons choisir un nombre de bins universel pour toutes les tailles des images, pour cela nous avons choisi le descripteur DT_{12} pour les raisons suivantes :

— La relation de corrélation positive entre le nombre de bins et le taux de reconnaissance.

— La non-monotonie dans la classification utilisant le descripteur DT_{15}, car dans la classification de ALOI-angle de vue en utilisant les KNN, le descripteur DT_{12} a un meilleur taux de reconnaissance que celui de DT_{15}.

— DT_{15} à un temps de réponse remarquablement grand par rapport à celui de DT_{12}, et aussi 75 plus de valeur en taille de descripteur, tandis que la différence de taux de reconnaissance n'est pas grande entre les deux, c'est pourquoi nous avons préféré utiliser le DT_{12}.

4.7.1.4 Influence de la base d'apprentissage sur les performances du descripteur de teinte

Dans cette partie du test, nous nous concentrons sur l'influence de la taille de la base d'apprentissage sur la catégorisation des images d'objets à travers le descripteur de teinte proposé. Pour cela, nous avons fait un test sur 50 objets. Pour COIL-100 et ALOI-angle de vue, chaque objet était représenté par 50 images de test, tandis que pour l'éclairage ALOI, chaque objet était représenté par 25 images de test.

Les tests sont réalisés en faisant varier le nombre d'images d'apprentissage pour chaque objet, nous avons pris : 1 image d'apprentissage pour chaque objet (donc 50 images dans la base d'apprentissage), 3 images d'apprentissage pour chaque objet (donc 150 images dans la base d'apprentissage), 5 images d'apprentissage pour chaque objet (donc 250 images dans la base d'apprentissage), 7 images d'apprentissage pour chaque objet (donc 350 images dans la base d'apprentissage), 9 images d'apprentissage pour chaque objet (donc 450 images dans la base d'apprentissage), 11 images d'apprentissage pour chaque objet (donc 550 images dans la base d'apprentissage).

La classification a été réalisée à l'aide de classificateurs SVM et KNN, les résultats sont présentés dans les figures 4.18, 4.19 et 4.20. Tandis que le tableau 4.5

représente l'influence de la variance de la taille de la base d'apprentissage sur le temps de calcul du descripteur de teinte.

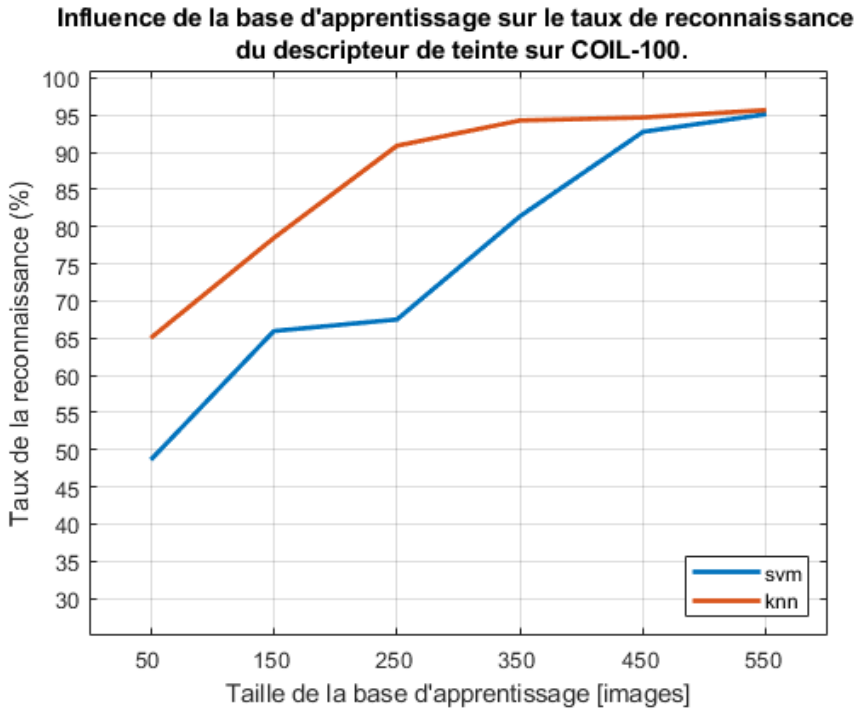


FIGURE 4.18 – Influence de la base d'apprentissage sur le taux de reconnaissance du descripteur de teinte sur COIL-100.

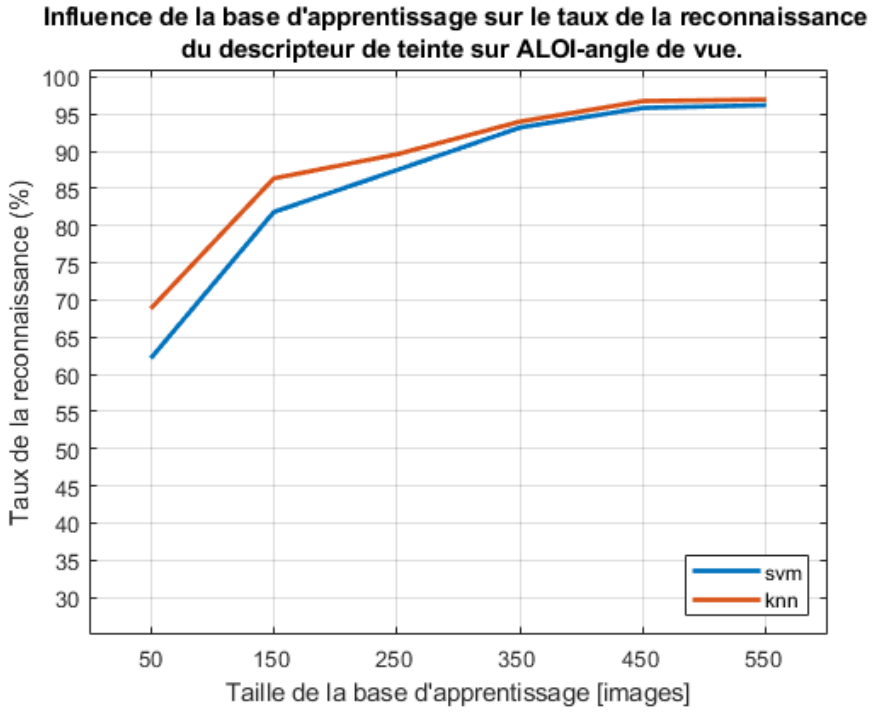


FIGURE 4.19 – Influence de la base d'apprentissage sur le taux de reconnaissance du descripteur de teinte sur ALOI-angle de vue.

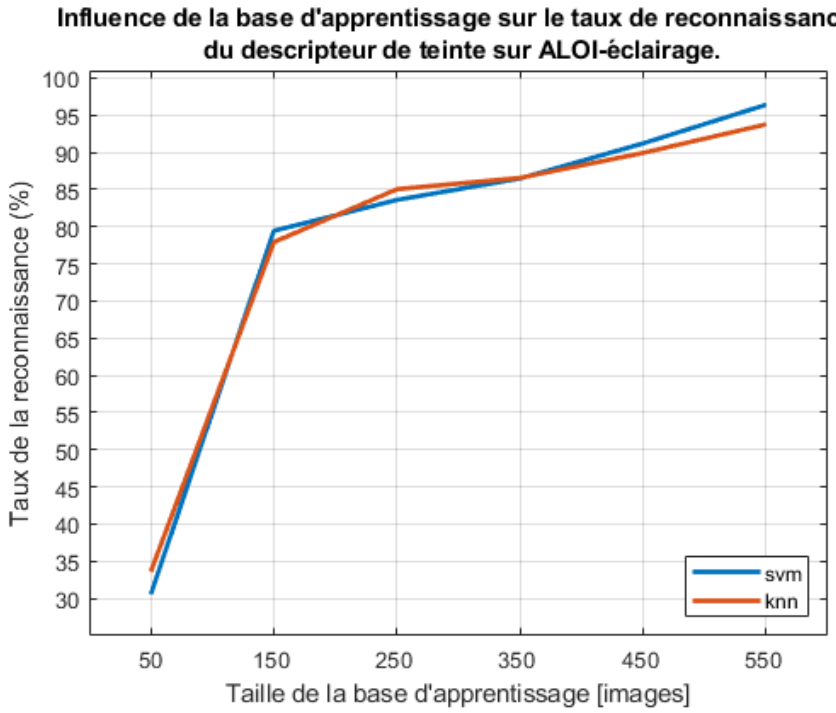


FIGURE 4.20 – Influence de la base d'apprentissage sur le taux de reconnaissance du descripteur de teinte sur ALOI-éclairage.

TABLE 4.5 – Influence de la base d'apprentissage sur le temps de réponse du descripteur de teinte.

		SVM		KNN	
		T.T[S]	T.M[S]	T.T[S]	T.M[S]
COIL-100	50 images	709	0.28	18	0.01
	150 images	840	0.34	18	0.01
	250 images	4686	1.87	19	0.01
	350 images	5074	2.03	20	0.01
	450 images	3175	1.27	42	0.02
	550 images	2576	1.03	92	0.04
ALOI-angle de vue	50 images	923	0.37	86	0.03
	150 images	949	0.38	82	0.03
	250 images	1012	0.40	78	0.03
	350 images	1262	0.50	84	0.03
	450 images	1724	0.69	106	0.04
	550 images	2103	0.84	399	0.16
ALOI-éclairage	50 images	428	0.34	47	0.04
	150 images	2331	1.86	37	0.03
	250 images	541	0.43	42	0.04
	350 images	791	0.63	50	0.04
	450 images	794	0.64	148	0.12
	550 images	907	0.73	238	0.19

La figure 4.18, la figure 4.19 et la figure 4.20 montrent qu'il existe une relation de corrélation positive entre la taille de la base de données d'apprentissage et le taux de reconnaissance des descripteurs de teinte, cela signifie que l'augmentation de la taille de la base de données d'apprentissage améliore le taux de reconnaissance du descripteur proposé. Sinon, cette augmentation peut ralentir la tâche de catégorisation et également augmenter le stockage et la taille de la mémoire requis. Par conséquent, nous devons ajuster la taille de la base de données d'apprentissage pour trouver un équilibre entre le taux de reconnaissance et les autres performances (temps de calcule et taille de stockage requis) à cette fin, nous recommandons d'utiliser de 30 % à 35 % des base de test dans les tâches d'apprentissage.

La taille de la base d'apprentissage n'a pas une influence sur la taille de descripteur vu que nous avons utilisé le descripteur DT_{12} seulement. Pour le temps de calcul, les résultats affichés dans le tableau 4.5 montrent que l'évolution du temps de catégorisation en fonction de la taille d'apprentissage n'est pas monotone pour les classificateurs SVM, donc la base d'apprentissage n'a pas d'influence directe sur le temps de catégorisation dans le cas de la classification basée sur les classificateurs SVM contrairement à la classification basé sur KNN.

4.7.1.5 Influence de la base de test sur les performances du descripteur de teinte

Dans cette partie de test, nous nous concentrons sur l'influence de la taille de test sur la catégorisation d'objets en utilisant le descripteur de teinte. Pour COIL-100 et ALOI-angle de vue chaque objet était représenté par 50 images de test, tandis que pour ALOI-éclairage chaque d'objets était représenter par 25 images de test. Les tests sont effectués en faisant varier le nombre d'objets inclus dans le test, nous avons pris : 10 objets (donc 500 images dans la base de test pour COIL-100 et ALOI-angle de vue et 250 images pour ALOI-éclairage), 20 objets, 30 objets 40 objets et 50 objets. La classification a été effectuée à l'aide de classificateurs SVM et KNN et les résultats sont présentés dans les figures 4.21, 4.22 et 4.23, tandis que

le tableau 4.6 représente l'influence de la variance de la taille de la base de test sur le temps de calcul du descripteur de teinte.

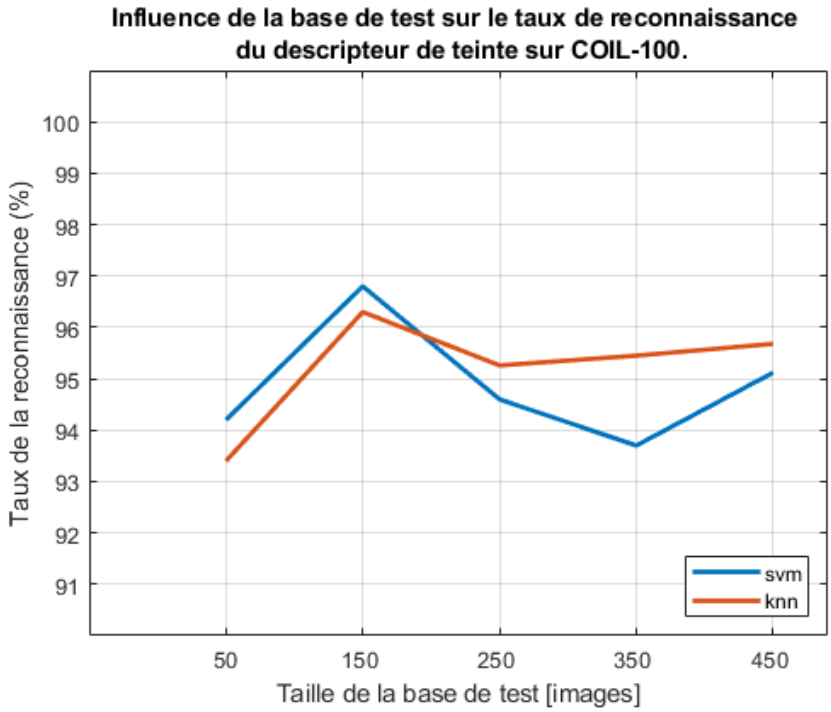


FIGURE 4.21 – Influence de la base de test sur le taux de reconnaissance du descripteur de teinte sur COIL-100.

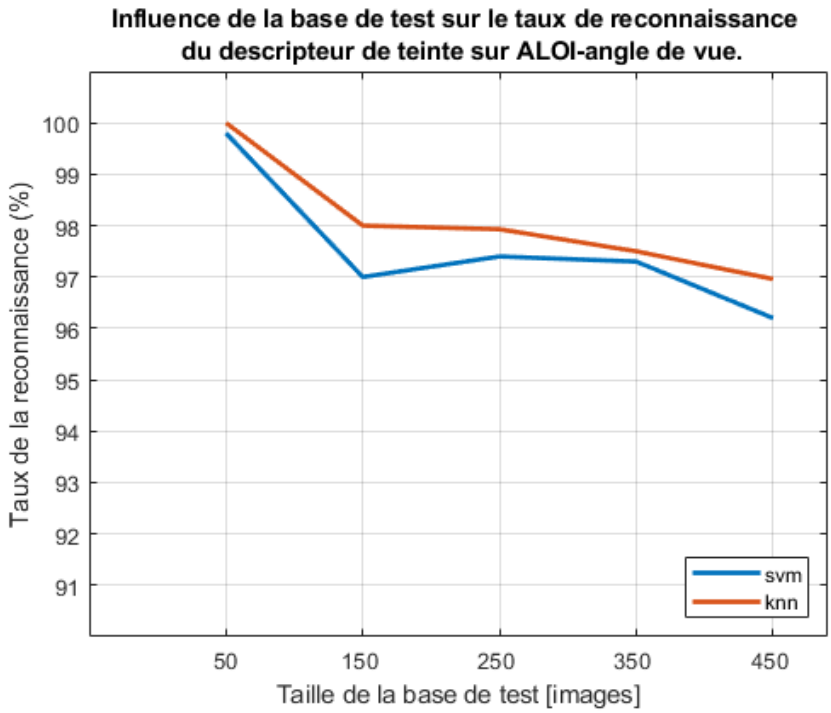


FIGURE 4.22 – Influence de la base de test sur le taux de reconnaissance du descripteur de teinte sur ALOI-angle de vue.

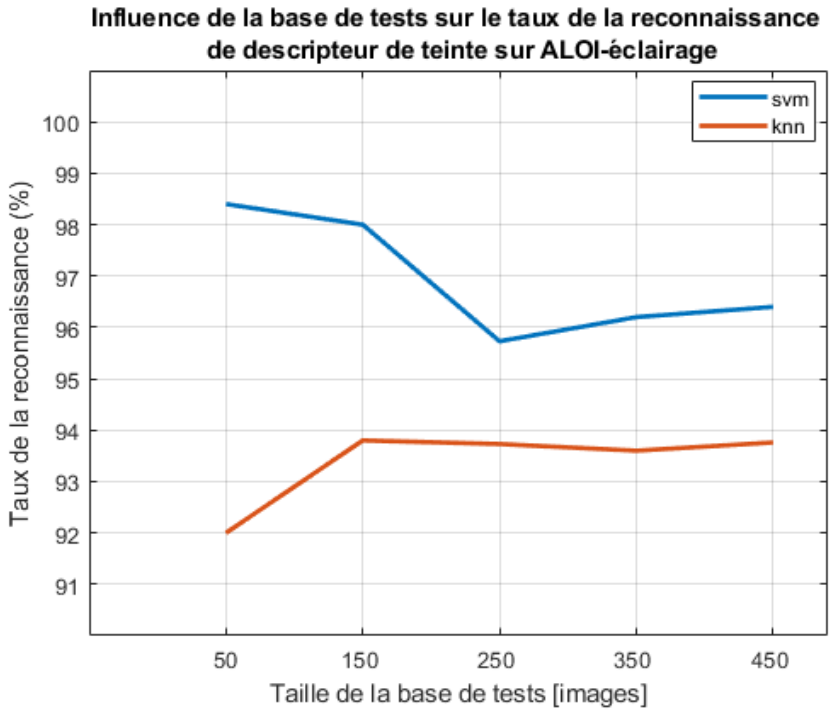


FIGURE 4.23 – Influence de la base de test sur le taux de reconnaissance du descripteur de teinte sur ALOI-éclairage.

TABLE 4.6 – Influence de la base des tests sur le temps de réponse du descripteur de teinte.

		SVM		KNN	
		T.T[S]	T.M[S]	T.T[S]	T.M[S]
COIL-100	500 images	34	0.07	3	0.01
	1000 images	136	0.14	7	0.01
	1500 images	373	0.25	11	0.01
	2000 images	2818	1.40	16	0.01
	2500 images	2576	1.03	92	0.04
ALOI-angle de vue	500 images	54	0.11	18	0.04
	1000 images	195	0.20	37	0.04
	1500 images	504	0.34	63	0.04
	2000 images	1028	0.52	81	0.04
	2500 images	2103	0.84	399	0.16
ALOI-éclairage	500 images	29	0.12	8	0.03
	1000 images	94	0.19	15	0.03
	1500 images	240	0.32	123	0.16
	2000 images	551	0.55	135	0.14
	2500 images	907	0.73	238	0.19

Sur la figure 4.21 et la figure 4.23, les deux courbes, SVM et KNN, ne sont pas monotones, tandis que sur la figure 4.22, les deux courbes, SVM et KNN, sont dé-

croissantes, ce qui signifie que l'augmentation du nombre d'objets inclus dans le test a diminué le taux de reconnaissance. Ce phénomène revient à l'utilisation de la couleur dans la catégorisation des objets par notre descripteur de teinte. La couleur des nouveaux objets inclus dans le test affecte le taux de reconnaissance, si ces nouveaux objets ont des couleurs différentes des objets qui sont déjà dans le test, le taux de reconnaissance augmentera, sinon le taux de reconnaissance diminuera. À notre avis, ce phénomène est logique et il existe dans toutes les méthodes de catégorisation des images d'objets, de plus il ne remet pas en cause l'efficacité de notre descripteur de teinte. Cette efficacité fera l'objet de l'étude comparative dans la section suivante.

Pour le temps de catégorisation (tableau 4.6), la logique a été respectée dans ce test, plus le nombre d'objets inclus dans le test augmente, plus le temps nécessaire pour terminer la tâche de catégorisation augmente également.

Pour la taille du descripteur, et similaire au test précédent, elle est restée fixe puisque nous n'utilisons que le descripteur DT_{12} seulement.

4.7.1.6 Influence du classificateur sur les performances du descripteur de teinte

Dans cette partie du test, nous visons à utiliser les résultats des tests précédents (l'influence de la variation du nombre de cellules, l'influence de la variation du nombre de bins, l'influence de la variation de la taille de la base d'apprentissage et l'influence de la variation de la taille de la base de test) pour étudier l'influence du classifieur sur les performances de notre descripteur de teinte en termes de taux de reconnaissance et de temps de calcul du système de catégorisation d'objets.

Pour le temps de réponse, dans tous les tests les classificateurs KNN sont clairement rapide par rapport aux les classificateurs SVM. Pour le taux de reconnaissance, dans les tests de COIL-100 et ALOI-angle de vue, c'est vrai que les résultats en utilisant KNN ont été généralement meilleur, mais la différence avec les SVM n'été pas grande (moins de 1%). Cependant les KNN nous ont déçus dans classifi-

cation d'ALOI-éclairage (figure 4.23), où le taux de reconnaissance des SVM a été clairement meilleur par rapport aux KNN avec une grande différence (plus de 3%). À cause de cette différence, nous avons choisi d'utiliser les classificateurs SVM dans notre article [42].

4.7.2 Tests comparatifs

Dans cette partie, nous visons à faire une étude comparative entre le descripteur de teinte proposé (HUE D) [42] et d'autres méthodes existantes : l'histogramme de gradient orienté (HOG) [24], histogrammes de teinte (Hue H) [121], histogrammes de l'adversaire (OPP) [121] et le descripteur d'image local de réponse de filtre de Gabor (Gabor) [138]. Nous nous concentrons d'abord sur le taux de reconnaissance de chaque méthode, puis nous passons à la comparaison du temps de calcule et de la taille de ces descripteurs.

4.7.2.1 Configuration de tests

L'environnement logiciel et matériel est le même que celui mentionné ci-dessus et utilisé dans l'étude paramétrique. Ce test est divisé en deux parties, dans la première partie (test de stabilité face aux changements géométriques),nous utilisant la base de données COIL-100 et la base ALOI-angle de vue. Alors que dans la deuxième partie (test de stabilité contre les changements de conditions d'éclairage), nous utilisons la base ALOI-éclairage.
Les paramètres des descripteurs utilisés dans cette partie de test sont les suivants : Pour HOG, 3 est considéré comme le nombre de compartiments $N = M = 3$. Pour les histogrammes adverses et les histogrammes de teinte, 36 est utilisé comme nombre de bins, 2 comme facteur de lissage et lambda = 1. Pour le descripteur Gabor nous utilisons un descripteur complet qui utilise une grille 5×5 pour la mise en commun spatiale, tandis que, pour le descripteur de teinte proposé, le nombre de cellules est 25 et le nombre de bins est 12. Le descripteur d'image locale de réponse de filtre de Gaborne sera pas testé sur la base COIL-100, car ce descripteur ne prend pas en

charge la taille des images de cette base de données.

4.7.2.2 Test de stabilité contre les changements géométriques

Dans cette partie nous avons effectué un test sur 50 objets, chaque objet a été représenté avec 72 images, 22 images sont utilisées dans la phase d'apprentissages et 50 images dans le test. Les classificateurs sont formés pour utiliser toutes les images sauf une. L'image utilisée dans l'ensemble d'apprentissage n'est pas utilisée dans le test. Les tableaux 4.7 et 4.8 montrent le pourcentage de taux de reconnaissance (Taux.R [%]), le temps total de tests en seconde (T.T [S]) et le temps moyen en seconde (T.M [S]) en utilisant des classificateurs SVM. Alors que Les tableaux 4.9 et 4.10 montrent les mêmes résultats en utilisant des classificateurs KNN.

TABLE 4.7 – Test de stabilité contre les changements de conditions géométriques sur COIL-100 à l'aide de SVM.

	Taux.R [%]	T.T [S]	T.M [S]
HOG [24]	84.04	988	0.39
OPP [121]	87.28	4947	1.98
Hue h [121]	89.28	4949	1.98
HUE D [42]	95.12	2576	1.03

TABLE 4.8 – Test de stabilité contre les changements de conditions géométriques sur ALOI-angle de vue à l'aide de SVM.

	Taux.R [%]	T.T [S]	T.M [S]
HOG [24]	85.36	1175	0.47
OPP [121]	79.40	11792	4.71
Hue h [121]	80.12	20074	8.02
Gabor [138]	74.36	75405	30.02
HUE D [42]	96.20	2103	0.84

TABLE 4.9 – Test de stabilité contre les changements de conditions géométriques sur COIL-100 de vue à l'aide de KNN.

	Taux.R [%]	T.T [S]	T.M [S]
HOG [24]	81.76	42	0.02
OPP [121]	87.04	309	0.12
Hue h [121]	92.48	133	0.05
HUE D [42]	95.68	92	0.04

TABLE 4.10 – Test de stabilité contre les changements de conditions géométriques sur ALOI-angle de vue à l'aide de KNN.

	Taux.R [%]	T.T [S]	T.M [S]
HOG [24]	85.08	266	0.11
OPP [121]	84.96	99	0.04
Hue h [121]	84.96	105	0.04
Gabor [138]	86.76	970	0.38
HUE D [42]	96.96	399	0.16

D'après les résultats présentés dans les quatre derniers tableaux, le descripteur de teinte proposée a le meilleur taux de reconnaissance dans toutes les expériences, 95.12 % sur COIL-100 en utilisant le SVM, 96.20 % sur ALOI-angle de vue en utilisant SVM, 95.68 % sur COIL-100 en utilisant KNN et 96.96 % sur ALOI-angle de vue en utilisant KNN.

La supériorité de la couleur dans la tâche de reconnaissance et de catégorisation des objets a été confirmée par ces tests. On voit que dans tous les tests, les résultats du descripteur utilisant des couleurs (histogramme de teinte [121], histogrammes de l'adversaire [121] et le descripteur de teinte [42]) sont meilleurs que les autres descripteurs. L'efficacité de la méthode cellules-bins face aux changment dans les conditions géométriques a également été confirmée par ces tests. Il est facile de remarquer que notre descripteur de teinte proposé a un taux de reconnaissance plus élevé que les autres méthodes dans tous les tests. Alors que dans les résultats du HOG, cette méthode de cellule et de bin n'était pas suffisante pour obtenir de bons résultats, en raison de l'utilisation du gradient qui n'était pas suffisant pour permettre au descripteur d'avoir une bonne distinction entre les objets.

D'après les résultats des tests COIL-100 (Tableau 4.7 et Tableau 4.9), on peut

remarquer que la combinaison de la couleur avec la méthode des cellules et des bins proposée dans le descripteur de teinte a amélioré le taux de reconnaissance HOG d'environ 11% dans les tests utilisant les classificateurs SVM et de 14% dans les tests utilisant les classificateurs KNN. Mais le temps de catégorisation est passé de 0.39 seconde à 1.03 seconde dans les tests utilisant les classificateurs SVM, et de 0.02 second à 0.04 seconde dans les tests de classificateurs KNN. Même chose dans les résultats des tests ALOI-view (Tableau 4.8 et Tableau 4.10), le taux de reconnaissance de descripteur de teinte s'est amélioré d'environ 11% dans les tests de classificateurs SVM et de 12% dans les tests de classificateurs KNN. Alors que le temps de catégorisation est passé de 0.47 seconde à 0.83 seconde dans les tests de classificateurs SVM, et de 0.11 second à 0.16 seconde dans les tests de classificateurs KNN.

À partir résultat des histogrammes d'adversaire et des histogrammes de teinte, ainsi que du résultat de Gabor sur les ensembles de données ALOI-angle de vue, nous pouvons conclure que ces descripteurs n'ont pas une stabilité suffisante contre les changements de conditions géométriques. En utilisant la méthode des cellules et des bins, nous avons réussi à renforcer la stabilité de notre descripteur de teinte, c'est pourquoi il y a une grande amélioration dans le taux de reconnaissance de descripteur de teinte par rapport à ces descripteurs dans tous les tests. Nous notons également le temps de catégorisation très élevé de ces descripteurs par rapport au temps de catégorisation de descripteur proposé, ce qui démontre les difficultés des tâches de catégorisation utilisant ces descripteurs en présence de changements de conditions géométriques. Nous rappelons que nous n'avons pas fait de test de descripteur Gabor sur COIL-100, car ce descripteur ne prend pas en charge la taille d'image COIL-100 (128×128 pixels).

Le taux de reconnaissance de descripteur de teinte propose ainsi que son temps de catégorisation en utilisant les classificateurs KNN sont meilleur que ceux en utilisant les classificateur SVM sur les deux bases de données, COIL-100 et ALOI-angle de vue.

4.7.2.3 Test de stabilité contre les changements de condition d'éclairage

Cette partie du test est effectuée sur les mêmes 50 objets utilisés dans le test ALOI-angle de vue, mais cette fois nous utilisons la partie ALOI-éclirage. Chaque objet a été représenté avec 11 images dans l'ensemble d'apprentissage et 25 images dans les tests. Chaque image a été encodée à l'aide de descripteurs HOG (HOG) [24], d'histogrammes de l'adversaire (OPP) [121], d'histogrammes de teinte (Hue H) [121], de descripteurs de Gabor (Gabor) [138] et du descripteur de teinte (HUE D) [42] et le descripteur de teinte proposé. Les paramètres des descripteurs sont les mêmes que ceux de la partie précédente du test.

Dans cette partie du test également, nous nous concentrons sur le taux de reconnaissance des méthodes et le temps moyen de catégorisation d'un objet. Ce dernier est calculé en divisant le temps total de catégorisation de chaque méthodes sur le nombre total d'images à catégoriser (1250 images dans cette partie du test). L'environnement logiciel et matériel est le même que celui cité précédemment.

Les classificateurs sont formés pour utiliser toutes les images sauf une, cette dernière est utilisée dans le test. Le tableau 4.11 montre le pourcentage de taux de reconnaissance (Taux.R [%]), le temps total de tests en seconde (T.T [S]) et le temps moyen en seconde (T.M [S]) en utilisant les classificateurs SVM. Alors que le tableau 4.12 montre ces résultats en utilisant des classificateurs KNN.

TABLE 4.11 – Test de stabilité contre les changements de conditions d'éclairage sur ALOI-éclairage à l'aide de SVM.

	Taux.R [%]	T.T [S]	T.M [S]
HOG [24]	81.84	341	0.27
OPP [121]	88.48	1913	1.37
Hue h [121]	91.76	910	0.73
Gabor [138]	91.92	2048	1.67
HUE D [42]	96.40	907	0.73

TABLE 4.12 – Test de stabilité contre les changements de conditions d'éclairage sur ALOI-éclairage à l'aide de KNN.

	Taux.R [%]	T.T [S]	T.M [S]
HOG [24]	82.64	134	0.11
OPP [121]	89.04	445	0.36
Hue h [121]	93.52	231	0.18
Gabor [138]	91.76	414	0.33
HUE D [42]	93.76	238	0.19

Les résultats présentés dans les tableaux 4.11 et 4.12 prouvent que le descripteur de teinte proposé à de meilleures performances que toutes les autres méthodes. Il a un taux de reconnaissance de 96.40% en utilisant les classificateurs SVM et de 93.76% en utilisant les classificateurs KNN.

La propriété théorique des teintes a été prouvée par les résultats présentés dans le tableau 4.11 et le tableau 4.12, du fait de la stabilité de cette composante face aux changements de conditions d'éclairage, les descripteurs qui utilisant les valeurs de teinte (l'histogramme de teinte et le descripteur de teinte) étaient plus stable et obtient les taux de reconnaissance les plus élevés par rapport aux autres méthodes.

La combinaison de la teinte avec la méthode des cellules et des bins nous a permis de résoudre le problème de la stabilité manquante contre les ombres et les changements de géométrie d'éclairage tels que l'ombrage, la couleur de l'éclairage et la diffusion de la lumière, les reflets des objets sous une source de lumière blanche et la diffusion d'une source de lumière blanche. En résolvant tous ces problèmes dans le descripteur de teinte proposé, nous avons réussi à augmenter le taux de reconnaissance de 81.84% à 96,40% en utilisant des classificateurs SVM, et de 82,64% à 93,76% en utilisant des classificateurs KNN.

Comparé à l'histogramme de teinte [121], le descripteur de teinte proposé a obtenu un meilleur taux de reconnaissance en utilisant les deux classificateurs SVM et KNN, et ce qui confirme à nouveau ses performances exceptionnelles. Cette haute performance est due à utilisation de la totalité de l'image dans le descripteur de teinte avec la méthode de cellules et de bins, ainsi qu'à l'utilisation de la constance de couleur Gray-Edge a résolu le problème de manque d'invariance face

aux changements de couleur d'éclairage et de diffusion de la lumière posés dans l'histogramme de teinte.

Les résultat du descripteur de Gabor [138] sont bien meilleurs dans cette partie du test (Tableau 4.11 et Tableau 4.12) par rapport aux changements de conditions géométriques (tableau 4.8 et tableau 4.10) ce qui confirme que ce descripteur est conçu pour contrer uniquement les changements de condition d'éclairage.

Le descripteur de teinte proposé résout les inconvénients des histogrammes de l'adversaire en utilisant la teinte qui est invariante au changement d'intensité lumineuse (changement dans les ombres et les changements de géométrie d'éclairage tels que l'ombrage), au décalage d'intensité lumineuse (reflets des objets sous une source de lumière blanche et à la diffusion d'une source blanche) et au changement et le décalage de l'intensité lumineuse (combinaisons des deux conditions ci-dessus). L'utilisation de la constance de la couleur Gray-Edge a également résolu le problème de l'invariance manquante au changement de la couleur de la lumière (changement de la couleur de l'éclairage et de la diffusion de la lumière). Cette invariance supplémentaire a amélioré le taux de reconnaissance de l'histogramme de l'adversaire d'environ 12 % en utilisation les SVM (tableau 4.12) et d'environ 5% en utilisant les KNN (tableau 4.12).

Dans cette partie du test (Tableau 4.11 et Tableau 4.12), et contrairement aux tests de COIL-100 et ALOI-angle de vue, le taux de reconnaissance du descripteur de teinte utilisant les classificateurs SVM est meilleur que en utilisant des classificateurs KNN. Bien que les classificateurs KNN aient été plus rapides dans tous les tests, à notre avis, les KNN ont échoué dans les tests d'ALOI-éclairage.

Tableau 4.13 récapitulé les tailles des différents descripteurs utilisés dans cette étude comparative, l'histogramme de gradient orienté (HOG), histogrammes de teinte (Hue H), histogrammes de l'adversaire (OPP), le descripteur d'image locale de réponse de filtre de Gabor (Gabor), et le descripteur de teinte (HUE D).

TABLE 4.13 – Comparaison entre les tailles des différents descripteurs.

	Taille du descripteur
HOG [24]	81 valeurs
OPP [121]	4608 valeurs
Hue h [121]	4608 valeurs
Gabor [138]	92928 valeurs
HUE D [42]	300 valeurs

Le tableau 4.13 confirme la performance exceptionnelle de descripteur de teinte en termes de la taille du descripteur. Avec 300 valeurs de caractérisation seulement, ce descripteur est idéal pour économiser la taille de mémoire et de stockage requis.

4.8 Conclusion

Dans ce chapitre, nous avons présenté une nouvelle méthode pour la reconnaissance et la catégorisation d'objets couleur faits à la main avec un fond uniforme. Dans un premier temps, nous avons proposé d'utiliser la composante de teinte dans le système de couleur HSV pour tirer parti de l'invariance ce cette composante de chrominance contre les changements des conditions d'éclairage, et aussi d'applique la constance de couleur Gray-Edge pour éliminer l'influence du changement de la couleur d'éclairage sur les couleurs de l'objet. Par la suite nous avons proposé d'utiliser la méthode des cellules et des bins pour rajouter au descripteur proposé l'invariance nécessaire contre les changements des conditions géométriques.
Dans un second temps, nous avons exposé les performances obtenues par notre descripteur et nous les avons comparées à celles des descripteurs classiques de la littérature.

CHAPITRE 5

Conclusion générale

« L'expérience est une observation provoquée dans le but de faire naître une idée.»

Claude Bernard

Conclusion générale

L'objectif de cette thèse était de réaliser de la catégorisation des objets couleur. Nous nous sommes attachés à mettre au point un processus de catégorisation d'objet rapide et fiable. Ce processus devait s'inscrire dans un système complet de reconnaissance, traitant en temps réel les informations provenant d'une caméra. Moins complexe au niveau du temps de calcul et de l'allocation mémoire. Notre cahier des charges définissait trois propriétés principales :

— Efficacité : le système proposé doit être invariant aux translations, rotations de l'objet, invariant à la résolution spatiale des images, invariant aux Modifications d'illumination, invariante aux Déformations, invariant aux fonds chargés, invariants à l'occultation et aussi doit pouvoir discriminant. En un seul mot, le système proposé doit avoir un taux de reconnaissance d'objets acceptable.

— Vitesse : l'analyse doit être efficiente, de manière qu'il respecte le compromis temps de calcul/ qualité de catégorisation.

— Optimisation de stockage : le système proposé doit avoir la capacité d'optimiser l'utilisation des espaces mémoire.

À cette fin, un nouveau modèle de reconnaissance d'objets en couleur a été proposé qui est analysé en théorie et évalué en pratique dans le but de reconnaître des objets multicolores.

Pour respecter ces propriétés, nous avons mené notre étude en deux étapes. La première étape consistait en une analyse des modèles de reconnaissance d'objets existants. Nous avons tout d'abord étudié les différentes techniques utilisées pour chaque étape dans le processus de reconnaissance d'objets, afin d'établir un ensemble des critères (dérivés également de notre cahier des charges) permettant de gérer les contraintes présentées lors de l'utilisation de ces techniques. Pour étudier ces contraintes, nous avons présenté une taxonomie des différentes étapes d'un al-

gorithme générique pour la reconnaissance d'objets. Partant de cette analyse, nous avons pu déterminer les parties sur lesquelles nous devions se concentrer dans un algorithme de reconnaissance d'objets pour respecter les critères définis dans le cahier de charge. Dans ce contexte, nous avons choisi de nous concentrer sur les deux premières parties d'un système de reconnaissance d'objets (extraction des primitives, description de ces primitives).

Pour l'étape d'extraction des primitives, nous avons proposé un système de catégorisation d'objets basé sur la couleur des objets, nous avons choisi d'utiliser les valeurs de la teinte dans l'espace de couleur HSV. Le descripteur orienté proposé profite de l'invariance des teintes contre les ombres, les changements de géométrie d'éclairage tels que l'ombrage, les reflets des objets sous une source de lumière blanche et à la diffusion d'une source blanche, et résout son manque d'invariance contre le changement de la couleur de l'éclairage et de la diffusion de la lumière par l'utilisant la constance de couleur Gray-Edge.

Pour la deuxième étape de description des primitives, nous avons proposé d'utiliser la méthode des cellules et des bins dans la construction de descripteurs de teinte proposer pour rajouter à ce dernier l'invariance nécessaire contre les transformations géométriques et photométriques et donc augmenter le taux de reconnaissance. Cette idée est basée sur la division de l'image en vingt-cinq cellules superposées de 50% pour faire fonctionner le descripteur sur les cellules locales. Le descripteur cellule est construit en calculant la valeur de teinte sur chaque pixel et en accumulant la magnitude des teintes dans douze cases appellent bins. En regroupant ces valeurs de magnitude de bins dans un vecteur, nous obtenons un vecteur de caractérisation pour la cellule, après la normalisation de la fonction, et en regroupant toutes les caractéristiques des vingt-cinq cellules dans un seul vecteur, nous obtenons le descripteur final de l'images.

Pour la tâche de classification, nous avons proposé l'utilisation de deux classificateurs puissants et très utilisés, les classificateurs SVM (Support Vector Machine) et KNN (k-plus proche voisin). La méthode proposée a été évaluée sur deux en-

sembles de données accessibles au public, Columbia Object Image Library (COIL-100) et The Amsterdam Library of Object Images (ALOI). Le modèle proposé excelle par sa capacité à distinguer des objets similaires avec une grande précision, sa stabilité face à tous types de changements des conditions géométriques et photométriques et aussi sa robustesse dans la catégorisation de tous types d'objets fabriqués à la main.

Les tests ont prouvé non seulement les performances exceptionnelles de ce descripteur de teinte proposé par rapport aux méthodes existantes en termes de taux de reconnaissance, mais aussi sa rapidité et sa capacité à optimiser l'utilisation des espaces mémoire. Cela montre également que ce modèle de reconnaissance est prometteur et pourrait faire l'objet d'applications industrielles.

La modification du type de données (par exemple en utilisant la fonctionnalité proposée pour détecter des piétons (humains) ou d'autres objets tels que voiture, chien, chat, etc., ou en utilisant un fond non uniforme) fera l'objet de nos futurs travaux. Nous souhaitons également que la fonctionnalité proposée dans cette thèse soit à la base d'une application qui remplacera l'utilisation du code à barres dans les magasins commerciaux (centre commercial, pharmacies. . . , etc.).

Annexes

ANNEXE A

Abréviations et notations

Abréviations

ALOI Amsterdam Library of Object Images

BRIEF Binary Robust Independent Elementary Features

CCD Charged Coupled Device

CDW Class Dependent Weighted

CIE Commission internationale de l'Éclairage

CMFs Color Matching Functions

CNN Convolutional Neural Network

COIL Columbia Object Image Library

CPU Central Processing Unit

CRF Conditional Random Fields

CSSDs Curvature Scale Space Descriptors

DOG Difference Of Gaussians

DPS Distribution de puissance spectrale

EBR Edge Based Region detector

FAST Features from Accelerated Segment Test

FELICM Fuzzy Edge and Local Information C-Mean

FPGA Field Programmable Gate Arrays

GLCM Grey Level Co-occurrence Matrix

GMRF Gaussiens Markoviens Random Fields

GPU Graphics Processing Unit

GW Global Weighted

HOG Histograms Of Oriented Gradients

HSL Hue Saturation Luminance

HSV Teinte Saturation Valeur

ISH Intensity Saturation Hue

KNN K-Nearest Neighbors

LBP Local Binary Patterns

LGN Corps Géniculé Latéral

LOG Laplacian Of Gaussian

LP Linear Programming

MPEG Moving Picture Experts Group

MRF Markoviens Random Fields

MSE Mean Square Error

OCR Optical Character Recognition

PDF Probability Density Function

PSNR Peak Signal to Noise Ratio

QP Quadratic Programming

RMN Relative Motion Network

RGB Rouge Vert Bleu

SIFT Scale Invariant Features Transform

SPM Spatial Pyramid Matching

SSD Sum of Squared Differences

SURF Speeded Up Robust Features

SUSAN Smallest Univalue Segment Assimilating Nucleus

SVM Support Vector Machine

TDNN Time Delay Neural Network

Notations relatives aux illuminants

f	Fréquence d'un photon
λ	Longueur d'onde d'un photon
E	Énergie d'un photon

Notations relatives aux Matériaux

$R(\lambda)$	Facteur de réflectance d'un Matériau
$T(\lambda)$	Facteur de transmittance d'un Matériau
$L_r(\lambda)$	Quantité d'énergies lumineuses réfléchies par l'objet
$L_{dr}(\lambda)$	Diffuseur parfait par réflexion
$I_{transmise}$	Intensité transmise de la lumière
$I_{incidente}$	Intensité incidente de la lumière

Notations relatives à la mesure de la couleur

R	Couleur primaire rouge
G	Couleur primaire verte
B	Couleur primaire bleue
RGB	Système de couleurs primaires rouge vert et bleue
R_P	Couleur primaire rouge définie par la CIE
G_P	Couleur primaire verte définie par la CIE
B_P	Couleur primaire bleue définie par la CIE
$R_P G_P B_P$	Système de couleurs primaires rouge verte et bleue défini par la CIE
$\bar{r}(\lambda)$	Fonction colorimétrique qui correspond à la couleur primaire R_P
$\bar{g}(\lambda)$	Fonction colorimétrique qui correspond à la couleur primaire G_P
$\bar{b}(\lambda)$	Fonction colorimétrique qui correspond à la couleur primaire B_P
HSV	Système de couleurs Hue Saturation Valeur
H	Composant de teinte dans le système de couleurs HSV
S	Composant de Saturation dans le système de couleurs HSV
V	Composant de Valeur dans le système de couleurs HSV
O_1	Premier composant dans le système de couleurs d'adversaire
O_2	Deuxième composant dans le système de couleurs d'adversaire
O_3	Troisième composant dans le système de couleurs d'adversaire
O_{1x}	dérivée du premier ordre de O_1
O_{2x}	dérivée du premier ordre de O_2
ang_x^o	Angle d'dversaire dans le système de couleurs d'adversaire
∂ang_x^O	Poids de l'angle de l'adversaire dans le système de couleur de l'adversaire

CR	Réponse du capteur de caméra sensible à la couleur rouge
CG	Réponse du capteur de caméra sensible à la couleur vert
CB	Réponse du capteur de caméra sensible à la couleur bleu
SC_R	sensibilité du capteur de caméra sensible à la couleur rouge
SC_G	sensibilité du capteur de caméra sensible à la couleur vert
SC_B	sensibilité du capteur de caméra sensible à la couleur bleu

Notations relatives aux images

p	Pixel
I	Image
I^u	Image prise sous une source de lumière inconnue
I^c	Image transformée prise sous une source de lumière inconnue
I_{gamma}	Image corrigée
I_{bords}	Image en bord découpés
X	Largeur de l'image
Y	Hauteur de l'image
$I_x(x, \sigma_D)$	Dérivées du premier ordre de l'image le long de x
$I_x(y, \sigma_D)$	Dérivées du premier ordre de l'image le long de y
$D^{u,c}$	matrice diagonale qui cartographie les couleurs qui sont prises sous une source de lumière inconnue u à leurs couleurs correspondantes sous l'illuminant canonique c

ANNEXE B

Théorème de Mercer

Le théorème de Mercer donne les conditions pour qu'une fonction - appelée fonction noyau - soit équivalente à un produit scalaire.

Rappelons tout d'abord la définition d'une matrice définie positive.

— **Matrice définie positive :** Une matrice M de dimension $(n \times n)$ dans $\mathbb{R}$ est définie positive SSI :

$$\forall v \in \mathbb{R}^n \quad v^T M v \geq 0 \tag{B.1}$$

Le théorème de Mercer fournit la condition nécessaire et suffisante pour qu'une fonction soit un noyau, à savoir

— **Condition de Mercer :** La fonction $K(x,y) : \chi \times \chi \to \mathbb{R}$ est un noyau SSI

$$G = K\left(x_i, x_j\right)_{i,j=1}^{n} \tag{B.2}$$

est définie positive.

ANNEXE C

Constance de la couleur

La constance des couleurs est la capacité de reconnaître les couleurs des objets indépendamment de la couleur de la source lumineuse [31]. L'obtention de la constance des couleurs est importante pour de nombreuses applications de vision par ordinateur, telles que la récupération d'images, la classification d'images, la reconnaissance d'objets couleur et le suivi d'objets.

C.1 Hypothèse GREY-WORLD

Les valeurs d'image, $f = (R,G,B)^T$, pour une surface Lambertienne dépendent de la source de lumière $e(\lambda)$, où λ est la longueur d'onde, la réflectance de surface $s(\lambda)$ et les fonctions de sensibilité de la caméra $c(\lambda) = (R(\lambda),G(\lambda),B(\lambda))$ sont donnés par :

$$f = \int_{\omega} e(\lambda)\, s(\lambda)\, c(\lambda)\, d\lambda \tag{C.1}$$

Où ω est le spectre visible et les polices en gras sont appliqués pour les vecteurs. Nous supposons que la scène est éclairée par une seule source lumineuse. Le but de la constance des couleurs est d'estimer la couleur de la source lumineuse $e(\lambda)$

ou sa projection sur les noyaux RGB :

$$e = \begin{pmatrix} R_e \\ G_e \\ B_e \end{pmatrix} = \int_{\omega} e(\lambda)\, c(\lambda)\, d\lambda \tag{C.2}$$

Étant donné les valeurs de l'image $f(x)$, où x est la coordonnée spatiale dans l'image. La tâche de la constance des couleurs n'est pas réalisable sans autres hypothèses.

Buchsbaum [16] propose l'hypothèse du monde gris qui suppose que la réflectance moyenne dans une scène est achromatique. Dans l'ouvrage d'origine, l'hypothèse est utilisée pour déduire que la réflectance moyenne pour les régions à ondes courtes, à ondes moyennes et à ondes longues est égale. Van de Weijer et al. [123], ont utilisé une définition plus forte de la réflectance achromatique d'une scène (comme également utilisée dans [28]) :

$$\frac{\int s(\lambda, x)\, dx}{\int dx} = g(\lambda) = k \tag{C.3}$$

Ce qui évite de faire d'autres hypothèses. Buchsbaum [16], par exemple, avait besoin de faire d'autres hypothèses sur les fonctions de base pour les sensibilités de la caméra, les réflectances de surface et les spectres de la source lumineuse. Le constant k est comprise entre 0 pour aucun réflectance (noir) et 1 pour la réflectance totale (blanc) de la lumière incidente, et l'intégrale est sur le domaine de la scène. Pour une telle scène à réflectance achromatique, elle considère que la couleur réfléchie est égale à la couleur de la source lumineuse, car :

$$\frac{\int f(x)\, dx}{\int dx} = \frac{1}{\int dx} \int \int_{\omega} e(\lambda)\, s(\lambda, x)\, c(\lambda)\, d\lambda dx \tag{C.4}$$

$$= \int_{\omega} e(\lambda)\, c(\lambda) \left(\frac{\int s(\lambda, x)\, dx}{\int dx} \right) d\lambda \tag{C.5}$$

$$= k \int_{\omega} e(\lambda)\, c(\lambda)\, d\lambda = ke \tag{C.6}$$

Où van de Weijer et al. [123], ont appliqué le théorème de Fubini pour échanger l'ordre d'intégration. La couleur de la source lumineuse normalisée est calculée avec $\hat{e} = ke/|ke|$. Plus récemment Finlayson et Trezzi [28], on proposer une autre méthode basée sur la norme de Minkowski appelée Nuances de gris (Shades of Grey) et est calculée par :

$$\left(\frac{\int (f(x))^p dx}{\int dx} \right)^{\frac{1}{p}} = ke \tag{C.7}$$

Pour $p = 1$, l'équation est égale à l'hypothèse de Monde du gris.

Le calcul de la norme, donné par l'équation C.7, est une opération de moyenne globale, qui ignore la corrélation locale importante entre les pixels. Cette corrélation locale peut être utilisée pour réduire l'influence du bruit. Le lissage local comme étape de prétraitement s'est avéré bénéfique pour les algorithmes de constance des couleurs, comme discuter dans l'étude de Barnard [4]. Pour exploiter cette corrélation locale, Weijer et al. [123] introduisons un lissage local avec un filtre gaussien, G^{σ} avec un écart type σ :

$$\left(\frac{\int (f^{\sigma}(x))^p dx}{\int dx} \right)^{\frac{1}{p}} = ke \tag{C.8}$$

Où : $f^{\sigma} = f \otimes G^{\sigma}$, avec $\otimes$ est le Produit tensoriel.

C.2 Hypothèse GREY-EDGE

Comme alternative à l'hypothèse de Gray-World, Weijer et al. [123] proposons l'hypothèse de GREY-EDGE : la moyenne des différences de réflectance dans une scène est achromatique :

$$\frac{\int |s_x^{\sigma}(\lambda, x)|\, dx}{\int dx} = g(\lambda) = k \tag{C.9}$$

L'indice x indique la dérivée spatiale à l'échelle σ. Avec l'hypothèse GREY-EDGE, la couleur de la source lumineuse peut être calculée à partir de la dérivée de couleur

moyenne dans l'image donnée par :

$$\frac{\int |f_x(x)|\,dx}{\int dx} = \frac{1}{\int dx}\int\int_{\omega} e(\lambda)\,|s_x(\lambda,x)|\,c(\lambda)\,d\lambda dx \tag{C.10}$$

$$= \int_{\omega} e(\lambda)\left(\frac{\int |s_x(\lambda,x)|\,dx}{\int dx}\right) c(\lambda)\,d\lambda \tag{C.11}$$

$$= k\int_{\omega} e(\lambda)\,c(\lambda)\,d\lambda = ke \tag{C.12}$$

Où $|f_x(x)| = (|R_x(x)|,|G_x(x)|,|B_x(x)|)^T$. L'hypothèse de Gray-Edge provient de l'observation que la distribution des couleurs des images dérivées forme, une forme ellipsoïdale relativement régulière, dont le grand axe coïncide avec la couleur de la source lumineuse [122].

Semblable à la constance des couleurs basée sur Gray-World, l'hypothèse Gray-Edge peut également être adaptée pour incorporer la norme de Minkowski :

$$\left(\frac{\int |f_x^{\sigma}(x)|^p dx}{\int dx}\right)^{\frac{1}{p}} = ke \tag{C.13}$$

La constance des couleurs basée sur cette équation suppose que le p-ème norme de Minkowski de la dérivée de la réflectance dans une scène est achromatique. On distingue deux cas particuliers. Pour $p = 1$, l'illuminant est dérivé par une opération de moyenne normale sur les dérivées des canaux. Pour $p = \infty$, l'illuminant est calculé à partir de la dérivée maximale de la scène. La ressemblance entre la dérivation de la constance des couleurs de l'hypothèse Gray-World et Gray-Edge est apparente. Les deux méthodes peuvent être combinées dans un cadre unique de méthodes de constance des couleurs basées sur des caractéristiques d'image de bas niveau dérivées de l'hypothèse générale suivante :

$$\left(\frac{\int |\partial^n f^{\sigma}(x)|^p dx}{\int dx^n}\right)^{\frac{1}{p}} = ke^{n,p,\sigma} \tag{C.14}$$

La division par $\int dx$ a été incorporée dans le constant k. À côté des hypothèses déjà discutées (Gray-World, norme Minkowski et le nouveau Gray-Edge proposé), il est évident que ce cadre inclut également une constance de couleur basée sur un ordre supérieur. Les dérivées d'ordre supérieur ont des correspondances avec le mécanisme centre-surround des yeux humains pour la constance des couleurs telles qu'exploité dans l'algorithme bien connu centre surround rétines [60]. L'influence des intensités de couleur pourrait être pondérée en fonction de leur distance au centre du champ récepteur généralement calculée par une différence de fonctions gaussiennes.

L'estimation de l'illuminant de l'équation C.14 décrite un cadre pour l'estimation de l'illuminant à bas niveau. Ce cadre produit différentes estimations de la couleur de l'illuminant en fonction de trois variables :

1. L'ordre n de la structure de l'image est le paramètre déterminant si la méthode est un algorithme Grey-World ou Gray-Edge. Les méthodes Gray-World sont basées sur les valeurs RGB, tandis que les méthodes Gray-Edge sont basées sur les dérivées spatiales d'ordre n. Dans cet article, nous étudierons la constance des couleurs basée sur l'ordre supérieur jusqu'à l'ordre $n = 2$.

2. La norme Minkowski p qui détermine les poids relatifs des multiples mesures à partir desquelles la couleur finale de l'illuminant est estimée. Une norme Minkowski élevée met l'accent sur des mesures plus importantes tandis qu'une norme Minkowski faible répartit également les poids entre les mesures.

3. L'échelle des mesures locales indiquée par σ. Pour une estimation de premier ordre ou supérieure, cette échelle locale est combinée avec l'opération de différenciation calculée avec la dérivée gaussienne. Pour les méthodes du monde gris d'ordre zéro, cette échelle locale est imposée par une opération de lissage gaussien.

Références

[1] S. Abbasi, F. Mokhtarian, and J. Kittler. Enhancing css-based shape retrieval for objects with shallow concavities. *Image and vision computing*, 18(3) :199–211, 2000.

[2] S. Adam, J. Ogier, C. Cariou, R. Mullot, J. Gardes, and Y. Lecourtier. Utilisation de la transformée de fourier-mellin pour la reconnaissance de formes multi-orientées et multi-échelles : application a l'analyse automatique de documents techniques. *Traitement du signal*, 18(1) :17, 2001.

[3] P. Alessi, E. Carter, M. Fairchild, R. Hunt, C. McCamy, B. Kránicz, J. Moore, L. Morren, J. Nobbs, Y. Ohno, et al. Colorimetry. *Tech. Rep. CIE 015 : 2004*, 2004.

[4] K. Barnard, L. Martin, A. Coath, and B. Funt. A comparison of computational color constancy algorithms. ii. experiments with image data. *IEEE transactions on Image Processing*, 11(9) :985–996, 2002.

[5] S. Bauer, U. Brunsmann, and S. Schlotterbeck-Macht. Fpga implementation of a hog-based pedestrian recognition system. In *Proc. MPC-Workshop*, pages 49–58, 2009.

[6] S. Bauer, S. Köhler, K. Doll, and U. Brunsmann. Fpga-gpu architecture for kernel svm pedestrian detection. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops*, pages 61–68. IEEE, 2010.

[7] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. Speeded-up robust features (surf). *Computer vision and image understanding*, 110(3) :346–359, 2008.

[8] P. R. Beaudet. Rotationally invariant image operators. In *Proc. 4th Int. Joint Conf. Pattern Recog, Tokyo, Japan, 1978*, 1978.

[9] I. Biederman. An invitation to cognitive science, vol. 2 : Visual cognition, 1995.

[10] É. Biémont. La lumière. 1996.

[11] F. W. Billmeyer and M. Saltzman. *Principles of color technology*. Wiley New York, 1981.

[12] C. Bonnefoy and J. Dichamp. La couleur en prothèse faciale (1re partie). *Actualités odonto-stomatologiques*, (254) :97–112, 2011.

[13] A. Bosch, A. Zisserman, and X. Muñoz. Scene classification using a hybrid generative/discriminative approach. *IEEE transactions on pattern analysis and machine intelligence*, 30(4) :712–727, 2008.

[14] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE transactions on pattern analysis and machine intelligence*, 33(9) :1820–1833, 2010.

[15] A. Broggi, M. Bertozzi, A. Fascioli, and M. Sechi. Shape-based pedestrian detection. In *Proceedings of the IEEE Intelligent Vehicles Symposium 2000 (Cat. No. 00TH8511)*, pages 215–220. IEEE, 2000.

[16] G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Franklin institute*, 310(1) :1–26, 1980.

[17] J. Canny. A computational approach to edge detection. *IEEE Transactions on pattern analysis and machine intelligence*, (6) :679–698, 1986.

[18] R. Chellappa and S. Chatterjee. Classification of textures using markov random field models. In *ICASSP'84. IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 9, pages 694–697. IEEE, 1984.

[19] C. Chen, L. Pau, P. Wang, and S. Wang. Texture analysis. *Handbook of Pattern Recognition and Computer Vision*, pages 207–248, 1998.

[20] T.-W. Chen, Y.-L. Chen, and S.-Y. Chien. Fast image segmentation based on k-means clustering with histograms in hsv color space. In *2008 IEEE 10th Workshop on Multimedia Signal Processing*, pages 322–325. IEEE, 2008.

[21] C.-W. Chong, P. Raveendran, and R. Mukundan. A comparative analysis of algorithms for fast computation of zernike moments. *Pattern Recognition*, 36(3) :731–742, 2003.

[22] F. C. Crow. Summed-area tables for texture mapping. In *Proceedings of the 11th annual conference on Computer graphics and interactive techniques*, pages 207–212, 1984.

[23] N. Dalal. *Finding people in images and videos*. PhD thesis, 2006.

[24] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, volume 1, pages 886–893. IEEE, 2005.

[25] R. Durikovic and R. Kimura. Gpu rendering of the thin film on paints with full spectrum. In *Tenth International Conference on Information Visualisation (IV'06)*, pages 751–756. IEEE, 2006.

[26] R. K. Ferrell, S. S. Gleason, and K. W. Tobin Jr. Application of fractal encoding techniques for image segmentation. In *Sixth International Conference on Quality Control by Artificial Vision*, volume 5132, pages 69–77. International Society for Optics and Photonics, 2003.

[27] G. D. Finlayson, S. S. Chatterjee, and B. V. Funt. Color angular indexing. In *European Conference on Computer Vision*, pages 16–27. Springer, 1996.

[28] G. D. Finlayson and E. Trezzi. Shades of gray and colour constancy. In *Color and Imaging Conference*, volume 2004, pages 37–41. Society for Imaging Science and Technology, 2004.

[29] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern recognition*, 26(1) :167–174, 1993.

[30] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS intercommission conference on fast processing of photogrammetric data*, pages 281–305. Interlaken, 1987.

[31] D. A. Forsyth. A novel algorithm for color constancy. *International Journal of Computer Vision*, 5(1) :5–35, 1990.

[32] B. V. Funt and G. D. Finlayson. Color constant color indexing. *IEEE transactions on Pattern analysis and Machine Intelligence*, 17(5) :522–529, 1995.

[33] G. Gales, A. Crouzil, and S. Chambon. Détection de points d'intérêt pour la mise en correspondance par propagation. *actes du Congrès Reconnaissance des Formes et Intelligence Artificielle, RFIA, support électronique, Caen*, 2010.

[34] J.-M. Geusebroek, G. J. Burghouts, and A. W. Smeulders. The amsterdam library of object images. *International Journal of Computer Vision*, 61(1) :103–112, 2005.

[35] J.-M. Geusebroek, R. Van den Boomgaard, A. W. M. Smeulders, and H. Geerts. Color invariance. *IEEE Transactions on Pattern analysis and machine intelligence*, 23(12) :1338–1350, 2001.

[36] T. Gevers, J. Van De Weijer, and H. Stokman. Color feature detection, 2006.

[37] C. Goutte and E. Gaussier. A probabilistic interpretation of precision, recall and f-score, with implication for evaluation. In *European conference on information retrieval*, pages 345–359. Springer, 2005.

[38] M. Grand-Brochier. *Descripteurs 2D et 2D+ t de points d'intérêt pour des appariements robustes*. PhD thesis, 2011.

[39] C. Gu, J. J. Lim, P. Arbeláez, and J. Malik. Recognition using regions. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1030–1037. IEEE, 2009.

[40] Y. Gu, N. Narendran, T. Dong, and H. Wu. Spectral and luminous efficacy change of high-power leds under different dimming methods. In *Sixth International Conference on Solid State Lighting*, volume 6337, page 63370J. International Society for Optics and Photonics, 2006.

[41] J. Guild. The colorimetric properties of the spectrum. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 230(681-693) :149–187, 1931.

[42] R. Hamdini, N. Diffellah, and A. Namane. Robust local descriptor for color object recognition. *Traitement du Signal*, 36(6) :471–482, 2019.

[43] R. M. Haralick. Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5) :786–804, 1979.

[44] R. M. Haralick, K. Shanmugam, and I. H. Dinstein. Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, (6) :610–621, 1973.

[45] C. G. Harris, M. Stephens, et al. A combined corner and edge detector. In *Alvey vision conference*, volume 15, pages 10–5244. Citeseer, 1988.

[46] G. Healey and D. Slater. Global color constancy : recognition of objects

by use of illumination-invariant properties of color distributions. *JOSA A*, 11(11) :3003–3010, 1994.

[47] P.-Y. Hsiao, S.-Y. Lin, and S.-S. Huang. An fpga based human detection system with embedded platform. *Microelectronic Engineering*, 138 :42–46, 2015.

[48] M.-K. Hu. Visual pattern recognition by moment invariants. *IRE transactions on information theory*, 8(2) :179–187, 1962.

[49] D. Hubel. L'oeil, le cerveau et la vision. *Pour la Science, Belin*, 1994.

[50] R. S. Hunter and R. W. Harold. *The measurement of appearance*. John Wiley & Sons, 1987.

[51] A. K. Jain, S. K. Bhattacharjee, and Y. Chen. On texture in document images. In *Proceedings 1992 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 677–678, 1992.

[52] J.-M. Jolion. editeur. les systèmes de vision. traité ic2, traitement. *Hermes Paris*, 2000.

[53] V. Jumb, M. Sohani, and A. Shrivas. Color image segmentation using k-means clustering and otsu's adaptive thresholding. *International Journal of Innovative Technology and Exploring Engineering (IJITEE)*, 3(9) :72–76, 2014.

[54] Y. Ke and R. Sukthankar. Pca-sift : A more distinctive representation for local image descriptors. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.*, volume 2, pages II–II. IEEE, 2004.

[55] A. Khotanzad and Y. H. Hong. Invariant image recognition by zernike moments. *IEEE Transactions on pattern analysis and machine intelligence*, 12(5) :489–497, 1990.

[56] L. Kitchen and A. Rosenfeld. Gray-level corner detection. *Pattern recognition letters*, 1(2) :95–102, 1982.

[57] E. KOEN. Evaluation of color descriptors for object and scene recognition. In *IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, Alaska, USA, June 2008*, 2008.

[58] C.-H. Kuo and R. Nevatia. How does person identity recognition help multi-person tracking ? In *CVPR 2011*, pages 1217–1224. IEEE, 2011.

[59] J. Lafferty, A. McCallum, and F. C. Pereira. Conditional random fields : Probabilistic models for segmenting and labeling sequence data. 2001.

[60] E. H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the national academy of sciences*, 83(10) :3078–3080, 1986.

[61] E. H. Land and J. J. McCann. Lightness and retinex theory. *Josa*, 61(1) :1–11, 1971.

[62] D. Larlus. *Création et utilisation de vocabulaires visuels pour la catégorisation d'images et la segmentation de classes d'objets*. PhD thesis, 2008.

[63] S.-H. Lee, M. Bang, K.-H. Jung, and K. Yi. An efficient selection of hog feature for svm classification of vehicle. In *2015 International Symposium on Consumer Electronics (ISCE)*, pages 1–2. IEEE, 2015.

[64] J. P. LEWIS. Fast templatematching templatematching. *Pattern Recognition*, 10(11) :120–123, 1995.

[65] L. Leyrit. *Reconnaissance d'objets en vision artificielle : application à la reconnaissance de piétons*. PhD thesis, 2010.

[66] J. Li and N. M. Allinson. A comprehensive review of current local features for computer vision. *Neurocomputing*, 71(10-12) :1771–1787, 2008.

[67] X. Li and X. Guo. A hog feature and svm based method for forward vehicle detection with single camera. In *2013 5th International Conference on Intel-*

ligent Human-Machine Systems and Cybernetics, volume 1, pages 263–266. IEEE, 2013.

[68] T. Lindeberg and J. Gårding. Shape-adapted smoothing in estimation of 3-d shape cues from affine deformations of local 2-d brightness structure. *Image and vision computing*, 15(6) :415–434, 1997.

[69] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee, 1999.

[70] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2) :91–110, 2004.

[71] G. Lowe. Sift-the scale invariant feature transform. *Int. J*, 2 :91–110, 2004.

[72] L. Macaire, V. Ultre, and J.-G. Postaire. Determination of compatibility coefficients for color edge detection by relaxation. In *Proceedings of 3rd IEEE International Conference on Image Processing*, volume 3, pages 1045–1048. IEEE, 1996.

[73] E. Maggio, F. Smeraldi, and A. Cavallaro. Combining colour and orientation for adaptive particle filter-based tracking. In *BMVC*, 2005.

[74] S. G. Mallat. A theory for multiresolution signal decomposition : the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7) :674–693, 1989.

[75] B. S. Manjunath and W.-Y. Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on pattern analysis and machine intelligence*, 18(8) :837–842, 1996.

[76] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada. Color and texture descriptors. *IEEE Transactions on circuits and systems for video technology*, 11(6) :703–715, 2001.

[77] D. Marr. *Vision : A computational investigation into the human representation and processing of visual information*. Henry Holt and Co., Inc., New York, NY, USA, 1982.

[78] K. Mikolajczyk, B. Leibe, and B. Schiele. Local features for object class recognition. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 2, pages 1792–1799. IEEE, 2005.

[79] K. Mikolajczyk and C. Schmid. Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1) :63–86, 2004.

[80] A. Milan, S. Roth, and K. Schindler. Continuous energy minimization for multitarget tracking. *IEEE transactions on pattern analysis and machine intelligence*, 36(1) :58–72, 2013.

[81] F. Mindru, T. Tuytelaars, L. Van Gool, and T. Moons. Moment invariants for recognition under changing viewpoint and illumination. *Computer Vision and Image Understanding*, 94(1-3) :3–27, 2004.

[82] H. P. Moravec. Obstacle avoidance and navigation in the real world by a seeing robot rover. Technical report, Stanford Univ Ca Dept of Computer Science, 1980.

[83] F. Nadia and H. Kamel. Personal identification based on texture analysis of arabic handwriting text. In *2006 2nd International Conference on Information & Communication Technologies*, volume 1, pages 1302–1307. IEEE, 2006.

[84] S. Nene, S. Nayar, and H. Murase. Columbia object image library (coil-100) department of computer science, columbia university ; new york, ny. Technical report, USA : 1996. Technical Report CUCS.

[85] S. Nicolas, Y. Kessentini, T. Paquet, and L. Heutte. Handwritten document segmentation using hidden markov random fields. In *Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, pages 212–216. IEEE, 2005.

[86] F. Nourbakhsh, P. B. Pati, and A. Ramakrishnan. Text localization and extraction from complex gray images. In *Computer Vision, Graphics and Image Processing*, pages 776–785. Springer, 2006.

[87] N. Ohta and A. Robertson. *Colorimetry : fundamentals and applications*. John Wiley & Sons, 2006.

[88] T. Ojala, M. Pietikainen, and T. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7) :971–987, 2002.

[89] S. E. Palmer. *Vision science : Photons to phenomenology*. MIT press, 1999.

[90] C. P. Papageorgiou, M. Oren, and T. Poggio. A general framework for object detection. In *Sixth International Conference on Computer Vision (IEEE Cat. No. 98CH36271)*, pages 555–562. IEEE, 1998.

[91] G. Pass and R. Zabih. Histogram refinement for content-based image retrieval. In *Proceedings Third IEEE Workshop on Applications of Computer Vision. WACV'96*, pages 96–102. IEEE, 1996.

[92] G. Pass, R. Zabih, and J. Miller. Comparing images using color coherence vectors. In *Proceedings of the fourth ACM international conference on Multimedia*, pages 65–73, 1997.

[93] Platon. *La République Livre VII*. IV av JC.

[94] H. Possegger, T. Mauthner, P. M. Roth, and H. Bischof. Occlusion geodesics for online multi-object tracking. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1306–1313, 2014.

[95] Y.-L. Qiao, Z.-M. Lu, C.-Y. Song, and S.-H. Sun. Document image segmentation using gabor wavelet and kernel-based methods. In *2006 1st International Symposium on Systems and Control in Aerospace and Astronautics*, pages 5–pp. IEEE, 2006.

[96] J. Quinlan. *Induction of decision trees*. Learn. 1, 1986.

[97] T. R. Reed and J. H. Dubuf. A review of recent texture segmentation and feature extraction techniques. *CVGIP : Image understanding*, 57(3) :359–372, 1993.

[98] Y. M. Ro, M. Kim, H. K. Kang, B. Manjunath, and J. Kim. Mpeg-7 homogeneous texture descriptor. *ETRI journal*, 23(2) :41–51, 2001.

[99] E. Rosten and T. Drummond. Machine learning for high-speed corner detection. In *European conference on computer vision*, pages 430–443. Springer, 2006.

[100] Y. Rui, A. C. She, and T. S. Huang. Modified fourier descriptors for shape representation-a practical approach. In *Proc of First International Workshop on Image Databases and Multi Media Search*, pages 22–23. Citeseer, 1996.

[101] C. L. Sabharwal and S. R. Subramanya. Indexing image databases using wavelet and discrete fourier transform. In *Proceedings of the 2001 ACM symposium on Applied computing*, pages 434–439, 2001.

[102] K. Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 :1582–1596, 01 2010.

[103] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE transactions on pattern analysis and machine intelligence*, 19(5) :530–535, 1997.

[104] R. Sève. Physique de la couleur : de l'apparence colorée à la technique colorimétrique. 1996.

[105] D. Slater and G. Healey. The illumination-invariant recognition of 3d objects using local color invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(2) :206–210, 1996.

[106] S. M. Smith and J. M. Brady. Susan—a new approach to low level image processing. *International journal of computer vision*, 23(1) :45–78, 1997.

[107] A. Stockman and L. T. Sharpe. The spectral sensitivities of the middle-and long-wavelength-sensitive cones derived from measurements in observers of known genotype. *Vision research*, 40(13) :1711–1737, 2000.

[108] S. Sural, G. Qian, and S. Pramanik. Segmentation and histogram generation using the hsv color space for image retrieval. In *Proceedings. International Conference on Image Processing*, volume 2, pages 589–592. IEEE, 2002.

[109] M. J. Swain and D. H. Ballard. Color indexing. *International journal of computer vision*, 7(1) :11–32, 1991.

[110] M. R. Teague. Image analysis via the general theory of moments. *JOSA*, 70(8) :920–930, 1980.

[111] G. R. Terrell and D. W. Scott. Variable kernel density estimation. *The Annals of Statistics*, pages 1236–1265, 1992.

[112] T. Toyoda and O. Hasegawa. Texture classification using extended higher order local autocorrelation. In *Proceedings of the 4th International Workshop on Texture Analysis and Synthesis*, pages 131–136, 2005.

[113] M. Trajković and M. Hedley. Fast corner detection. *Image and vision computing*, 16(2) :75–87, 1998.

[114] A. Trémau. Eléments de base de la colorimétrie. Technical report, Tech. rep., Laboratoire LIGIV-Université Jean Monnet, tremeau@ ligiv. org, 2004.

[115] M. Tuceryan. Moment-based texture segmentation. *Pattern recognition letters*, 15(7) :659–668, 1994.

[116] M. Tuceryan and A. K. Jain. Texture segmentation using voronoi polygons. *IEEE transactions on pattern analysis and machine intelligence*, 12(2) :211–216, 1990.

[117] M. Tuceryan and A. K. Jain. The handbook of pattern recognition and computer vision , chapter 2.1. *Texture Analysis. World Scientific Co.,*, pages 207–248, 1998.

[118] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings. 1991 IEEE computer society conference on computer vision and pattern recognition*, pages 586–587, 1991.

[119] T. Tuytelaars and L. Van Gool. Content-based image retrieval based on local affinely invariant regions. In *International Conference on Advances in Visual Information Systems*, pages 493–500. Springer, 1999.

[120] T. Tuytelaars and L. Van Gool. Matching widely separated views based on affine invariant regions. *International journal of computer vision*, 59(1) :61–85, 2004.

[121] K. Van De Sande, T. Gevers, and C. Snoek. Evaluating color descriptors for object and scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, 32(9) :1582–1596, 2009.

[122] J. Van de Weijer, T. Gevers, and A. D. Bagdanov. Boosting color saliency in image feature detection. *IEEE transactions on pattern analysis and machine intelligence*, 28(1) :150–156, 2005.

[123] J. Van De Weijer, T. Gevers, and A. Gijsenij. Edge-based color constancy. *IEEE Transactions on image processing*, 16(9) :2207–2214, 2007.

[124] J. Van De Weijer and C. Schmid. Coloring local feature extraction. In *European conference on computer vision*, pages 334–348. Springer, 2006.

[125] J. Van de Weijer and C. Schmid. Computer vision–eccv 2006 : 9th european conference on computer vision, graz, austria, may 7-13, 2006. proceedings, part ii. *by Aleö Leonardis, Horst Bischof, and Axel Pinz. Berlin, Heidelberg : Springer Berlin Heidelberg*, pages 334–348, 2006.

[126] L. Van Gool, T. Moons, and D. Ungureanu. Affine/photometric invariants for planar intensity patterns. In *European Conference on Computer Vision*, pages 642–651. Springer, 1996.

[127] V. Vapnik. A note one class of perceptrons. *Automation and remote control*, 1964.

[128] P. Viola, M. Jones, et al. Robust real-time object detection. *International journal of computer vision*, 4(34-47) :4, 2001.

[129] J. Von Kries. Influence of adaptation on the effects produced by luminous stimuli. *handbuch der Physiologie des Menschen*, 3 :109–282, 1905.

[130] L. Wang and G. Healey. Using zernike moments for the illumination and geometry invariant classification of multispectral texture. *IEEE Transactions on Image Processing*, 7(2) :196–203, 1998.

[131] W. Wright. A re-determination of the mixture curves of the spectrum. *Transactions of the Optical Society*, 31(4) :201, 1930.

[132] W. D. Wright. A re-determination of the trichromatic coefficients of the spectral colours. *Transactions of the Optical Society*, 30(4) :141, 1929.

[133] P. Wu, B. S. Manjunath, S. Newsam, and H. Shin. A texture descriptor for browsing and similarity retrieval. *Signal processing : Image communication*, 16(1-2) :33–43, 2000.

[134] X. Yan, X. Wu, I. A. Kakadiaris, and S. K. Shah. To track or to detect ? an ensemble framework for optimal selection. In *European Conference on Computer Vision*, pages 594–607. Springer, 2012.

[135] J. Yang, P. A. Vela, Z. Shi, and J. Teizer. Probabilistic multiple people tracking through complex situations. In *11th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2009.

[136] A. Yao, D. Uebersax, J. Gall, and L. Van Gool. Tracking people in broadcast sports. In *Joint Pattern Recognition Symposium*, pages 151–161. Springer, 2010.

[137] J. H. Yoon, M.-H. Yang, J. Lim, and K.-J. Yoon. Bayesian multi-object tracking using motion context from multiple objects. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 33–40. IEEE, 2015.

[138] S. Zambanini and M. Kampel. A local image descriptor robust to illumination changes. pages 11–21, 06 2013.

[139] D. Zhang and G. Lu. A comparative study of curvature scale space and fourier descriptors for shape-based image retrieval. *Journal of Visual Communication and Image Representation*, 14(1) :39–57, 2003.

[140] A. Znaidia, T. Zaharia, and F. Preteux. Une évaluation des descripteurs visuels mpeg-7 pour la recherche d'images par le contenu. *Hammamet, Tunisia, May*, 2009.

الملخص
التعرف على فئات الصور مهم للوصول إلى المعلومات المرئية حول طبيعة الكائنات وأنواع المشاهد. في هذه الأطروحة، نقترح نهجًا جديدًا لتصنيف الأشياء الملونة المصنوعة يدويا في الصور ذات الخلفية الموحدة باستخدام المعلومات القوية التي يوفرها اللون. يعتمد هذا النهج على مزيج من ثبات اللون "الحواف رمادية"، مكون الصبغة في عالم الالوان (صبغة، تشبع، قيمة) وكذلك فكرتي الخلايا والصناديق المستخدمة في واصف الرسوم البيانية للتدرجات الموجهة. يستفيد الواصف الموجه المقترح من ثبات الصبغة في مواجهة تغيرات شدة الضوء، تحول الضوء وتحول شدة الضوء، كما يعالج افتقارها إلى الثبات في مواجهة تغير لون الاضاءة باستخدام ثبات اللون "الحواف رمادية". بالإضافة إلى ذلك، أدى استخدام فكرتي الخلايا والصناديق في منهجية بناء الواصف المقترح إلى تعزيز قدرته على مواجهة التحولات الهندسية والضوئية وبالتالي زيادة معدل التعرف. للقيام بمهمة التصنيف قمنا بالاستعانة بتقنيتي شعاع الدعم الآلي وكي أقرب جار، وهما طريقتا تصنيف قويتان معروفتان بمرونتهما وقدرتهما على التعميم. تم اختبار الطريقة المقترحة على مكتبتي بيانات متاحتين للجمهور، وهما مكتبة كولومبيا ومكتبة أمستردام لصور الاشياء الملونة. أثبتت الاختبارات ليس فقط الأداء الاستثنائي للطريقة المقترحة في هذه الرسالة مقارنة بالطرق الحالية من حيث معدل التعرف، ولكن أيضًا سرعتها وقدرتها على تحسين استخدام مساحات الذاكرة.
الكلمات المفتاحية: الصور، الكائنات، اللون، الصبغة، شعاع الدعم الآلي، ...

Abstract :
Image category recognition is important to access visual information on the level of objects and scene types. In this thesis, we propose a new approach for color object recognition using the powerful information provided by the color. This approach is based on the combination of Gray-Edge color constancy, hue components in HSV (Hue, Saturation, Value) color space and cell and bin ideas used in the HOG (Histograms of Oriented Gradients) descriptors. The proposed oriented descriptor benefits of the invariance of hues against light intensity change, light intensity shift and light intensity change and shift, and solve its missing of invariance against light color change by using Gray-Edge color constancy. Moreover, the use of cells and bins in this proposed descriptor building strengthens its invariance to the geometric and photometric transformation and increases the recognition rate. The classifiers SVM (Support Vector Machine) and KNN (k-Nearest Neighbors) classifiers, which are two strong classification methods known for their flexibility and their power of generalization are used for the training and categorization steps. The proposed method is evaluated on two publicly available datasets including Columbia Object Image Library COIL-100 and The Amsterdam Library of Object Images ALOI. Tests have confirmed not only the exceptional performance of the proposed method compared to existing methods in terms of recognition rate, but also its rapidity and its optimization of using storage and memory spaces.
Key words— Object categorization, Color object recognition, HSV, Visual information, SVM, K-NN.

Résumé :
La reconnaissance des catégories des images est importante pour accéder aux informations visuelles des objets et des types de scènes. Dans cette thèse, nous proposons une nouvelle approche pour la catégorisation d'objet couleur en utilisant les informations puissantes fournies par la couleur. Cette approche est basée sur la combinaison de la constance des couleurs « Gray-Edge », la composante de teinte dans l'espace colorimétrique teinte, saturation et valeur ou HSV (en anglais ; Hue, Saturation, Value) et des idées de cellules et des bacs utilisés dans les descripteurs basés sur les histogrammes des gradients orientés ou HOG (en anglais ; Histograms of Oriented Gradients). Le descripteur orienté proposé profite de l'invariance des teintes contre le changement d'intensité lumineuse, le décalage d'intensité lumineuse et le changement/décalage d'intensité lumineuse, et résout son manque d'invariance contre le changement de couleur de la lumière en utilisant la constance de couleur Gray-Edge. De plus, l'utilisation de cellules et de bacs dans la méthodologie de construction de descripteur proposé a renforcé son invariance aux transformations géométriques et photométriques, et aussi augmente le taux de reconnaissance. Les classificateurs à machine de supports de vecteur ou SVM (en anglais pour Support Vector Machine) et classificateurs de plus proches voisins ou KNN (en anglais pour K-Nearest Neighbours), qui sont deux méthodes de forte classification connues pour leurs flexibilités et leurs pouvoirs de généralisation sont utilisés pour la classification. La méthode proposée est évaluée sur deux ensembles de données accessibles au public, dont la Bibliothèque d'images d'objets Columbia (en anglais pour Columbia Object Image Library coil-100) et la bibliothèque d'images d'objets d'Amsterdam ALOI (en anglais pour Amsterdam Library of Object Images). Les tests ont montré non seulement les performances exceptionnelles de la méthode proposée dans cette thèse par rapport aux méthodes existantes en termes de taux de reconnaissance, mais aussi sa rapidité et sa capacité à optimiser l'utilisation des espaces mémoire.
Mots clés— Catégorisation d'objet, Reconnaissance d'objet couleur, HSV, Informations visuelles, SVM, K-NN.