



**République Algérienne Démocratique et Populaire**  
**Ministère de l'Enseignement Supérieur et**  
**de la Recherche Scientifique**  
**Projet de Fin d'Études**



**Université Echahid Hamma Lakhdar d'El-Oued**

Faculté de Technologie

Mémoire de fin d'étude

En vue de l'obtention du diplôme de

**MASTER ACADEMIQUE**

**Domaine: technologie**

**Filière: génie électrique**

**Spécialité: Systèmes de communication**

**Thème**

**Amélioration la ré-identification des personnes  
par attributs**

**✚ Présenté par:**

**✚ Assem Elhacene CHELBI**

**✚ Adel CHAIB**

**Devant le jury composé de:**

**Dr. Fouaze MOUSSI**

Superviseur

**Dr. Said GHENDIR**

Président

**Dr. Nabil OURAR**

Examineur

**Année universitaire: 2023 /2024**

*Pour toute ma famille.*

*À tous ceux qui ont été les bougies qui ont illuminé notre chemin :*

*À notre chère famille, qui a toujours été une source de soutien et de force :*

*- À nos parents bien-aimés, sans votre amour et vos sacrifices, nous n'aurions jamais atteint ce jour. Nous vous remercions du fond du cœur et demandons à Dieu de vous protéger et de vous accorder santé et longue vie. Notre amour pour vous ne peut être décrit par des mots.*

*- À nos frères, vous faites partie de nous et de notre vie, et vous êtes toujours à nos côtés à chaque étape.*

*- À nos pères, qui ont été nos plus grands soutiens. Votre amour et votre dévouement nous poussent toujours en avant.*

*- À nos chères mères, les lanternes de sagesse et de conseil dans nos vies.*

*- À tous les membres de notre précieuse famille, nos tantes adorées et nos oncles respectés. Votre soutien est inestimable.*

*À mes chers amis, qui ont partagé avec moi cette aventure académique dans ses moindres détails.*

*À toutes les personnes qui souffrent en silence et qui ne peuvent pas s'exprimer, ce travail vous est dédié, j'espère qu'il sera pour vous une lueur d'espoir pour l'avenir.*

*À tous les petits oiseaux de notre famille, vous êtes la joie et la beauté de la vie.*

*Je vous remercie tous du fond du cœur, car c'est grâce à vous que cette réussite a été possible.*

*Avec tout mon amour et ma gratitude,*

*-chelbi assem elhacene et chaib adel.*

## Résumé

La reconnaissance des individus, ou Re-ID, est une tâche essentielle dans le domaine de la sécurité intelligente. Elle consiste à identifier les personnes sur des images capturées par différentes caméras de surveillance avec des champs de vision non superposés. Dans un environnement réel, la Re-ID représente un défi en raison des variations dans les conditions d'éclairage, les angles de vue, les postures et les ombres. Deux problèmes fondamentaux essentiels pour la Re-ID sont la représentation des caractéristiques et l'apprentissage des métriques. La représentation des caractéristiques doit être efficace et robuste contre les variations d'éclairage et d'angles de vue, tandis qu'il est nécessaire d'apprendre une métrique discriminante pour faire correspondre différentes images d'individus. Dans ce travail, nous utiliserons une représentation efficace des caractéristiques appelée Apparence Locale Maximale (LOMO) et le descripteur du histogramme des orientations de gradient (GOG), en plus d'une méthode d'apprentissage de sous-espace et d'apprentissage des métriques appelée Analyse Discriminante Quadratique Inter-Vues (XQDA) avec la distance de Mahalanobis pour améliorer les résultats de classification.

Les résultats expérimentaux obtenus (LOMO+XQDA=44.34%) et (GOG+XQDA=43.13%) ont montré l'efficacité des deux méthodes dans l'amélioration de la capacité de reconnaissance d'identité des personnes en utilisant la base de données VIPeR.

**Mots-clés:** Reconnaissance des individus (Re-ID), LOMO, GOG, XQDA, Mahalanobis.

## Abstract

Person recognition, or Re-ID, is a critical task in intelligent security applications. It involves identifying individuals in images captured by different non-overlapping surveillance cameras. In real-world scenarios, Re-ID poses challenges due to variations in lighting conditions, viewing angles, postures, and shadows. Two fundamental issues in Re-ID are feature representation and metric learning. Effective feature representation needs to be robust against lighting and viewpoint variations, while learning a discriminative metric is essential for matching different individual images. In this work, we utilize an effective feature representation called Local Maximum Occurrence (LOMO) and Gradient Orientation Histogram (GOG) descriptors, alongside a subspace learning and metric learning method known as Cross-View Quadratic Discriminant Analysis (XQDA) with Mahalanobis distance to enhance classification results. Experimental results (LOMO+XQDA=44.34%) and (GOG+XQDA=43.13%) demonstrate the efficacy of both methods in improving person identity recognition using the VIPeR database. Keywords: Person recognition (Re-ID), LOMO, GOG, XQDA, Mahalanobis.

Dédicace.....	I
Résumé.....	II
Liste des Figures.....	III
Liste des tableaux.....	III
Liste des abréviation.....	III
Introduction Générale.....	III
<b>Chapitre1 biométrique et ré_ identification de personne.....</b>	
I.1. Introduction .....	1
I.2. La biométrie.....	2
I.2.1. Reconnaissance faciale.....	2
I.2.2. Vérification vocale.....	2
I.2.3. Reconnaissance des empreintes digitales.....	3
I.2.4. Géométrie de la main.....	3
I.2.5. Reconnaissance de l'iris.....	4
I.2.6. Dynamique de frappe.....	4
I.2.7 Signature.....	5
I.2.8. Reconnaissance de la démarche.....	5
I.2.9. Analyse rétinienne.....	5
I.3. Etapes d'un système de Re-identification des personnes.....	6
I.3.1. Détection des personnes à ré-identification.....	6
I.3.2. Extraction des primitives.....	6
I.3.3. Appariement des primitives.....	6
I.3.4. Re-identification de la personne.....	6
I.4. Comparaison entre l'identification et la Re-ID.....	7
I.5. Différents scénarios de la Re-ID.....	8
I.6. Problématique du système de Re-ID.....	8
I.6.1. Conditions d'illumination .....	8
I.6.2. Occultations .....	8
I.6.3. Changement de l'angle de vue .....	9
I.6.4. Changement de pose .....	9
I.6.5. Détection des personnes .....	9
I.7. Applications de la Re-ID des personnes.....	9
I.8. Biométrie douce (Soft Biométrie).....	10
I.8.1. Caractéristiques de la Biométrie Douce.....	10

I.8.2. Avantage de la biométrie douce.....	11
I.8.3. Applications de la Biométrie Douce.....	11
I.8.4. Techniques et applications des biométries souples.....	11
I.8.5. Description sémantique.....	12
I.9. Ensembles de données de référence.....	12
I.10. Conclusion.....	13
<b>Chapitre2 Méthode.....</b>	<b>14</b>
II.1.Introduction.....	15
II.2.Travaux connexes.....	15
II.3.Descripteur LOMO.....	16
II. 4. Caractéristique LOMO.....	17
II.4.1. Gestion des variations d'éclairage.....	16
II.4.2. Gestion des changements de point de vue.....	17
II.5. Analyse discriminante quadratique inter_vues.....	19
II.5.1 Réexamen de l'approche Bayesian Face et KISSME.....	19
II.5.2.XQDA.....	20
II.6. Rôle de Retinex.....	21
II.7. Role de l'occurrence maximale locale.....	22
II.8. Descripteur GOG.....	22
II.9. Descripteur Gaussien Hiérarchique.....	24
II.9.1. Caractéristiques des pixels.....	24
II.9.2. Région Gaussiens.....	25
II.9.3. Mapping de l'espace tangent et demi-vectorisation.....	25
II.9.4. Gaussienne de région sur l'espace tangent.....	26
II.9.5. Fusion du descripteur de différentes espaces de couleurs.....	27
II.9.6. Normalisation de GOG.....	28
II.10. Analyse de Mahalanobis.....	28
II.11.Conclusion.....	28
<b>Chapitre3 Implémentations et Résultats.....</b>	<b>29</b>
III.1. Introduction.....	30
III.2. Base de données.....	30
III.3Approche proposée.....	30
III.4. Attributs.....	31
III.5. Détection des attributs SB.....	31
III.6. Détection d'attributs.....	32

---

III.7. Résultats sans biométrie douce (soft biometrics).....	32
III.8. Résultats avec biométrie douce (soft biometrics).....	33
III.8.1 Descripteur LOMO avec SB.....	33
III.8.2 Descripteur GOG avec SB.....	33
III.9. Comparaison des résultats entre LOMO et GOG.....	33
III.10. Conclusion.....	34
Conclusion Générale.....	35
Bibliographie.....	36

Liste des Figures

**Chapiter I**

**Fig.I.1:** Every prend des mesures pour lutter contre l'insécurité ..... 1

**Fig.I.2.1:** Modalité faciale ..... 2

**Fig.I.2.2:** Modalité vocale ..... 3

**Fig.I.2.3:** Modalité empreinte digitale ..... 3

**Fig.I.2.4:** Modalité de la géométrie de la main ..... 4

**Fig.I.2.5:** Modalité de l'iris ..... 4

**Fig.I.2.6:** Modalité des signataires ..... 5

**Fig.I.2.7:** Modalité de la démarche ..... 5

**Fig.I.2.8:** Modalité de la rétine ..... 6

**Fig.I.3:** Etapes d'un système générique de Re-ID dans un réseau de deux caméras..... 7

**Fig.I.4:** Scénarios de la Re-ID ..... 8

**Fig.I.5:**Exemple de facteurs de complexité de la Re-ID; (a) conditions d'illumination; (b) occultation; (c) angle de vue; (d) détection des personnes ..... 9

**Fig.I.6:** Le fichier informatif pour la personne disparue "Kristine Nicole Hamilton", parmi les autres, on peut clairement identifier les traits biométriques souples..... 11

**Chabiter2**

**Fig. II.1.** (a) Exemples de paires d'images de la base de données VIPeR [9]. (b) Images traitées dans (a) par Retinex. Les images dans la même colonne représentent la même personne ..... 18

**Fig.II.2.** Illustration de la méthode d'extraction des caractéristiques LOMO..... 19

**Fig. II.3.1.** Répartitions de  $\Omega I$  et  $\Omega E$  dans une dimension projetée..... 20

**Fig.II.6.** Importance de la distribution hiérarchique : (a) Régions ayant la même distribution (moyenne/covariance) des caractéristiques des pixels (chaque couleur indique le même vecteur de caractéristiques). (b) Patches locaux à l'intérieur des régions ayant une distribution différente des caractéristiques des pixels. (c) Les régions peuvent être distinguées par les distributions des distributions au niveau des patches..... 22

**Fig.II.6.1.** Importance de la moyenne: (a) Images originales. (b) Images montrant les valeurs moyennes RGB de patchs de pixels de  $10 \times 10$  de (a). (c) Images avec la moyenne retirée (chaque valeur RGB est mise à l'échelle dans la plage[0,255] pour la visualisation). Il est facile de déterminer les mêmes personnes à partir de(b),tandis que c'est difficile à partir de (c)..... 23

**Fig. II.8.** Descripteur GOG..... 24

**Chapitre III**

<b>Fig. III.1:</b> Quelques exemples de l'ensemble de données VIPeR.....	<b>30</b>
<b>Fig.III.2.</b> Schéma diagrammatique de l'approche proposée.....	<b>30</b>
<b>Fig. III.3:</b> Back pack, jeans, Carrying, Redshirt et Male sont cinq attributs SB utilisés dans ce travail.....	<b>31</b>
<b>Fig. III.4:</b> Un exemple de résultats de sélection de bandes pour la détection d'attributs dans VIPeR.....	<b>31</b>
<b>Fig.III.5:</b> Résultats soft biometrics 5ATTRIBU.....	<b>32</b>
<b>Fig. III.6:</b> Le différence entre les résultats de LOMO et GOG.....	<b>34</b>

**Liste des tableaux**

<b>Tableau.1:</b> Comparaison entre l'identification et la Re-ID.....	<b>4</b>
<b>Tableau.2:</b> Exemple de différents ensembles de données et leurs caractéristiques.....	<b>14</b>
<b>Tableau.3:</b> Soft biometrics Attributes.....	<b>32</b>
<b>Tableau.4:</b> Résultats de la Re-ID des personnes sur de données VIPeR(LOMO).	<b>33</b>
<b>Tableau 5:</b> Résultats de la Re-ID des personnes sur de données VIPeR(GOG)	<b>34</b>
<b>Tableau.6.</b> Performance de la Ré-Identification des Personnes avec LOMO+XQDA et GOG+XQDA à Différentes Dimensions.....	<b>34</b>

## Liste des abréviations

**Re-ID:** ré-identification.

**CMDA:** Criterion Multilinear Discriminant Analysis.

**PCA:** Principal component analysis.

**XQDA:** Cross-view Quadratic Discriminant Analysis.

**LOMO:** Local Maximal Occurrence Feature.

**GOG:** Gaussian Of Gaussian.

**SILTP:** Scale Invariant Local Ternary Pattern.

**KISSME:** KISS Metric Learning for Person Re-identification.

**LBP:** Local Binary Pattern.

**LDA:** Linear Discriminant Analysis.

**SDALF:** Symmetry-Driven Accumulation of Local Features.

**ID:** Identification.

**HSV:** Hue, Saturation, Valeur.

**CNIL:** Commission Nationale de l'Informatique et des Libertés.

**SB:** Soft Biometric.

**VIPeR:** View Invariant Pedestrian Recognition.

## Introduction Générale

Dans ce mémoire, nous abordons l'utilisation des techniques de vision par ordinateur pour améliorer le processus de ré-identification des personnes dans des contextes de surveillance et de sécurité. La ré-identification des personnes à partir de leurs caractéristiques est une méthode innovante visant à améliorer la précision de l'identification entre différentes caméras ou scènes. Dans ce contexte, cette méthode se distingue des approches traditionnelles en intégrant des caractéristiques physiques comme la couleur des vêtements, permettant ainsi une identification plus précise des individus.

L'étude repose sur l'utilisation d'une méthode biométrique douce employant les descripteurs LOMO et GOG, qui sont des outils efficaces pour extraire les caractéristiques locales et structurelles des images. Le descripteur LOMO se concentre sur l'analyse des motifs locaux en calculant la fréquence des valeurs maximales des distributions de couleurs et de textures dans différentes parties de l'image. Cette approche aide à capturer les variations subtiles des motifs locaux et à gérer des défis tels que les changements de lumière et de clarté. En revanche, le descripteur GOG repose sur la représentation de l'image à l'aide d'un modèle statistique basé sur des distributions gaussiennes multi-niveaux, permettant ainsi de capturer les relations complexes entre les pixels.

Les expériences sont réalisées sur l'ensemble de données VIPeR, une référence bien connue dans la recherche en vision par ordinateur utilisée pour évaluer l'efficacité des techniques de ré-identification. Les résultats de l'étude montrent que la combinaison des descripteurs LOMO et GOG avec l'algorithme XQDA améliore considérablement la capacité de discrimination entre les individus en réduisant les dimensions superflues et en conservant les caractéristiques les plus distinctives. Cette combinaison renforce la performance des systèmes de sécurité, de surveillance et de gestion des accès, contribuant ainsi à atteindre des niveaux élevés de précision et d'efficacité dans la ré-identification des personnes.

D'un point de vue pratique, ces résultats peuvent être appliqués dans divers domaines tels que la sécurité et la surveillance dans les aéroports, les bâtiments gouvernementaux et les lieux publics. De plus, ces techniques peuvent être utilisées dans des applications commerciales pour améliorer l'expérience des clients en suivant leurs mouvements et leurs préférences.

La mémoire est organisée comme suit :

Dans le **premier chapitre**, intitulé "Identification biométrique et ré-identification des personnes", nous présenterons un aperçu complet des concepts de l'identification biométrique et de leur importance dans la ré-identification des individus. Nous passerons en revue les principales technologies biométriques et analyserons les sujets liés à la ré-identification des individus.

Dans le **deuxième chapitre**, intitulé "Méthodes et techniques", nous nous concentrerons sur la présentation d'une variété de méthodes et de techniques utilisées dans le domaine de l'identification biométrique et de la ré-identification des individus. Nous fournirons une analyse approfondie des méthodes utilisées et de leurs applications potentielles.

Quant au **troisième chapitre**, intitulé "Mise en œuvre et résultats", nous approfondirons l'étude de la mise en œuvre des méthodes et des techniques discutées dans les chapitres précédents. Nous

présenterons en détail la mise en œuvre de ces méthodes et les résultats potentiels pouvant être obtenus grâce à leur utilisation.

# **CHAPITRE 1**

biométrique et Re-ID de personne

## I.1. Introduction

La surveillance vidéo connaît actuellement une expansion considérable, tant sur le plan technologique qu'économique. Elle est devenue un élément essentiel des politiques de sécurité gouvernementales, répondant au besoin de chaque citoyen en matière de sécurité face à la montée de la délinquance et de la criminalité. Les attentats de septembre 2001 aux États-Unis et de 2005 à Londres ont contribué à une augmentation significative du nombre de caméras de surveillance. À ce jour, Londres détient le record du nombre de caméras de vidéosurveillance par habitant. Prenons l'exemple des cambriolages de logements en France, qui a connu une augmentation de 7% par rapport à 2012 [1]. En réponse, le gouvernement français a mis en place les moyens nécessaires pour lutter contre l'insécurité, la vidéosurveillance en étant l'un d'entre eux. En effet, selon le ministère de l'Intérieur, le taux d'élucidation des crimes et délits commis sur la voie publique peut être multiplié par deux dans les villes sous surveillance vidéo [2]. La vidéosurveillance est désormais indispensable pour surveiller à la fois les espaces publics et privés.

Dans ce cadre, des réseaux de caméras sont déployés dans les rues, les centres commerciaux, les transports en commun, les bureaux, les aéroports, les immeubles résidentiels, etc. En 2012, la Commission Nationale de l'Informatique et des Libertés (CNIL) a signalé l'installation d'environ 935 000 caméras en France [03]. Dans la ville d'Évry, qui est souvent confrontée à des incidents de délinquance fréquemment relayés par les médias, 65 caméras surveillent la ville dans l'espoir de dissuader définitivement les criminels. Les images capturées sont renvoyées à un centre de contrôle et affichées sur des écrans (fig. I.1).



**Fig. I.1: Every prend des mesures pour lutter contre l'insécurité [24].**

Un système de vidéo surveillance consiste principalement à surveiller simultanément tous ces écrans. Cependant, l'augmentation du nombre de caméras installées rend la tâche d'exploitation manuelle des données produites par ces caméras difficile. Pour faciliter l'exploration de ces données par les agents de sécurité, il est donc nécessaire de rendre la vidéosurveillance plus intelligente en automatisant certaines de ses fonctions. Parmi celles-ci, on peut citer la détection des objets, la détection des personnes, la reconnaissance d'événements et d'actions humaines, le suivi des personnes, etc. Une autre application consiste à reconnaître les personnes qui quittent le champ de vision d'une caméra et réapparaissent dans celui d'une autre. Le système de vidéo surveillance doit alors être capable de ré-identifier la personne et de poursuivre le suivi.

## I.2. La biométrie

### *Qu'est-ce que la Biométrie Douce ?*

La biométrie se réfère à l'identification d'une personne à travers une ou plusieurs caractéristiques physiologiques (empreintes digitales, visage, iris, contour de la main, etc.) ou comportementales (signature, démarche, etc.). Étymologiquement, la biométrie humaine est associée à l'anthropologie physique [7]. Une autre définition de la biométrie est formulée par Roethenbaugh [6] : « La biométrie s'applique à des particularités ou des caractéristiques humaines uniques et mesurables, permettant la reconnaissance ou la vérification automatique de l'identité ».

Les systèmes biométriques sont de plus en plus utilisés depuis quelques années. L'apparition de l'ordinateur et sa capacité à traiter et à stocker les données ont permis la création des systèmes biométriques informatisés. Il existe plusieurs caractéristiques physiques uniques pour un individu, ce qui explique la diversité des systèmes appliquant la biométrie, selon que l'on prend en compte :

### I.2.1. Reconnaissance faciale

La reconnaissance faciale est généralement considérée comme le principal moyen par lequel les gens se reconnaissent mutuellement. Les approches les plus populaires de la reconnaissance faciale reposent soit sur [8] :

- . La localisation et la forme des caractéristiques faciales, telles que les yeux, les sourcils, le nez, les lèvres, le menton, et leurs relations spatiales.
- . L'analyse globale de l'image du visage qui représente un visage comme une combinaison pondérée de plusieurs visages canoniques.
- . En pratique, un système fiable de reconnaissance faciale devrait automatiquement :
  - . Détecter si un visage est présent dans l'image acquise.
  - . Localiser le visage s'il y en a un.
  - . Reconnaître le visage indépendamment de sa pose générale et dans différentes conditions ambiantes.

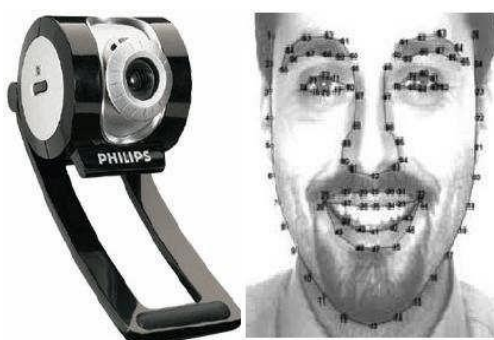


Fig. I.2.1: Modalité faciale [25].

### I.2.2. Vérification vocale

La voix est une combinaison de caractéristiques biométriques physiques et comportementales. Le processus d'authentification vocale est basé sur l'extraction et la modélisation de caractéristiques spécifiques du discours [10]. Ces caractéristiques physiques de la voix d'un individu sont basées sur la forme et la taille des voies vocales, de la bouche, des cavités nasales et des lèvres utilisées dans la symémoire du son. Les caractéristiques physiques de la voix humaine

sont invariantes pour un individu, mais l'aspect comportemental de la parole change au fil du temps en raison de facteurs tels que l'âge, les conditions médicales (comme le rhume), l'état émotionnel, etc. Le principal inconvénient du système de reconnaissance basé sur la voix est que les caractéristiques de la parole sont sensibles à de nombreux facteurs tels que le bruit de fond [8].



**Fig. I.2.2:** Modalité vocale [25].

### I.2.3. Reconnaissance des empreintes digitales

Les êtres humains utilisent les empreintes digitales pour l'identification personnelle depuis de nombreuses décennies. Les empreintes digitales sont l'une des technologies biométriques les plus matures utilisées dans les divisions de la police scientifique du monde entier pour les enquêtes criminelles [11].

Une empreinte digitale est le motif de crêtes et de vallées à la surface d'une extrémité du doigt, dont la formation est déterminée au cours des sept premiers mois du développement fœtal. Il a été déterminé de manière empirique que les empreintes digitales des jumeaux identiques sont différentes, tout comme les empreintes sur chaque doigt d'une même personne [8]. L'un des principaux inconvénients des systèmes d'identification par empreintes digitales est que de petites blessures et brûlures affectent fortement l'empreinte digitale [10].



**Fig. I.2.3:** Modalité empreinte digitale [25].

### I.2.4. Géométrie de la main

Les systèmes de reconnaissance basés sur la géométrie de la main reposent sur plusieurs mesures prises à partir de la main humaine, notamment sa forme, la taille de la paume, ainsi que les longueurs et largeurs des doigts [9]. La technique est très simple, relativement facile à utiliser et peu coûteuse. Les facteurs environnementaux tels que le temps sec ou les anomalies individuelles telles que la peau sèche n'affectent pas la précision de l'authentification des systèmes basés sur la géométrie de la main.

Cependant, la géométrie de la main n'est pas réputée pour être très distinctive, et les systèmes de reconnaissance basés sur la géométrie de la main ne peuvent pas être étendus pour des systèmes nécessitant l'identification d'un individu parmi une grande population [8].



**Fig. I.2.4:** Modalité de la géométrie de la main [25].

### I.2.5. Reconnaissance de l'iris

L'iris est la région annulaire de l'œil délimitée par la pupille et la sclérotique (le blanc de l'œil) de chaque côté. La texture complexe de l'iris porte des informations très distinctives utiles pour la reconnaissance personnelle. Chaque iris est distinctif, et même les iris de jumeaux identiques sont différents [9].

Les systèmes basés sur l'iris ont les taux de fausses correspondances les plus bas parmi toutes les méthodes biométriques actuellement disponibles, et c'est la technique la moins intrusive parmi les biométries basées sur l'œil. C'est l'un des rares systèmes biométriques, en plus des empreintes digitales, qui fonctionne bien en mode "identification" (comparaison un à plusieurs) [12].



**Fig.I.2.5:** Modalité de l'iris [25].

### I.2.6. Dynamique de frappe

La dynamique de frappe est une autre technique précoce dans laquelle beaucoup de temps et d'efforts ont été investis, y compris par certaines grandes entreprises de technologie de l'information [13].

La dynamique de frappe, ou analyse, est également appelée rythme de frappe. Il s'agit d'une méthode automatisée d'analyse de la manière dont un utilisateur tape sur un terminal ou un clavier, examinant des dynamiques telles que la vitesse, la pression, le temps total nécessaire pour taper des mots particuliers et le temps écoulé entre la frappe de certaines touches. Plus précisément, l'analyse de la frappe mesure deux variables distinctes: le " temps de résidence ", qui est la durée pendant laquelle une personne maintient une touche particulière enfoncée, et le " Temps de vol ", qui est le temps écoulé entre les touches.

Cette technique fonctionne en surveillant les entrées du clavier des milliers de fois par seconde dans le but d'identifier l'utilisateur par ses habitudes de rythme de frappe [12].

### I.2.7 Signature

La signature personnelle a été acceptée dans les transactions gouvernementales, légales et commerciales comme une méthode d'authentification. En raison des assistants personnels numériques (PDAs) et des tablettes PC, la signature en ligne peut émerger comme le choix biométrique privilégié sur ces appareils.

La signature est une biométrie comportementale qui évolue au fil du temps et est influencée par les conditions physiques et émotionnelles des Fig (I.2.6)



Fig. I.2.6: Modalité des signataires [8].

### I.2.8. Reconnaissance de la démarche

La démarche est la manière dont une personne marche, et c'est l'une des rares caractéristiques biométriques qui peuvent être utilisées pour reconnaître les gens à distance. La plupart des algorithmes de reconnaissance de la démarche tentent d'extraire la silhouette humaine afin de dériver les attributs spatio-temporels d'un individu en mouvement. Certains algorithmes utilisent le flux optique associé à un ensemble de points en mouvement dynamiquement extraits sur le corps humain pour décrire la démarche d'un individu [11]. Cependant, la démarche d'un individu est affectée par plusieurs facteurs, y compris le choix des chaussures, la nature des vêtements, des affections aux jambes, la surface de marche, etc...



Fig. I.2.7: Modalité de la démarche [25].

### I.2.9. Analyse rétinienne

Des recherches menées dans les années 1930 suggéraient que les motifs des vaisseaux sanguins à l'arrière de l'œil humain étaient uniques à chaque individu, faisant de l'analyse rétinienne l'une des plus anciennes biométries connues [12].

La rétine est une fine couche de cellules à l'arrière du globe oculaire des vertébrés. C'est la partie de l'œil qui convertit la lumière en signaux nerveux.

Le principe de la biométrie rétinienne consiste à capturer et à analyser les motifs des vaisseaux sanguins sur le nerf fin à l'arrière du globe oculaire qui traite la lumière entrant par la pupille. Ces vaisseaux sanguins présentent un motif unique, d'un œil à l'autre et d'une personne à l'autre. Les motifs rétiniens sont des traits hautement distinctifs. Chaque œil a son propre motif de vaisseaux sanguins totalement unique; même les yeux de jumeaux identiques sont distincts [12]. Bien que

chaque motif reste normalement stable tout au long de la vie d'une personne, il peut être affecté par des maladies telles que le glaucome, le diabète, l'hypertension et le syndrome d'immunodéficience acquise (SIDA).

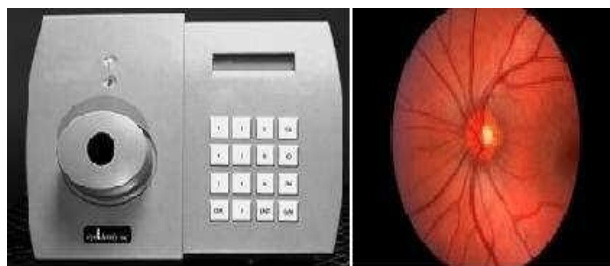


Fig. I.2.8: Modalité de la rétine [25].

### I.3. Etapes d'un système de Re-identification des personnes

La Re-ID des personnes au sein d'un réseau composé de deux caméras aux champs de vision disjoints, avec un intervalle de temps court entre les enregistrements des deux caméras (environ quelques minutes). Un système générique de Re-ID se compose de quatre étapes principales figure (I.3): la détection des personnes à ré-identifier, l'extraction des primitives, l'appariement des primitives et la Re-ID des personnes. Dans la suite, nous détaillerons le principe de chaque étape.

#### I.3.1. Détection des personnes à ré-identification

Cette étape dépend largement de la complexité de la scène. Dans une scène simple, où généralement une seule personne traverse, la détection peut être réalisée par une méthode de faible complexité telle que la suppression du fond ou la détection du mouvement. Lorsque la scène devient complexe, avec plusieurs personnes traversant, cette étape devient plus difficile et peut nécessiter l'utilisation d'une méthode de suivi des personnes. Pour cette mémoire, nous ne considérons pas ce cas complexe, mais plutôt nous utilisons des images générées par un algorithme de détection des personnes. Dans des scénarios complexes, cet algorithme n'est pas parfait et peut conduire à des détections de personnes incorrectes. En effet, les images peuvent par exemple contenir plusieurs personnes, seulement une partie du corps, etc.

#### I.3.2. Extraction des primitives

Cette étape consiste à décrire la personne par un ensemble de primitives approprié au type de la scène. Dans la littérature, ces primitives sont extraites des images et décrivent pour la plupart l'apparence de la personne par une description de couleur, de texture et/ou de forme.

#### I.3.3. Appariement des primitives

Une fois les primitives extraites des données enregistrées, elles sont appariées en définissant un score de correspondance. L'appariement des primitives peut être fondé sur une méthode supervisée ou non-supervisée.

#### I.3.4. Re-identification de la personne

La Re-ID de personne a été largement étudiée en tant que problème spécifique de recherche d'une personne à travers des caméras non chevauchantes [14, 15]. Avec une personne d'intérêt en tant que requête, l'objectif de la Re-ID est de déterminer si cette personne est apparue ailleurs à un

moment distinct capturé par une autre caméra [16]. La personne de la requête peut être représentée par une image [17, 18], une séquence vidéo [19, 20], voire une description textuelle [21, 22]. En raison de la demande urgente en matière de sécurité publique et de l'augmentation du nombre de caméras de surveillance sur les campus universitaires, les parcs à thème, les rues, etc., la Re-ID de personne est impérative dans la conception des systèmes de surveillance vidéo intelligents. En raison de son impact en recherche et de son importance pratique, la Re-ID est une communauté vision en rapide croissance [23]. La Re-ID de personne est basée sur la tâche de détection des piétons. Tout d'abord, un ensemble de galerie de piétons est formé en collectant les images de piétons recadrés ou la signature d'image de piéton extraite de chaque scène de caméra dans le réseau. Ensuite, la similarité ou la distance est mesurée par rapport à l'image de la requête. Enfin, l'image correspondante la mieux adaptée selon la similarité mesurée est considérée [24].

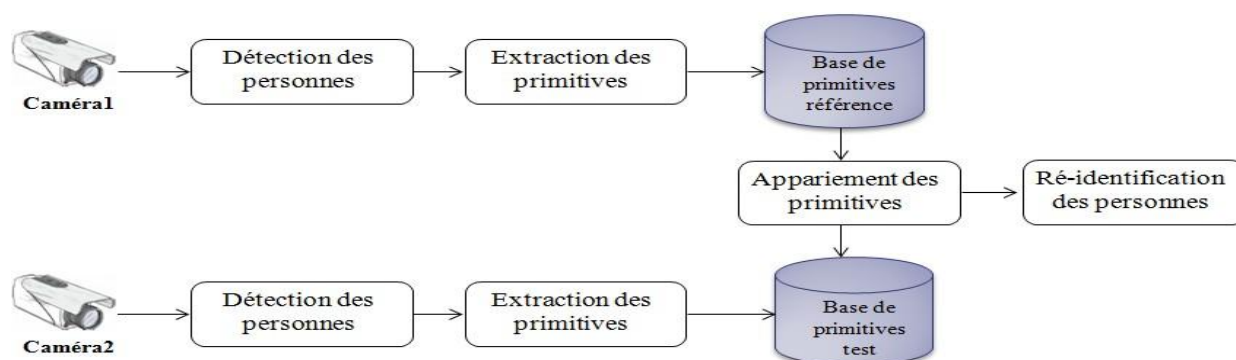


Fig. I.3: Etapes d'un système générique de Re-ID dans un réseau de deux caméras [24].

#### I.4. Comparaison entre l'identification et la Re-ID

Comme mentionné précédemment, la Re-ID est similaire à une tâche biométrique, car elle vise à associer une identité de référence à une identité test inconnue. La distinction principale entre l'identification et la Re-ID réside dans le fait qu'un système d'identification inclut une phase d'enrôlement où l'identité de la personne est enregistrée. En revanche dans la Re-ID, l'objectif est de déterminer si une personne traversant le champ de vision d'une caméra réapparaît dans un autre contexte.

Concernant la difficulté des deux systèmes, le niveau de difficulté de la Re-ID est différent de celui de la tâche d'identification biométrique classique. La table (I.1) montre une comparaison entre l'identification et la Re-ID selon trois critères: nombre d'échantillons référence disponibles par personne, type des conditions d'acquisition et temps entre l'acquisition des données de test et de référence. Les deux premiers critères rendent la tâche d'identification moins complexe alors que le troisième rend la tâche de Re-ID moins complexe.

Table(I.1): Comparaison entre l'identification et la Re-ID [24].

Critère de comparaison	Identification	Re-ID
Echantillons référence par Personne	Plusieurs	Un seul
Conditions d'acquisition	Conditions contrôlées (éclairage, fond, etc.)	Conditions aléatoires: peuvent être simples mais sont généralement Complexes
Temps entre l'acquisition des données de test et de référence	Beaucoup de temps	Généralement dans le Même jour

## I.5. Différents scénarios de la Re-ID

Selon les conditions d'acquisition des données, la Re-ID peut correspondre à plusieurs scénarios de degrés de complexité différents. On en cite deux: simple et complexe (fig. I.3).

Dans le scénario simple, des contraintes sont fixées sur le passage de la personne dans le champ de vue de la caméra. De plus, les conditions du milieu d'enregistrement sont contrôlées. En effet, la personne doit marcher seule dans une direction bien définie sans porter aucun objet. La base de données est composée d'un nombre réduit de personnes filmées, à deux instants différents, par une seule caméra fixe dans des conditions d'éclairage stables (la même caméra est utilisée en référence et en test).

Dans le scénario complexe, on n'a aucune contrainte sur le passage de la personne dans le champ de vue de la caméra. De plus, les conditions du milieu d'enregistrement ne sont pas contrôlées. En effet, plusieurs personnes peuvent passer simultanément en portant ou non des objets de petites dimensions (sac, téléphone, etc.) ou de grandes dimensions (poussette, matériel acheté, etc.). En outre, la base de données est composée d'un nombre de personnes largement plus important que dans le scénario simple ; elles sont filmées par deux caméras installées dans deux endroits différents et dans des conditions d'éclairage non contrôlées.



Fig. I.4: Scénarios de la Re-ID [24].

## I.6. Problématique du système de Re-ID

La correspondance des signatures des personnes dans un réseau de caméras à champs de vues disjoints est la tâche la plus difficile pour tous les systèmes d'analyse vidéo. En effet, un système de Re-ID doit faire face à la variation de l'apparence causée par différents facteurs tels que la variation des conditions d'illumination, l'occultation, la variation des angles de vue, la variabilité du fond, la pose, la qualité de détection des personnes, ... etc. voir la figure (I.4).

### I.6.1. Conditions d'illumination

La variation des conditions d'illumination est un défi critique d'un système de vidéo surveillance. Cette variation peut être due à la diversité des paramètres optiques des caméras ou des conditions d'éclairage du milieu d'enregistrement. Ainsi, les conditions d'illumination changent d'une façon significative l'apparence d'une personne.

### I.6.2. Occultations

Le passage des personnes dans le champ de vue de la caméra sans aucune contrainte peut produire des occultations. Ces occultations peuvent être causées par des objets, par d'autres personnes ou par des structures de l'environnement. En cas d'occultation, si certaines

caractéristiques importantes de la personne sont invisibles, sa Re-ID devient difficile et peut échouer.

### I.6.3. Changement de l'angle de vue

Selon l'angle de vue à partir duquel la personne est observée, des parties du corps humain peuvent apparaître différemment. En outre, des parties observées dans une caméra peuvent être cachées dans une autre.

### I.6.4. Changement de pose

Le changement de pose causé par la variation des angles de vue et par l'articulation du corps humain conduit à des apparences significativement différentes de la même personne vue par différentes caméras.

### I.6.5. Détection des personnes

Bien que cette mémoire ne se concentre pas sur la détection des personnes avant de les ré-identifier, il est important de noter que cette tâche demeure un défi majeur pour un système de Re-ID. La détection des personnes peut être réalisée à l'aide d'un algorithme statique [4] ou d'un algorithme de suivi des personnes [5]. Une détection insuffisante peut résulter du passage des personnes au sein d'une foule, de perturbations de l'arrière-plan, etc. De plus, elle est influencée par d'autres facteurs tels que les conditions d'éclairage, les poses, les angles de vue, etc. La Re-ID d'une personne devient complexe si certaines de ses parties sont détectées dans une caméra et non dans une autre.



**Fig.I.5:** Exemple de facteurs de complexité de la Re-ID; (a) conditions d'illumination; (b) occultation; (c) angle de vue; (d) détection des personnes [24].

## I.7. Applications de la Re-ID des personnes

Les stratégies de Re-ID des personnes ont un potentiel énorme dans de nombreux domaines différents, de la défense et la surveillance au commerce de détail et aux soins de santé.

- **Cross-camera person tracking:** Computer vision requires the ability to monitor people through several cameras, conduct crowd movement analysis, and recognize movements in order to comprehend a scene [24]. Typically, cross-camera tracking consists of two steps, object tracking in single camera and data association between multiple cameras. Consequently, besides the difficulty in single camera tracking, the main challenge of cross-camera tracking is to associate the trajectories of human movement in different views [47].

• **Suivi par détection:** La Re-ID individuelle peut être utile même dans une configuration de suivi à une seule caméra. Le suivi de plusieurs personnes est une tâche difficile, notamment dans des scénarios complexes et encombrés avec des occultations fréquentes et des interactions individuelles. L'idée principale est de détecter les personnes, d'estimer leurs schémas de mouvement et d'associer les détections dans différents cadres. Cette étape de liaison, appelée association de données, est en fait une forme de Re-ID [24]. Cela découle directement du développement de puissantes méthodes discriminatives en apprentissage automatique et de leur application à la détection avec un entraînement hors ligne. Beaucoup de ces algorithmes peuvent être adaptés pour un entraînement en ligne, où chaque détection réussie fournit plus d'informations sur la cible [48].

• **Récupération de personne:** Dans ce cas, la Re-ID est liée à une tâche de reconnaissance. Une requête spécifique avec un individu cible est donnée, et toutes les instances correspondantes dans une large base de données sont recherchées. Pour la récupération d'image, la tâche de Re-ID est utilisée, et elle retourne généralement des listes classées, des objets similaires et ainsi de suite [24].

• **Interaction homme-machine:** Au cours des dernières décennies, les scientifiques se sont davantage concentrés sur le développement de technologies visant à améliorer les interactions entre les humains et les machines en intégrant de nouvelles procédures innovantes avec les ordinateurs. Cela a entraîné une percée significative dans la vision par ordinateur et a permis aux ordinateurs de fonctionner ou de contrôler la machine en surveillant les gestes et les expressions des humains [49].

• **Analyse à long terme du comportement et des activités humaines:** L'analyse des habitudes de consommation des clients, par exemple, en les observant toucher, interviewer et essayer des articles dans les magasins sous diverses caméras de surveillance. Un autre exemple est la recherche sur les soins de santé gériatriques, qui examine les actions à long terme des personnes âgées pour aider les médecins à poser des diagnostics plus précis [24].

## I.8. Biométrie douce (Soft Biométrie)

### *Qu'est-ce que la Biométrie Douce?*

La biométrie douce se réfère à l'utilisation de caractéristiques biométriques qui, contrairement aux caractéristiques biométriques dures (comme les empreintes digitales, l'iris ou la reconnaissance faciale), ne sont pas nécessairement uniques à chaque individu, mais peuvent fournir une identification fiable lorsqu'elles sont combinées. Ces caractéristiques incluent des traits physiques, comportementaux et contextuels. La biométrie douce est particulièrement utile dans les systèmes de surveillance où la collecte de données biométriques dures est impraticable ou invasive.

#### I.8.1. Caractéristiques de la Biométrie Douce

a) Caractéristiques Physiques:

- **Apparence générale:** Couleur et style des vêtements, taille et forme du corps.
- **Accessoires:** Lunettes, chapeaux, sacs, bijoux, etc.

b) Caractéristiques Comportementales:

- **Démarche:** La façon dont une personne marche, incluant la vitesse, la longueur du pas, et la posture.

- **Gestes et mouvements:** Mouvements typiques des bras, des mains, ou du corps entier.
- c) Caractéristiques Contextuelles:
  - **Localisation:** Endroits fréquentés par une personne.
  - **Interactions sociales:** Interactions avec d'autres personnes ou objets.

### I.8.2. Avantage de la biométrie douce

- **Non-Intrusivité:** La biométrie douce peut être capturée passivement par des caméras de surveillance sans nécessiter la coopération ou même la connaissance de l'individu.
- **Adaptabilité:** Elle fonctionne bien dans des environnements non contrôlés où les conditions d'éclairage, les angles de vue et les arrière-plans peuvent varier de manière significative.
- **Complémentarité:** Lorsqu'elle est combinée avec des biométries dures, elle peut renforcer la précision globale des systèmes d'identification et de surveillance.

### I.8.3. Applications de la Biométrie Douce

- **Surveillance et Sécurité:** Suivi des individus à travers des réseaux de caméras dans les lieux publics, les centres commerciaux, les aéroports, etc.
- **Analyse Comportementale:** Étude des comportements des individus dans des environnements spécifiques pour des applications telles que le marketing, l'urbanisme, et la gestion des foules.
- **Recherche de Personnes Disparues:** Utilisation de descriptions générales et de comportements pour identifier et localiser des personnes disparues dans des vidéos de surveillance.

### I.8.4. Techniques et applications des biométries souples

Nous utilisons déjà des biométries souples tous les jours sans même le savoir. Un exemple est illustré dans la figure (I.6) où parmi les informations "biométriques dures" (les deux photos), nous trouvons une description détaillée des caractéristiques physiques que nous avons précédemment incluses dans la liste des biométries souples (poids, taille, tatouages, etc...). Chacune des caractéristiques biométriques souples présentées précédemment peut facilement trouver des applications principalement dans les trois domaines suivants.



**Fig.I.6:** Le fichier informatif pour la personne disparue "Kristine Nicole Hamilton", parmi les autres, on peut clairement identifier les traits biométriques souples.

### I.8.5. Description sémantique

Un autre aspect important des biométries souples est leur caractéristique conforme à l'humain, ce qui les rend conformes aux labels de description naturelle humaine. Cela signifie qu'un ordinateur qui extrait des identifiants biométriques souples peut communiquer directement ces valeurs à un opérateur humain sans aucune forme de technique de traitement, et vice versa. Grâce à cette propriété, une meilleure interaction homme-machine est possible car la biométrie souple permet l'extraction d'informations sémantiques à partir de l'image/vidéo

### I.9. Ensembles de données de référence

Au cours des dernières années, plusieurs ensembles de données de référence pour la Re-ID des personnes ont été publiés. Plusieurs facteurs doivent être pris en compte lors de la création d'un ensemble de données de référence [46], tels que:

- Le nombre de piétons.
- Le nombre d'images par personne dans une vue de caméra.
- Les variations de résolutions, d'éclairages, de poses, d'occultations et d'arrière-plans dans la même vue de caméra et entre différentes vues de caméras.
- Le nombre de vues de caméras.

Certains ensembles de données publics qui ont été testés pour les algorithmes de Re-ID de personnes et leurs caractéristiques sont illustrés dans le tableau 2.

**Tableau 2: Exemple de différentes ensembles de données et leurs caractéristiques.**

Datasets	Release time	# Identifiés	# Image	# Cameras	Taille de la coupe
<b>VIPeR</b>	<b>2007</b>	<b>632</b>	<b>1264</b>	<b>2</b>	<b>128*48</b>
GRID	2009	1025	1275	8	Varier
PRID2011	2011	934	24541	2	128*64
CUHK01	2012	971	3884	2	160*60
CUHK02	2013	1816	7264	10	160*60
CUHK03	2014	1467	13164	10	Varier
iLIDS-VID	2014	300	42495	2	Varier
Market1501	2015	1501	32217	6	128*64
MARS	2016	12611	1191003	6	256*128
AIRPORT	2017	9651	39902	6	128*64
LPW	2018	2,731	592,438	3,4,4	-
ThermalWo-Rld	2018	516	15,118	20	-

## **I.10. Conclusion**

Dans ce chapitre, nous avons présenté les technologies utilisées dans les systèmes biométriques pour l'identification des personnes. Cette étude nous a permis de constater que la Re-ID de personne est un nouveau domaine qui a été développé pour assurer la sécurité du public. Nous pouvons suivre le chemin d'une personne et nous assurer qu'aucune activité illégale ou inappropriée n'est entreprise en utilisant des modèles de Re-ID et de suivi. La Re-ID est un problème difficile car les humains sont souvent capturés par de nombreuses caméras de basse résolution, dans des conditions d'occultation, dans des conditions d'éclairage médiocres et dans diverses positions. Dans ce cas, une modélisation robuste de l'apparence complète du corps d'une personne est nécessaire, surtout lorsque d'autres indicateurs biométriques traditionnels ne sont pas disponibles ou impossibles à capturer en raison de limitations de résolution du capteur ou de fréquence d'images. Les chercheurs tentent d'améliorer les performances du système de Re-ID de personne en développant une représentation robuste des caractéristiques et un apprentissage de la métrique discriminante.

# **CHAPITRE 2**

## Méthode

## II.1. Introduction

Les caractéristiques LOMO et GOG sont des outils efficaces pour améliorer la précision des systèmes de Re-ID des personnes dans le domaine de la vision par ordinateur. Les caractéristiques LOMO reposent sur l'extraction d'informations locales à partir des images en calculant la fréquence des valeurs maximales des distributions de couleurs et de textures dans des parties de l'image, ce qui permet de capturer les variations subtiles des motifs locaux et de gérer des défis tels que les changements de lumière et de clarté. En revanche, les caractéristiques GOG reposent sur la représentation de l'image à l'aide d'un modèle statistique basé sur des distributions gaussiennes multi-niveaux pour extraire des informations structurelles et texturales, permettant de capturer les relations complexes entre les pixels. Lorsqu'elles sont combinées avec l'algorithme XQDA, ces caractéristiques améliorent considérablement la capacité de discrimination entre les individus en réduisant les dimensions et en conservant les caractéristiques les plus distinctives. Ces outils sont puissants et efficaces dans les applications de sécurité, de surveillance et de gestion des accès, contribuant à atteindre une grande précision dans la ré-identification des personnes.

## II.2. Travaux connexes

Pour représenter les caractéristiques, plusieurs approches efficaces ont été proposées. Shengcai Liao [29] a découpé l'image horizontalement, extrait la caractéristique locale en utilisant un bloc coulissant, et comparé les caractéristiques locales une par une. Liang Zheng [30] utilise l'algorithme de détection des parties du corps humain (mains, jambes, torse, etc.) pour diviser les images en blocs, puis extrait les caractéristiques et mesure les blocs. Dans cet article, les images sont découpées horizontalement, mais la caractéristique du bloc n'est pas calculée pas à pas par bloc coulissant. La caractéristique d'un bloc entier est comptée directement, puis le temps de statistiques de caractéristiques est réduit. La caractéristique de couleur est la caractéristique visuelle la plus largement utilisée dans la recherche d'images. La principale raison en est que la couleur est souvent liée à l'objet ou à la scène contenue dans l'image. Son avantage est qu'elle n'est pas affectée par la rotation et la translation de l'image, et ne peut pas être affectée par le changement d'échelle de l'image avec l'aide de la normalisation. EH Land [31] décrit la classification des caractéristiques de couleur et l'extraction des caractéristiques invariantes à la couleur basées sur l'algorithme Retinex. Les résultats montrent que l'espace colorimétrique HSV peut satisfaire les caractéristiques visuelles humaines et présente certains avantages dans la reconnaissance des piétons. Afin de réduire l'influence de l'éclairage sur les images, cet article utilise l'algorithme mentionné par la théorie de Retinex pour extraire les caractéristiques de couleur originales des images en supprimant le changement de caractéristique des images causé par l'éclairage. La caractéristique de texture est une caractéristique visuelle qui reflète le phénomène homogène dans les images, qui incarne l'attribut inhérent partagé par la surface de l'objet. GL Yang et al. [42] ont proposé l'utilisation de SILTP pour transformer les caractéristiques de texture et les initialiser. Plus les caractéristiques extraites sont détaillées, plus l'image peut être distinguée. La caractéristique des motifs binaires locaux peut mieux décrire les détails de l'image. En même temps, nous devons tenir compte de l'impact de la lumière sur l'image. La caractéristique de texture SILTP est utilisée dans cet article. Cette caractéristique améliore la caractéristique de texture de LBP. Elle peut réduire l'influence du bruit de l'image et modifier l'influence de l'échelle d'éclairage sur la caractéristique de texture de l'image.

La conception des caractéristiques et l'apprentissage de la distance sont deux composantes clés de la Re-ID des personnes. Dans la conception des caractéristiques, plusieurs travaux ont été menés en se concentrant sur les propriétés caractéristiques des images de personnes. Symmetry-Driven Accumulation of Local Features (SDALF) [67] exploite la propriété symétrique d'une personne en obtenant les positions de la tête, du torse et des jambes pour traiter les variations de vue. L'apprentissage de la saillance non supervisé [62] estime des patches rares parmi différentes images, pour effectuer une correspondance d'apparences rares telles que des manteaux de couleurs rares, des bagages et des dossiers rares. Les descripteurs basés sur les attributs obtiennent une description linguistique des images de personnes [71]. Ces travaux ont été principalement réalisés dans des contextes non supervisés.

Notre utilisation de statistiques à deux niveaux (patch/region) pour la Re-ID des personnes est motivée par la méthode récemment proposée Local Maximal Occurrence (LOMO) [75], qui est une représentation de haute dimension des histogrammes de couleurs et du Local Ternary Pattern (SILTP). Cette méthode construit localement un histogramme des caractéristiques des pixels, puis prend ses valeurs maximales dans des bandes horizontales pour surmonter les variations de point de vue tout en maintenant une discrimination locale. En effet, LOMO décrit uniquement l'information moyenne des caractéristiques des pixels. La caractéristique Covariance-of-Covariance [81], où la covariance de la région est estimée sur les covariances locales des patches de caractéristiques des pixels, nous a motivés à ajouter des informations de covariance dans chaque hiérarchie.

### II.3. Descripteur LOMO

La Re-ID des personnes consiste à reconnaître des individus à travers différents capteurs d'un réseau de caméras. Alors que les indices liés à l'apparence vestimentaire sont largement utilisés, d'autres modalités pourraient être exploitées comme sources d'informations supplémentaires, telles que les mesures anthropométriques et la démarche. Dans ce travail, nous examinons si la précision de la Re-ID basée sur l'apparence vestimentaire peut être améliorée en les fusionnant avec les mesures anthropométriques extraites des données de profondeur, en utilisant des capteurs RGB-D, dans des environnements non contraints. Nous proposons également un cadre basé sur la dissimilarité pour construire et fusionner des descripteurs multimodaux d'images de piétons pour des tâches de Re-ID, en tant qu'alternative à la fusion au niveau des scores largement utilisée. L'évaluation expérimentale est réalisée sur deux ensembles de données comprenant des données RGB-D, dont l'un est un ensemble de données nouveau et disponible publiquement que nous avons acquis en utilisant des capteurs Kinect.

Un autre aspect de la Re-ID des personnes concerne la façon de développer une fonction de distance ou de similarité robuste pour aborder le problème complexe de la correspondance. De nombreux algorithmes d'apprentissage de métriques ont été proposés en tenant compte de cet aspect [51, 34]. En pratique, de nombreuses méthodes d'apprentissage de métriques précédentes [39, 27, 34, 49] suivent un processus en deux étapes pour l'apprentissage de la métrique. Tout d'abord, une analyse en composantes principales (PCA) est appliquée pour réduire la dimension, puis l'apprentissage de la métrique est effectué sur le sous-espace PCA. Cependant, cette approche en deux étapes peut ne pas être optimale pour l'apprentissage de la métrique dans un espace de basse dimension, car les échantillons de différentes classes peuvent déjà être mélangés après la première étape. Dans cet article, nous proposons une nouvelle représentation des caractéristiques appelée LOMO, ainsi qu'une méthode d'apprentissage de sous-espace et de métrique appelée XQDA. La fonction LOMO analyse l'occurrence horizontale des caractéristiques locales et maximise cette occurrence pour obtenir une

représentation stable face aux changements d'angle. De plus, nous constatons que l'application de la transformée de Retinex est utile pour traiter les variations d'éclairage dans la Re-ID des personnes. Pour apprendre une métrique discriminante, nous proposons d'apprendre un sous-espace discriminant de basse dimension par une analyse discriminante quadratique inter-vues, et simultanément, une métrique QDA est apprise sur le sous-espace dérivé. Nous montrons que le problème peut être formulé comme un quotient de Rayleigh généralisé, et une solution en forme fermée peut être obtenue par la décomposition des valeurs propres généralisée. Nous présentons également une méthode de calcul pratique pour XQDA, ainsi que sa régularisation et la sélection de dimension. La méthode proposée s'avère être efficace et efficiente à travers des expériences de Re-ID de personnes sur quatre bases de données publiques, et nous démontrons également comment les composants proposés entraînent des améliorations significatives.

La conception des caractéristiques et l'apprentissage de la distance sont deux composantes clés de la Re-ID des personnes. Dans la conception des caractéristiques, plusieurs travaux ont été menés en se concentrant sur les propriétés caractéristiques des images de personnes. Symmetry-Driven Accumulation of Local Features (SDALF) [67] exploite la propriété symétrique d'une personne en obtenant les positions de la tête, du torse et des jambes pour traiter les variations de vue. L'apprentissage de la saillance non supervisé [62] estime des patchs rares parmi différentes images, pour effectuer une correspondance d'apparences rares telles que des manteaux de couleurs rares, des bagages et des dossiers rares. Les descripteurs basés sur les attributs obtiennent une description linguistique des images de personnes [71]. Ces travaux ont été principalement réalisés dans des contextes non supervisés.

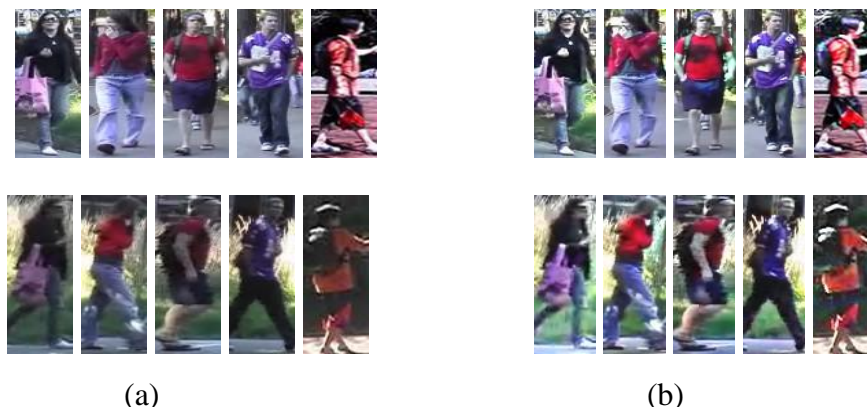
Notre utilisation de statistiques à deux niveaux (patch/region) pour la Re-ID des personnes est motivée par la méthode récemment proposée Local Maximal Occurrence (LOMO) [75], qui est une représentation de haute dimension des histogrammes de couleurs et du Local Ternary Pattern (SILTP). Cette méthode construit localement un histogramme des caractéristiques des pixels, puis prend ses valeurs maximales dans des bandes horizontales pour surmonter les variations de point de vue tout en maintenant une discrimination locale. En effet, LOMO décrit uniquement l'information moyenne des caractéristiques des pixels. La caractéristique Covariance-of-Covariance [81], où la covariance de la région est estimée sur les covariances locales des patchs de caractéristiques des pixels, nous a motivés à ajouter des informations de covariance dans chaque hiérarchie.

## II.4. Caractéristique LOMO

### II.4.1. Gestion des variations d'éclairage

La couleur joue un rôle crucial dans la description des images de personnes. Cependant, les conditions d'éclairage peuvent grandement différer entre les caméras, et les paramètres de réglage de la caméra peuvent varier également. Par conséquent, les couleurs perçues pour une même personne peuvent être très différentes selon les différentes vues des caméras. Par exemple, la Figure (II.1.a) illustre quelques échantillons d'images provenant de la base de données VIPeR [42]. On peut constater des variations significatives dans l'éclairage et l'apparence des couleurs des images de la même personne prises par les deux caméras. Dans cette étude, nous proposons d'appliquer l'algorithme Retinex [45, 44, 43] pour prétraiter les images de personnes. Le Retinex tient compte de la luminosité humaine et de la perception des couleurs, visant à produire une image en couleur cohérente avec la perception humaine de la scène. L'image restaurée présente généralement des couleurs vives, avec des détails améliorés notamment dans les zones ombragées. Nous mettons en

œuvre l'algorithme Retinex multi-échelles selon [43], qui combine le Retinex à petite échelle pour la compression de la plage dynamique et le Retinex à grande échelle pour la rendition tonale simultanée. De cette manière, l'algorithme gère automatiquement à la fois la cohérence des couleurs et la compression de la plage dynamique, offrant ainsi une bonne approximation de la perception visuelle humaine. Plus spécifiquement, nous utilisons deux échelles de Retinex centre/périphérie, avec  $\sigma = 5$  et  $\sigma = 20$ . De plus, nous calculons automatiquement les paramètres de gain/décalage de manière à ce que les intensités résultantes s'étendent linéairement dans l'intervalle  $[0, 255]$ .



**Fig.II.1: (a) Exemples de paires d'images de la base de données VIPeR. (b) Images traitées dans (a) par Retinex. Les images dans la même colonne représentent la même personne[9].**

La figure(II.1.(b)) montre quelques exemples des images traitées par notre implémentation de Retinex. En comparaison avec la figure II.2.1(a), on peut observer que les images Retinex de la même personne à travers les caméras présentent une meilleure cohérence en termes d'éclairage et de couleur. Cela facilite la Re-ID des personnes par rapport à l'utilisation des images d'origine. Avec les images Retinex, nous appliquons l'histogramme de couleur HSV pour extraire les caractéristiques de couleur.

#### II.4.2. Gestion des changements de point de vue

Les piétons apparaissent souvent sous différents angles de vue lorsqu'ils sont capturés par différentes caméras. Par exemple, une personne vue de face par une caméra peut apparaître de dos sous une autre caméra. Par conséquent, il est également difficile de faire correspondre les personnes capturées sous différents angles de vue. Pour remédier à cela, [50, 51] a suggéré de diviser une image de personne en six bandes horizontales, calculant ainsi un histogramme unique dans chaque bande. Cette approche a réussi à représenter de manière invariante au point de vue [50, 51, 47]. Cependant, elle peut également perdre des détails spatiaux au sein de chaque bande, ce qui affecte sa capacité de discrimination. Nous proposons d'utiliser des fenêtres coulissantes pour décrire les détails locaux d'une image de personne. Plus précisément, nous utilisons des sous-fenêtres de taille  $10 \times 10$ , avec un pas de chevauchement de 5 pixels pour localiser des patches locaux dans des images de taille  $128 \times 48$ . Dans chaque sous-fenêtre, nous extrayons deux échelles d'histogrammes SILTP ( $SILTP_{4,3}^{0,3}$  et  $SILTP_{4,5}^{0,3}$ ), ainsi qu'un histogramme HSV conjoint de  $8 \times 8 \times 8$ . Chaque bin d'histogramme représente la probabilité d'occurrence d'un motif dans une sous-fenêtre. Pour prendre en compte les variations d'angle de vue, nous examinons toutes les sous-fenêtres à la même position horizontale, et maximisons l'occurrence locale de chaque motif (c'est-à-dire le même bin d'histogramme) parmi ces sous-fenêtres. L'histogramme résultant acquiert ainsi une certaine invariance aux variations d'angle de vue, tout en capturant les caractéristiques de la région locale d'une personne. La fig(II.2) illustre la procédure d'extraction de la caractéristique LOMO que nous proposons.

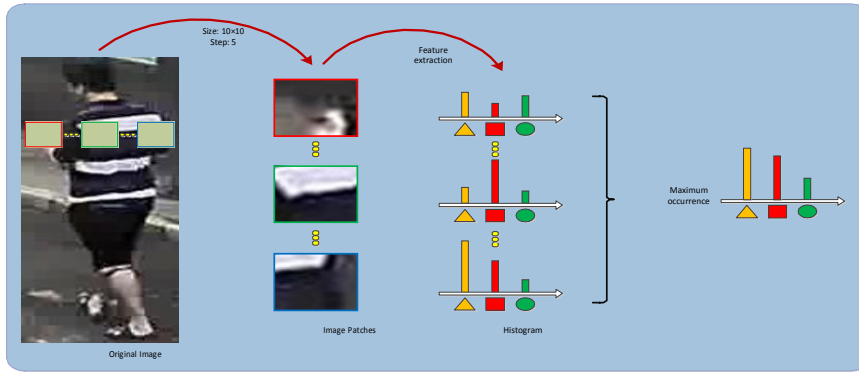


Fig. II.2: Illustration de la méthode d'extraction des caractéristiques LOMO[92].

Pour approfondir la considération de l'information multi-échelle, nous construisons une représentation pyramidale à trois échelles, qui réduit l'échantillonnage de l'image originale de  $128 \times 48$  par deux opérations locales de pooling moyen  $2 \times 2$ , et répète la procédure d'extraction de caractéristiques ci-dessus. En concaténant toutes les occurrences locales maximales calculées, notre descripteur final a une dimension de  $(8 \times 8 \times 8 \text{ bins de couleur} + 34 \times 2 \text{ bins SILTP}) \times (24 + 11 + 5 \text{ groupes horizontaux}) = 26,960$  dimensions. Enfin, nous appliquons une transformation logarithmique pour supprimer les valeurs élevées des bins, et normalisons à la fois les caractéristiques HSV et SILTP à une longueur unitaire. Étant donné que nous utilisons uniquement des caractéristiques HSV et SILTP simples, la méthode proposée d'extraction de caractéristiques est efficace à calculer.

## II.5. Analyse discriminante quadratique inter-vues

### II.5.1 Réexamen de l'approche Bayesian Face et KISSME

Considérons une différence d'échantillons  $\Delta = x_i - x_j$ .  $\Delta$  est appelée la différence intra personnelle si  $y_i = y_j$ , tandis qu'elle est appelée la différence extra personnelle si  $y_i \neq y_j$  [48]. En conséquence, deux classes de variations peuvent être définies : les variations intra personnelles  $\Omega_I$  et les variations extra personnelles  $\Omega_E$ . Par conséquent, de cette manière, le problème de classification multi classe peut être résolu en distinguant ces deux classes. Moghaddam et al. [48] ont proposé de modéliser chacune des deux classes avec une distribution gaussienne multi variée. Cela correspond à un modèle QDA avec  $\Omega_I$  et  $\Omega_E$  définis comme deux classes distinctes. De plus, il a été remarqué dans [48] que  $\Omega_I$  et  $\Omega_E$  ont toutes deux une moyenne nulle. L'algorithme résultant est appelé "Bayesian Face" appliqué à la reconnaissance faciale. De manière intéressante, dans [34], Kostinger et al. ont également dérivé une approche similaire appelée KISSME via le test du rapport de vraisemblance des deux distributions gaussiennes, et l'ont appliquée à la Re-ID de personnes.

Formellement, les algorithmes Bayesian Face et KISSME sont formulés comme suit. Sous la distribution gaussienne de moyenne nulle, les vraisemblances d'observer  $\Delta$  dans  $\Omega_I$  et  $\Omega_E$  sont définies comme suit:

$$P(\Delta|\Omega_I) = \frac{1}{(2\pi)^{D/2} |\Sigma_I|^{1/2}} e^{-\frac{1}{2}\Delta^T \Sigma_I^{-1} \Delta} \quad (1)$$

$$P(\Delta|\Omega_E) = \frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_E|^{\frac{1}{2}}} e^{-\frac{1}{2}\Delta^T\Sigma_E^{-1}\Delta} \quad (2)$$

où  $\Sigma_I$  et  $\Sigma_E$  sont les matrices de covariance de  $\Omega_I$  et  $\Omega_E$  respectivement, et  $n_I$  et  $n_E$  représentent le nombre d'échantillons dans les deux classes. En appliquant la règle bayésienne et le test du rapport de vraisemblance logarithmique, la fonction de décision peut être simplifiée comme suit :

$$f(\Delta) = \Delta^T(\Sigma_I^{-1} - \Sigma_E^{-1})\Delta \quad (3)$$

et donc, la fonction de distance dérivée entre  $x_i$  et  $x_j$  est

$$d(x_i, x_j) = (x_i - x_j)^T(\Sigma_I^{-1} - \Sigma_E^{-1})(x_i - x_j). \quad (4)$$

Par conséquent, apprendre la fonction de distance correspond à estimer les matrices de covariance  $\Sigma_I$  et  $\Sigma_E$ .

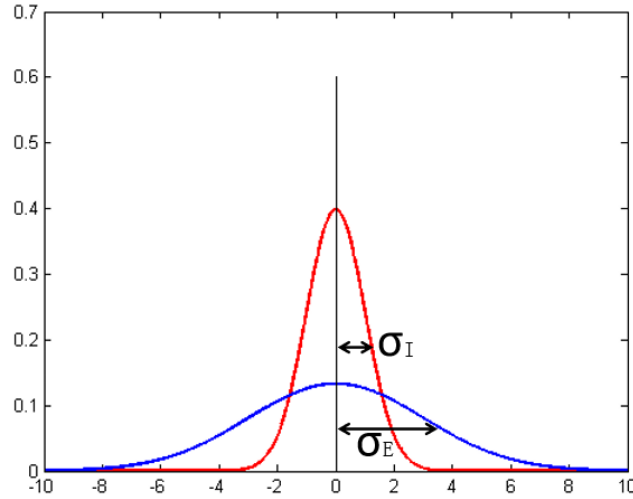


Fig. II.3.1. Répartitions de  $\Omega_I$  et  $\Omega_E$  dans une dimension projetée[92].

### II.5.2. XQDA :

Habituellement, les dimensions originales des caractéristiques  $d$  sont grandes, et un espace dimensionnel bas  $R^r$  (où  $r < d$ ) est préféré pour la classification. [48] a suggéré de décomposer  $\Sigma_I$  et  $\Sigma_E$  séparément pour réduire les dimensions. Dans [34], l'ACP a été appliquée, puis  $\Sigma_I$  et  $\Sigma_E$  ont été estimés dans le sous-espace de l'ACP. Cependant, les deux méthodes ne sont pas optimales car la réduction de dimension ne prend pas en compte l'apprentissage de la distance métrique.

Dans cet article, nous étendons les approches bayésiennes du visage et KISSME à l'apprentissage de métriques entre vues croisées, où nous cherchons à apprendre un sous-espace

$W = (w_1, w_2, \dots, w_r) \in R^{d \times r}$  avec des données entre vues croisées, et en même temps à apprendre une fonction de distance dans le sous-espace de dimension  $r$  pour la mesure de similarité entre vues croisées. Supposons que nous ayons un ensemble d'entraînement entre vues croisées  $\{X, Z\}$  de  $c$  classes, où  $X = (x_1, x_2, \dots, x_n) \in R^{d \times n}$  contient  $n$  échantillons dans un espace  $d$ -dimensionnel depuis une vue,  $Z = (z_1, z_2, \dots, z_m) \in R^{d \times m}$  contient  $m$  échantillons dans le même espace  $d$ -dimensionnel mais depuis l'autre vue. Le problème d'appariement entre vues croisées découle de nombreuses applications, telles que la reconnaissance de visages hétérogènes [46] et la Re-ID de personnes

invariante par rapport à la vue [32]. Notez que  $\mathbf{Z}$  est identique à  $\mathbf{X}$  dans le scénario d'appariement à vue unique. En considérant un sous-espace  $\mathbf{W}$ , la fonction de distance Eq. (4) dans le sous-espace de dimension  $\mathbf{r}$  est calculée comme suit :

$$d_W(\mathbf{x}, \mathbf{z}) = (\mathbf{x} - \mathbf{z})^T \mathbf{W} (\Sigma_I' - 1 - \Sigma_E' - 1) \mathbf{W}^T (\mathbf{x} - \mathbf{z}) \quad (5)$$

où  $\Sigma_I = \mathbf{W}^T \Sigma_I' \mathbf{W}$  et  $\Sigma_E = \mathbf{W}^T \Sigma_E' \mathbf{W}$ . Par conséquent, nous devons apprendre une matrice de noyau  $\mathbf{M}(\mathbf{W}) = \mathbf{W} (\Sigma_I^{-1} - \Sigma_E^{-1}) \mathbf{W}^T$ . Cependant, optimiser directement  $d_W$  est difficile car  $\mathbf{W}$  est contenu dans deux matrices inverses.

Rappelez-vous que  $\Omega_I$  et  $\Omega_E$  ont une moyenne nulle, alors, étant donné une base  $\mathbf{w}$ , les échantillons projetés des deux classes resteront centrés sur zéro, mais peuvent avoir des variances différentes, comme le montre la Fig II.3.1. Dans ce cas, le critère de Fisher traditionnel utilisé pour dériver l'ADL n'est plus approprié car les deux classes ont la même moyenne. Cependant, les variances  $\sigma_I$  et  $\sigma_E$  peuvent toujours être utilisées pour distinguer les deux classes. Par conséquent, nous pouvons optimiser la direction de projection  $\mathbf{w}$  de manière à ce que  $\sigma_E(\mathbf{w})/\sigma_I(\mathbf{w})$  soit maximisé.  $\sigma_I(\mathbf{w}) = \mathbf{w}^T \Sigma_I \mathbf{w}$  et  $\sigma_E(\mathbf{w}) = \mathbf{w}^T \Sigma_E \mathbf{w}$ , donc l'objectif  $\sigma_E(\mathbf{w})/\sigma_I(\mathbf{w})$  correspond au quotient généralisé de Rayleigh.

$$J(\mathbf{w}) = \frac{\mathbf{w}^T \Sigma_E \mathbf{w}}{\mathbf{w}^T \Sigma_I \mathbf{w}} \quad (6)$$

La maximisation de  $J(\mathbf{w})$  est équivalente à

$$\max_{\mathbf{w}} \mathbf{w}^T \Sigma_E \mathbf{w}, \text{ s.t. } \mathbf{w}^T \Sigma_I \mathbf{w} = 1, \quad (7)$$

Cela peut être résolu par le problème de décomposition en valeurs propres généralisées, de manière similaire à l'ADL. Autrement dit, la plus grande valeur propre de  $\Sigma_I^{-1} \Sigma_E$  est la valeur maximale de  $J(\mathbf{w})$ , et le vecteur propre correspondant  $\mathbf{w}_1$  est la solution. De plus, la solution orthogonale à  $\mathbf{w}_1$  et correspondant à la deuxième plus grande valeur de  $J(\mathbf{w})$  est le vecteur propre de la deuxième plus grande valeur propre de  $\Sigma_I^{-1} \Sigma_E$ , et ainsi de suite.

Par conséquent, avec  $\mathbf{W} = (\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_r)$ , nous apprenons un sous-espace discriminant, ainsi qu'une fonction de distance dans le sous-espace appris, comme défini dans l'équation (5). Nous appelons l'algorithme dérivé Analyse Discriminante Quadratique Inter-Vues (XQDA) pour refléter sa connexion avec l'ADQ et la sortie d'une métrique inter-vues.

## II.6. Rôle de Retinex

Le modèle Retinex est une théorie qui vise à expliquer et améliorer la perception des couleurs par les humains sous différentes conditions d'éclairage. Ce modèle est utilisé dans le traitement d'images pour améliorer la clarté des couleurs et le contraste, rendant les images plus proches de ce que l'œil humain perçoit réellement. Le modèle Retinex permet d'améliorer la qualité des images en corrigeant les éclairages inégaux et en supprimant les ombres, ce qui accentue les détails et les couleurs réelles de l'image. Ce modèle est largement utilisé dans les applications de photographie et de traitement d'images numériques pour améliorer l'apparence des images dans diverses conditions d'éclairage, garantissant ainsi que les couleurs des images soient claires et ressemblent aux couleurs naturelles perçues par les humains.

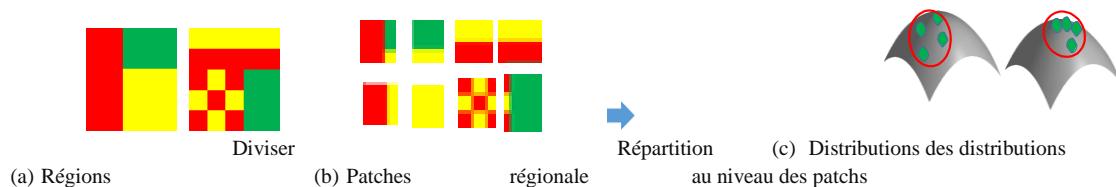
## II.7. Rôle de l'Occurrence Maximale Locale

La technique de l'Occurrence Maximale Locale (LOMO) dans l'analyse d'images vise à améliorer l'extraction des caractéristiques fines et des détails importants des images numériques. En identifiant les points avec des valeurs maximales locales, LOMO aide à augmenter la précision des modèles de reconnaissance et de classification, comme la reconnaissance faciale, même en présence de bruit ou de variations de luminosité. LOMO se distingue par son efficacité à gérer les changements de lumière et les angles de vue, ce qui la rend utile dans de nombreuses applications. Elle est également plus efficace sur le plan computationnel car elle se concentre sur les points saillants, réduisant ainsi le temps de traitement des images et permettant des applications en temps réel. En résumé, LOMO améliore la précision et la fiabilité de l'analyse d'images dans divers domaines.

## II.8. Descripteur GOG

La recherche de correspondance d'apparence des images de personnes, observées à partir de différentes vues de caméra et appelée Re-ID de personne, suscite un intérêt croissant en raison de sa diversité d'applications [55]. Dans cette tâche, les images de personnes sont capturées sous des angles variés, avec différentes conditions d'éclairage, résolutions, poses humaines et environnements. Ces variations intra-personnelles significatives dans les images de personnes présentent des défis sérieux. De plus, la similarité des vêtements entre différentes personnes ajoute des complexités supplémentaires.

Pour surmonter ces défis, les chercheurs se penchent activement sur le développement de descripteurs d'apparence [56, 57, 58, 61, 62, 63] et de méthodes pour les appairer [74, 78, 59, 60, 64]. Les descripteurs caractérisent l'information d'apparence, incluant la couleur et la texture, des vêtements humains. Un bon descripteur devrait



**Fig.II.6. Importance de la distribution hiérarchique: (a) Régions ayant la même distribution (moyenne/covariance) des caractéristiques des pixels (chaque couleur indique le même vecteur de caractéristiques). (b) Patches locaux à l'intérieur des régions ayant une distribution différente des caractéristiques des pixels. (c) Les régions peuvent être distinguées par les distributions des distributions au niveau des patchs[86].**

Être résistant face aux variations intra-personnelles tout en possédant une grande capacité discriminative pour distinguer différentes personnes est essentiel.

Les images de personnes ont une résolution faible et présentent d'importantes variations de pose. Par conséquent, il a été démontré que l'indice le plus crucial pour la Re-ID des personnes est l'information de couleur, notamment les histogrammes de couleur et les descripteurs de noms de couleur [61]. Étant donné qu'ils ne parviennent pas à différencier de manière suffisante les personnes de couleur similaire, des descripteurs texturaux tels que le Local Binary Pattern (LBP) et les réponses des banques de filtres sont souvent combinés avec des descripteurs de couleur [59, 60, 64].

Un descripteur de covariance [83] décrit une région d'intérêt en tant que covariance des caractéristiques des pixels. Il offre une manière naturelle de fusionner différentes modalités, telles

que la couleur et la texture, des caractéristiques des pixels en un méta-descripteur unique. Étant donné que le descripteur de covariance est obtenu en moyennant les caractéristiques à l'intérieur de la région, il permet de remédier aux effets du bruit et aux désalignements spatiaux. Par conséquent, il a été appliqué avec succès à la Re-ID des personnes [53, 54, 85].

Dans cet article, nous proposons un nouveau descripteur de région basé sur une distribution gaussienne hiérarchique des caractéristiques des pixels pour la Re-ID des personnes. Plus précisément, nous extrayons de manière dense des patches locaux à l'intérieur d'une région et considérons la région comme un ensemble de patches locaux. Nous modélisons d'abord la région comme un ensemble de multiples distributions gaussiennes, chacune représentant l'apparence d'un patch local. Nous appelons une telle distribution gaussienne représentant chaque patch local un patch gaussien. Les caractéristiques de l'ensemble des patches gaussiens sont à nouveau décrites par une autre distribution gaussienne. Nous appelons cette distribution gaussienne une gaussienne de région. Les paramètres de la gaussienne de région sont ensuite utilisés comme vecteur de caractéristiques pour représenter la région.



**Fig.II.6.1. Importance de la moyenne :** (a) Images originales. (b) Images montrant les valeurs moyennes RGB de patches de pixels de  $10 \times 10$  de (a). (c) Images avec la moyenne retirée (chaque valeur RGB est mise à l'échelle dans la plage  $[0,255]$  pour la visualisation). Il est facile de déterminer les mêmes personnes à partir de (b), tandis que c'est difficile à partir de (c) [86].

Notre motivation pour l'utilisation d'un modèle hiérarchique découle de la structure d'apparence des images de personnes. Les vêtements des personnes sont composés de parties locales, chacune ayant des structures de couleur/textures locales. L'arrangement spatial de ces parties détermine la structure d'apparence globale. Cependant, la plupart des descripteurs méta existants [68, 69, 56, 77, 82, 83] sont basés sur une distribution globale des caractéristiques des pixels à l'intérieur d'une région. Ainsi, la structure locale de l'image de la personne est perdue. En revanche, notre descripteur proposé décrit la distribution globale en utilisant la distribution locale des caractéristiques des pixels. En effet, il peut distinguer les textures qui ont la même distribution globale mais des structures locales différentes, comme illustré dans la Fig. 2.6.

Nous utilisons la distribution gaussienne comme composant de base de la hiérarchie. La motivation de l'utilisation de la distribution découle de l'importance de la couleur moyenne des parties locales. Bien que la représentation hiérarchique des descripteurs de covariance ait été proposée [72, 81], l'information de moyenne n'est pas incluse dans chaque hiérarchie. La perte d'information de moyenne est un problème crucial lorsqu'ils sont appliqués à la Re-ID des personnes. Cela est dû au fait que les vêtements qu'une personne porte ont tendance à se composer d'un petit nombre de couleurs dans chaque partie locale, et par conséquent, la couleur moyenne dans les parties locales a tendance à être l'information discriminative majeure des personnes. Comme le montre la Fig. 2.6.1, les images moyennes de couleur locale contiennent des informations fortement distinctives de différentes personnes. Nous nommons la méthode hiérarchique proposée "Gaussian Of Gaussian" (GOG) descriptor. Le descripteur GOG offre une manière conceptuellement simple et

cohérente de générer des caractéristiques discriminatives et robustes décrivant simultanément les informations de couleur et de texture. Les résultats d'expériences approfondies menées sur cinq ensembles de données publics révèlent que, malgré sa simplicité, notre descripteur proposé peut atteindre des performances étonnamment élevées en matière de Re-ID des personnes.

## II.9. Descripteur Gaussien Hiérarchique

Nous présentons le descripteur gaussien hiérarchique proposé appelé GOG dans la fig 2.8 Pour parvenir à la représentation des caractéristiques d'une image de personne, nous adoptons un modèle basé sur les parties [80]. Nous supposons que  $G$  régions d'une image de personne sont données à l'avance, ce sont généralement des bandes horizontales de l'image. Le descripteur proposé renvoie un vecteur de caractéristiques des régions. Le reste de cette section décrit les détails du descripteur.

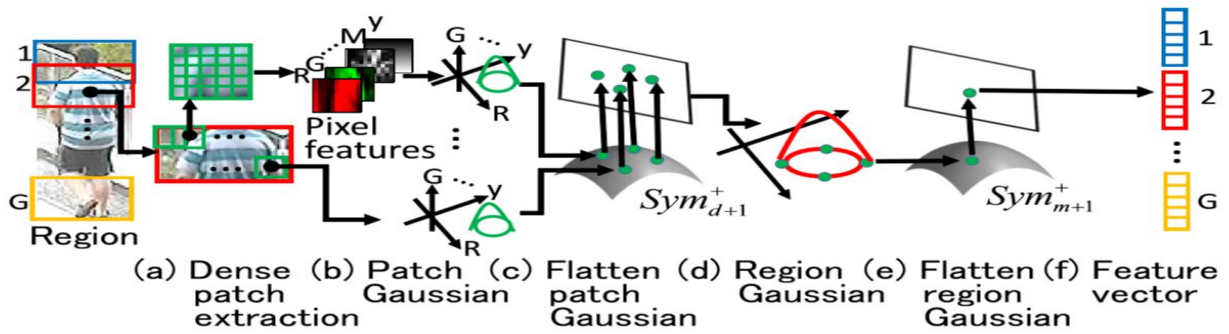


Fig II.8. Descripteur GOG [86].

### II.9.1. Caractéristiques des pixels

Intéressons-nous à l'une des  $G$  régions d'une image de personne. Pour décrire la structure locale de la région, nous extrayons de manière dense des patches carrés ( $k \times k$  pixels) avec des intervalles de  $p$  pixels (Fig. 2.8(a)). Afin de caractériser chaque pixel dans le patch, nous extrayons un vecteur de caractéristiques de dimension  $d$  ( $f_i$ ) pour chaque pixel  $i$ . Le vecteur de caractéristiques peut être de n'importe quel type, tel que la couleur, l'intensité, l'orientation du gradient et la réponse du filtre. Étant donné que le nombre de pixels dans chaque patch est petit, il est préférable que la dimension  $d$  soit faible pour estimer de manière robuste les matrices de covariance des gaussiennes de patch à l'étape suivante. Dans ce travail, nous extrayons des caractéristiques de pixels définies comme étant 8-dimensionnelles :

$$f_i = [y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B]^T, \quad (1)$$

où  $y$  est la position du pixel dans la direction verticale,  $M_\theta \in \{0^\circ, \dots, 270^\circ\}$  représente les magnitudes du gradient d'intensité des pixels le long de quatre orientations, et  $R, G, B$  sont les valeurs des canaux de couleur. Chaque dimension de  $f_i$  est étirée linéairement dans la plage  $[0, 1]$  pour égaliser les échelles des différentes valeurs des caractéristiques.

La position du pixel est introduite pour tirer parti de l'information spatiale dans chaque région. L'utilisation uniquement de la position verticale de l'image découle de l'analyse dans [56] ; les images de personnes ont tendance à être bien alignées dans la direction verticale, tandis que le changement de pose ou de point de vue entraîne un désalignement important dans la direction horizontale. Notez que l'on aimerait définir  $y_i$  depuis le haut (ou le centre) de la région actuelle comme dans [68]. Cependant, chaque pixel appartient à plusieurs régions et un tel réglage augmente la complexité computationnelle. Comme les images de personnes sont grossièrement alignées, nous

définissons directement  $y_i$  depuis le haut de l'image.

L'information sur le gradient est introduite pour décrire l'information texturale des vêtements. L'orientation du gradient  $O = \arctan(I_y/I_x)$  est calculée à partir des dérivées par rapport à  $x$  et  $y$ ,  $I_x$ ,  $I_y$ , de l'intensité  $I$ . Nous quantifions l'orientation en quatre bins ;  $O_\theta \in \{0^\circ, 90^\circ, 180^\circ, 270^\circ\}$ . Pour compenser la perte d'information due à la quantification, nous utilisons un vote doux dans les deux bins d'orientation voisins. Les poids de vote sont déterminés linéairement à partir des distances par rapport aux orientations quantifiées, comme dans le vecteur GO dans [70]. Pour se concentrer sur les arêtes à fort gradient, nous multiplions la magnitude du gradient  $M = \sqrt{I_x^2 + I_y^2}$  par l'orientation quantifiée  $O_\theta$  et obtenons la magnitude du gradient orienté ;  $M_\theta = MO_\theta$ . L'information couleur est l'indice le plus important pour la reconnaissance des personnes. Nous utilisons les valeurs des canaux de couleur de l'espace couleur le plus basique : RGB. D'autres espaces couleur, tels que Lab, HSV et YCbCr, pourraient être utilisés. En fait, nous étendrons nos caractéristiques de pixels à différents espaces couleur (Section 3.5)

### II.9.2. Région Gaussiens

Après avoir extrait les caractéristiques des pixels à l'intérieur d'un patch, nous les résumons ensuite à l'aide de la distribution paramétrique la plus classique qui a la moyenne et la covariance comme paramètres : la distribution gaussienne (Fig. 2.8 (b)). Pour chaque patch  $s$ , nous modélisons les vecteurs de caractéristiques comme la gaussienne du patch  $N(f; \mu_s, \Sigma_s)$  définie comme suit :

$$N(f; \mu_s, \Sigma_s) = \frac{\exp(-\frac{1}{2}(f-\mu_s)^T \Sigma_s^{-1}(f-\mu_s))}{(2\pi)^{d/2} |\Sigma_s|}, \quad (2)$$

où  $|\cdot|$  est le déterminant d'une matrice,  $\mu_s$  est le vecteur moyen, et  $\Sigma_s$  est la matrice de covariance du patch échantillonné  $s$ . Le vecteur moyen et la matrice de covariance sont respectivement estimés par :  $\mu_s = \frac{1}{n_s} \sum_{i \in L_s} f_i$  et  $\Sigma_s = \frac{1}{n_s-1} \sum_{i \in L_s} (f_i - \mu_s)(f_i - \mu_s)^T$ , où  $L_s$  est la zone du patch échantillonné  $s$  et  $n_s$  représente le nombre de pixels dans  $L_s$ .

Notez que les vecteurs moyens et les matrices de covariance échantillonnés de manière dense peuvent être calculés efficacement à l'aide d'images intégrales [84]. Étant donné que les régions peuvent se chevaucher, nous construisons les images intégrales des caractéristiques des pixels pour l'image globale de la personne plutôt que de les créer pour chaque région. Pour une description plus précise des distributions, un modèle de mélange gaussien (GMM) pourrait être utilisé. Étant donné qu'un patch local est censé être composé d'un petit nombre de couleurs/textures, nous supposons que la gaussienne unimodale est suffisante pour décrire la distribution de ses caractéristiques de pixels.

### II.9.3. Mapping de l'espace tangent et demi-vectorisation

Comme nous l'expliquerons dans la prochaine sous-section, notre descripteur est une représentation résumée des Gaussiennes de patches dans une région. Pour cette synthèse, des opérations mathématiques telles que la moyenne ou la covariance de la gaussienne sont nécessaires.

Du point de vue de la géométrie de l'information, l'espace des distributions de probabilité est considéré comme une variété riemannienne où l'opération euclidienne ne peut pas être appliquée directement. Une variété riemannienne peut être localement aplatie dans un espace euclidien en la projetant dans un espace tangent doté d'une métrique riemannienne. L'espace des matrices définies

positives symétriques (SPD) est également considéré comme une variété riemannienne, et cet espace est récemment bien compris. La métrique log-euclidienne pour les matrices SPD offre une méthode solide pour mapper un point sur la variété vers un espace euclidien tangent via un logarithme matriciel

Pour tirer avantage de la métrique log-euclidienne, nous incorporons les Gaussiennes de patch dans la matrice SPD de manière similaire au travail [73]. Selon une analyse de la littérature sur la géométrie de l'information [76], l'espace des Gaussiennes multivariées d-dimensionnelles peut être intégré dans l'espace des matrices SPD de dimension  $d + 1$ , noté  $Sym_{d+1}^+$ . Nous représentons la Gaussienne de patch d-dimensionnelle  $N(\mu_s, \Sigma_s)$  dans  $Sym_{d+1}^+$  sous la forme de  $P_s$ .

$$N(f; \mu_s, \Sigma_s) \sim P_s = |\Sigma_s|^{-1} \begin{bmatrix} \Sigma_s & \mu_s \mu_s^T & \mu_s \\ \mu_s^T & 1 & \end{bmatrix}.$$

Pour une théorie plus détaillée sur cet plongement, on peut se référer à la littérature [76]. La matrice de covariance du patch local devient souvent singulière en raison du manque de pixels suffisants à l'intérieur du patch. Nous évitons ce problème en ajoutant la matrice identité  $\text{Id}$  à  $\Sigma_s$  avec une petite valeur constante positive,  $\epsilon_s$ , comme suit :  $\Sigma_s \leftarrow \Sigma_s + \epsilon_s \text{Id}$ .

Afin de décrire la distribution de la région dans une opération euclidienne, nous cartographions ensuite chacune des Gaussiennes de patch  $P_s$  dans un espace tangent via un logarithme matriciel (Fig 2.8 (c)). Nous stockons ensuite la partie triangulaire supérieure de la matrice cartographiée sous forme de vecteur, car la matrice est symétrique. En considérant que les entrées hors-diagonale sont comptées deux fois lors du calcul de la norme [84], la matrice de la Gaussienne de patch  $P_s$  devient un vecteur  $g_s$  de dimension

$m = (d^2 + 3d)/2 + 1$ , défini comme suit :

$$g_s = \text{vec}(\log(P_s)) = [b_s(1,1), \sqrt{2}b_s(1,2), \dots, \sqrt{2}b_s(1,d+1), b_s(2,2), \sqrt{2}b_s(2,3), \dots, b_s(d+1,d+1)]^T$$

où  $\log(\cdot)$  est l'opérateur de logarithme matriciel et  $b_s(i,j)$  est l'élément  $(i,j)$  de  $B_s = \log(P_s)$ .

#### II.9.4. Gaussienne de région sur l'espace tangent

En raison de la variation de la pose des images de personnes, les positions des parties locales varient dans différentes observations. Ainsi, nous résumons les patches locaux dans une représentation sans ordre. Plus précisément, nous résumons les Gaussiennes de patch aplaties dans la section précédente en une distribution régionale (Fig. 2.8(d)). Pour cette synthèse, nous utilisons également une distribution gaussienne qui peut décrire non seulement la covariance mais aussi la moyenne. Encore une fois, un modèle de mélange gaussien (GMM) pourrait être utilisé pour décrire des distributions plus précises. Cependant, l'appariement entre les GMM n'est pas un problème trivial [73] et entraînera une complexité pour l'appariement entre les descripteurs de région. La synthèse avec une distribution gaussienne est réalisée en tenant compte d'une propriété spatiale des patches de la manière suivante.

Une image de personne contient souvent des régions d'arrière-plan qui diffèrent significativement d'un endroit à un autre. Pour atténuer l'effet des régions d'arrière-plan, nous introduisons un poids pour chaque patch de manière similaire aux histogrammes de couleur pondérés [67]. Dans la plupart des cas, la personne est centrée dans chaque image ; ainsi, une valeur plus élevée est attribuée aux patches qui sont plus proches de l'axe central  $y$  d'une image :  $w_s = \exp(-(x_s - x_c)^2/2\sigma^2)$  où

$x_c = W/2$ ,  $\sigma = W/4$ . Ici,  $x_s$  représente la coordonnée x du pixel central du patch s et W est la largeur de l'image. Ensuite, nous définissons le vecteur moyen pondéré et la matrice de covariance comme suit :

$$\boldsymbol{\mu}^G = \frac{1}{\sum_{s \in g} w_s} \sum_{s \in g} w_s \mathbf{G}_s, \quad (4)$$

$$\boldsymbol{\Sigma}^G = \frac{1}{\sum_{s \in g} w_s} \sum_{s \in g} w_s (\mathbf{G}_s - \boldsymbol{\mu}_s) (\mathbf{G}_s - \boldsymbol{\mu}_s)^T, \quad (5)$$

où G est la région dans laquelle les Gaussiennes de patch sont résumées. En utilisant le vecteur moyen et la matrice de covariance, nous représentons la région sous forme de la Gaussienne de région  $N(\mathbf{g}; \boldsymbol{\mu}^G, \boldsymbol{\Sigma}^G)$ . Pour l'appariement entre les descripteurs de région, il est pratique de linéariser la Gaussienne de région dans l'espace euclidien, car la plupart des méthodes d'appariement, telles que l'apprentissage de métriques, sont conçues dans un espace euclidien. À cette fin, nous intégrons la Gaussienne de région de dimension m dans les matrices SPD de dimension m + 1 de la même manière que l'équation (3) :

$N(\mathbf{g}; \boldsymbol{\mu}^G, \boldsymbol{\Sigma}^G) \sim Q$  où Q est une matrice SPD de taille  $(m + 1) \times (m + 1)$ . Ici, la matrice de covariance  $\boldsymbol{\Sigma}^G$  est régularisée comme suit :  $\boldsymbol{\Sigma}^G \leftarrow \boldsymbol{\Sigma}^G + \epsilon^G I_m$ . Nous mappons ensuite Q dans l'espace tangent de  $\text{Sym}^+_{m+1}$  en utilisant le logarithme matriciel et le demi-vectorisons pour former un vecteur de caractéristiques de dimension  $(m^2 + 3m)/2 + 1$ , que nous notons  $\mathbf{z}$  (Fig. 2.8(e)). En extrayant la Gaussienne de région pour chacune des régions G, nous obtenons les vecteurs de caractéristiques  $\{\mathbf{z}_g\}_g^G = \mathbf{1}$ . Afin de conserver l'emplacement spatial

de ces vecteurs, nous les concaténons et formons un vecteur de caractéristiques (Fig. 3(f)). Ensuite, la représentation des caractéristiques d'une image de personne devient:  $\mathbf{z} = [\mathbf{z}_1^T, \dots, \mathbf{z}_G^T]^T$

### II.9.5 Fusion du descripteur de différentes espaces de couleurs

Il a été prouvé que les descripteurs extraits de différentes espaces de couleurs ont des propriétés complémentaires, et leur fusion améliore les précisions de Re-ID [61]. Pour extraire plus d'informations sur la couleur dans les descripteurs GOG, nous remplaçons les valeurs des canaux RGB dans la caractéristique du pixel dans l'équation (1) par les valeurs de trois canaux de couleurs alternatives {Lab, HSV, nRGB} et fusionnons leurs descripteurs GOG. Ici, le nRGB est l'espace de couleurs normalisé (par exemple,  $nR = R/(R+G+B)$ ). Comme il y a une redondance dans cet espace, nous n'utilisons que {nR, nG} dans cet espace de couleur. Ainsi, la dimension de la caractéristique du pixel de chaque espace de couleur {RGB, Lab., HSV, nRnG} est  $d = \{8,8,8,7\}$  et donc la dimension du vecteur gaussien de patch est  $m = \{45,45,45,36\}$ .

Nous désignons le descripteur GOG  $\mathbf{z}$  extrait de l'équation (1) comme  $GOG_{RGB}$ , et les descripteurs extraits des canaux de couleurs alternatifs comme  $GOG_{Lab}$ ,  $GOG_{HSV}$  et  $GOG_{nRnG}$ , respectivement. La fusion est simplement réalisée en concaténant les descripteurs GOG sur différentes caractéristiques de pixels comme suit :

$$GOG_{Fusion} = [GOG_{RGB}^T, GOG_{Lab}^T, GOG_{HSV}^T, GOG_{nRnG}^T]^T.$$

Par conséquent, la dimensionalité du descripteur de fusion est de  $3$  (espaces de couleurs)  $\times 1081 = (45^2 + 3 \times 45)/2 + 1 \times G$  (régions)  $+ 1$  (espace de couleur)  $\times 703 = (36^2 + 3 \times 36)/2 + 1 \times G$  (régions).

## II.9.6 Normalisation de GOG

Pour des caractéristiques de grande dimension, la normalisation est un facteur important pour améliorer leurs performances [79]. Étant donné que le descripteur GOG est de grande dimension, nous normalisons le descripteur en utilisant la normalisation de la norme L2, qui est la normalisation la plus largement adoptée. Nous avons observé l'existence de dimensions qui ont généralement des valeurs élevées/petites communes entre différentes images dans le descripteur GOG. Cela est dû à l'utilisation de caractéristiques de pixels ayant différentes propriétés de distribution, par exemple, la magnitude du gradient se distribue de manière clairsemée dans les images, et l'intensité de couleur se distribue de manière plus uniforme. Dans de tels cas, la distance cosinus, c'est-à-dire la distance euclidienne après la normalisation, serait dominée par les dimensions biaisées. Pour remédier à de telles dimensions biaisées, nous supprimons le vecteur moyen des échantillons d'entraînement avant de normaliser le vecteur de caractéristiques. La normalisation de GOG devient donc comme suit :

$$z = (z - \bar{z}) / \|z - \bar{z}\|_2, \quad (6)$$

où  $z$  est la moyenne des descripteurs GOG. Pour le descripteur de fusion, nous normalisons chacun des descripteurs GOG extraits sur quatre espaces de couleur avant de les concaténer. Pour la représentation Bag-of-Words, une normalisation similaire est proposée pour refléter les mots manquants en commun pour la similarité cosinus. En revanche, nous l'utilisons pour remédier à l'effet des dimensions biaisées

## II.10. Analyse de Mahalanobis

La méthode de Mahalanobis est une approche d'analyse de données visant à mesurer les distances entre les échantillons dans un espace multidimensionnel. Son objectif principal est de déterminer une matrice  $M$  de manière à décrire efficacement les distances entre les échantillons. La distance entre deux échantillons  $x_i$  et  $x_j$  peut être définie par l'équation:

$$d_M(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j)$$

Où la matrice  $M$  est définie de manière à être positive semi-définie.

L'objectif principal de l'utilisation de Mahalanobis est d'estimer les distances entre les échantillons de manière appropriée, ce qui permet de mieux comprendre et analyser les données. En estimant la matrice  $M$ , il est possible d'améliorer la classification et la discrimination entre les échantillons dans l'espace de données multidimensionnel.

## II.11. Conclusion

Dans ce chapitre, nous avons examiné les approches proposées pour la Re-ID des personnes, à savoir "LOMO" et "GOG", ainsi que leurs avantages et leur fonctionnement pour améliorer le taux de reconnaissance correct des personnes en tenant compte des facteurs externes tels que l'éclairage, les angles de prise de vue variés et le bruit. Nous avons également exploré le principe de fonctionnement de l'approche d'apprentissage métrique (XQDA) et l'utilisation de la distance de Mahalanobis.

# **CHAPITRE 3**

## Implémentations et Résultats

### III.1. Introduction

Dans ce dernier chapitre, nous présenterons la plate forme logicielle que nous avons implémentée sous l’environnement MATLAB®2022, qui montre le principe du système complet des Re-ID du personne, ainsi qu’une explication des différentes tâches que nous avons réalisées dans le cadre de ce travail depuis le prétraitement, jusqu’à la recherche dans une base de données, en passant par la codification.

Dans ce chapitre nous allons présenter les différentes étapes de la conception de notre système. Nous développons en suite la partie d’extraction de paramètre.

Finalement, on utilise biométrie douce (SB) avec Re-ID à partir de ces deux technique (LOMO & GOG) de conversion étudié est appliqué sur les la base de données VIPER[93], et les résultats obtenus sont exposés.

### III.2. Base de données

Dans notre étude, nous avons exploité l’ensemble de données VIPeR[93], spécialement conçu pour reconnaître les individus indépendamment de l’angle de vue. Cet ensemble contient 1 264 images de 632 piétons, capturées par deux caméras sous différents angles, poses et éclairages. Toutes les images sont normalisées à une taille de 128\*48 pixels. Où l’angle de vue de la première caméra varie de 0 à 90 degrés, tandis que celui de la seconde caméra varie de 90 à 180 degrés. Chaque individu est représenté par une seule image de chaque caméra. Lors du test, chaque image capturée par la première caméra est comparée à celle de la seconde caméra. La Figure (III.1) présente un échantillon d’images de cet ensemble de données.



Fig.III.1: Quelques exemples de l’ensemble de données VIPeR.

### III.3 Approche proposée

La figure2 montre le diagramme schématique de l’approche de RE-ID proposée. Les attributs de biométrie douce (SB) sont d’abord détectés à partir d’une image de sonde en utilisant les classificateurs d’attribut formés. À l’étape suivante, les distances SB sont calculées entre la sonde et les images de la galerie. À l’étape suivante, calcule ré-identification de personne

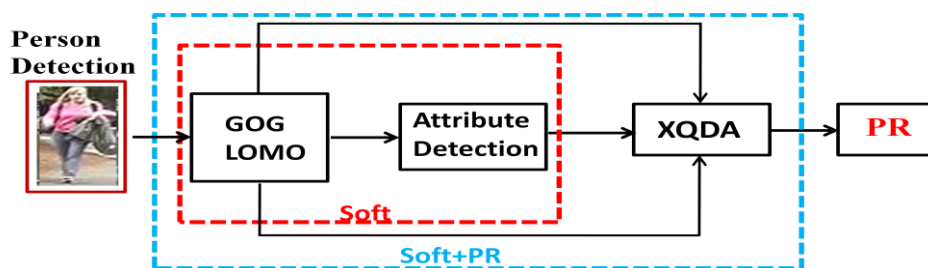


Fig.III.2. Schéma diagrammatique de l’approche proposée

**III.4. Attributs**

Nous avons défini l'espace suivant de  $N_a=5$  attributs binaires: *Back pack*, *jeans*, *Carrying*, *redshirt* et *Male*. Ce sont les cinq caractéristiques de SB, et chaque attribut est illustré dans la Figure(III.3).



**Fig.III.3:** Back pack, Jeans, Carrying, Redshirt et Male sont cinq attributs SB utilisés dans ce travail.

**III.5. Détection des attributs SB**

Jeans, Redshirt, Carrying, Back pack et Male, sont les cinq attributs (SB) avec des annotations binaires. Les auteurs dans [19], donnent la précision de ces attributs en utilisant la base de donnée VIPeR, qui comprend deux vues de caméra (CAM A et CAM B). Le Tableau (III.1), illustre les attributs SB utilisés dans ce travail et leurs numéros associés dans la base de données VIPeR, qui comprend 632 sujets. La détection des attributs est subdivisée en un problème de classification binaire. Les étiquettes binaires 1 et 0 sont associées à chaque attribut. Une valeur de 1 pour un attribut signale l'existence de cet attribut. Nous utilisons RF "Random Forest" pour entraîner les classificateurs d'attributs. Divers classificateurs sont entraînés pour différents attributs

**Table.III.1:** Soft biometrics Attributes

Attributes	Back pack	Jeans	Carrying	Redshirt	Male
# subjects	229	221	173	65	309

Sélection de parties Nous effectuons une sélection de caractéristiques pour choisir les parties spatiales (bandes) utilisées pour détecter toutes les attributs. La figure (III.4), illustre les parties largement intuitives liées à certains attributs. Ce résultat montre qu'une sélection automatique des caractéristiques pour les attributs spatialement localisés est possible. De manière importante, réduire l'espace des caractéristiques pour chaque attribut SB aide à atténuer le défi de la tendance à l'information (données) éparses pour tous les attributs.



**Fig.III.4:** Un exemple de résultats de sélection de bandes pour la détection d'attributs dans VIPeR.

### III.6. Détection d'attributs

La précision des classificateurs reste limitée, certaines caractéristiques ayant une précision de classification relativement faible. Par exemple, la détection de "Carrying" (port d'objets) a une précision de 60,96%, tandis que la détection des "jeans" atteint une précision de 74,85%. Ces résultats mettent en évidence l'importance de la sélection et de la pondération individuelle des caractéristiques, ainsi que les différences inhérentes entre elles. Il est intéressant de noter que certaines caractéristiques biométriques qui semblent visuellement subtiles peuvent néanmoins être identifiées avec une précision notable. Par exemple, la détection du "male" (genre masculin) a une précision de 49,03%, malgré la petite taille des images faciales.

Les détecteurs de performance inférieure ont tendance à être plus sensibles aux variations de pose. Par exemple, les "necks" (cou) ne peuvent être détectés que de face. De même, les "logos" (logos) sont généralement plus petits et moins visibles de l'arrière et des côtés en raison des petites régions distinctes impliquées, de l'arrière-plan encombré et de l'éclairage variable. De plus, la détection du "male" est une caractéristique difficile à discerner à distance, surtout avec une faible résolution, où le niveau de bruit de l'image peut être significatif.

En revanche, certaines caractéristiques comme les "jeans" sont parmi les plus détectées et les plus courantes dans la base de données, ce qui explique leur poids élevé dans les résultats globaux. Tandis que les caractéristiques comme le "back pack" (sac à dos) ont une précision de 63,60%, les "redshirt" atteignent une précision maximale de 81,78%. Ces disparités montrent clairement que certaines caractéristiques visuelles sont plus faciles à identifier que d'autres, en fonction de leur visibilité et de leur fréquence dans la base de données.

En somme, ces observations soulignent l'importance cruciale de la sélection minutieuse et de la pondération des caractéristiques dans les systèmes de reconnaissance. La capacité à détecter certaines caractéristiques de manière fiable, même lorsque d'autres sont plus difficiles à discerner, peut améliorer considérablement la précision globale du système.

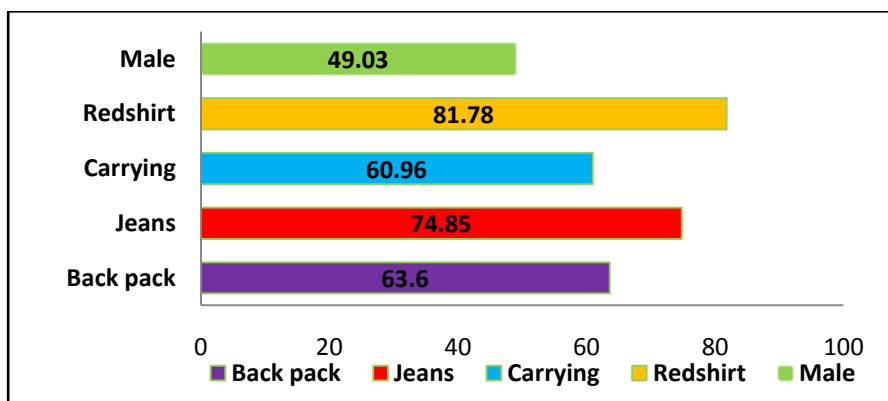


Fig.III.5: Résultats soft biometrics 5ATTRIBU

### III.7. Résultats sans biométrie douce (soft biometrics)

	Rank1%	Rank5%	Rank10%	Rank15%	Rank20%	XQDA Dims
LOMO+XQDA	39.81	67.88	80.32	87.31	91.20	270
GOG+XQDA	34.05	63.04	75.38	81.65	86.01	80

### III.8. Résultats avec biométrie douce (soft biometrics)

#### III.8.1 Descripteur LOMO avec SB

Les résultats sont montrés dans le tableau (III.2), le descripteur LOMO avec l'apprentissage de la métrique de distance XQDA obtient les meilleurs résultats, avec une précision de Rank1 pour le sac à dos (back pack) de 42,91% (XQDA=70). Deuxièmement, la meilleure performance de l'algorithme LOMO pour le taux de reconnaissance en Rank1 pour les jeans est de 44,34% (XQDA=100). Troisièmement, la précision de Rank1 pour le Carrying est d'environ 40,22% (XQDA=150). Quatrièmement, la meilleure performance de l'algorithme LOMO pour le taux de Rank1 pour les chemises (red shirt) est de 35,03% (XQDA=140). Cinquièmement, la meilleure performance pour les hommes (male) a atteint une précision de Re-ID de 41,46% en Rank1 pour LOMO (XQDA=400).

**Tableau.III.2:** Résultats de la Re-ID des personnes sur de données VIPeR(LOMO)

Ranks	Rank1	Rank5	Rank10	Rank15	Rank20	XQda Dims
<b>Back pack</b>	42.91	75.57	85.57	89.97	92.63	70
<b>Jeans</b>	44.34	76.77	87.37	91.42	93.70	100
<b>Carrying</b>	40.22	73.10	83.54	88.10	91.08	150
<b>Read shirt</b>	35.03	66.68	77.85	83.80	87.63	140
<b>Male</b>	41.46	67.09	79.11	84.18	86.71	400

#### III.8.2 Descripteur GOG avec SB

Les résultats sont montrés dans le Tableau (III.3) Le descripteur GOG avec l'apprentissage de la métrique de distance XQDA obtient les meilleurs résultats, avec une précision de Rank1 pour le Back Pack de 41,74 % (XQDA dims =100). Deuxièmement, la précision de Rank1 pour les jeans est de 39,97 % (XQDA dims=96). Troisièmement, la précision de Rank1 pour le Carrying 40,79 % (XQDA=80). Quatrièmement, la meilleure performance de l'algorithme GOG pour le taux de Rank1 pour les Red shirt est de 35,03 % (XQDA dims =33). Cinquièmement, la meilleure performance de l'algorithme GOG pour le Rank1, pour le Male est de 43,13 % (XQDA dims =140).

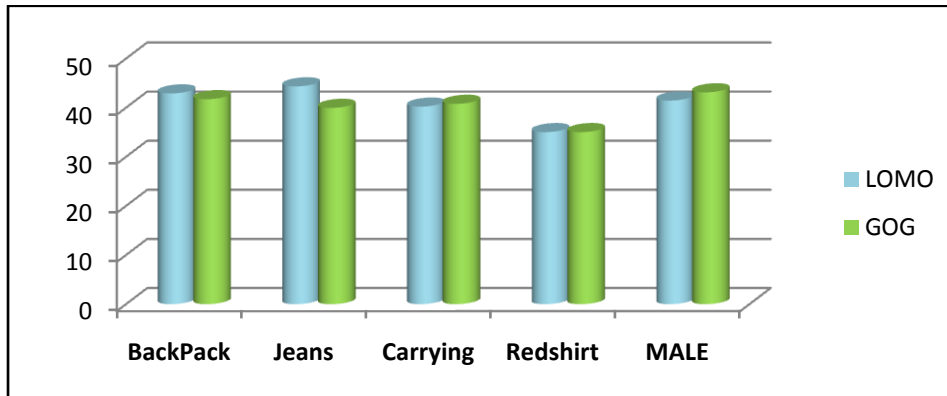
**Tableau.III.3:** Résultats de la Re-ID des personnes sur de données VIPeR(GOG)

Ranks	Rank1	Rank5	Rank10	Rank15	Rank20	XQda Dims
<b>Back pack</b>	41.74	72.91	84.05	89.05	91.77	100
<b>Jeans</b>	39.97	71.99	82.34	87.47	90.54	96
<b>Carrying</b>	40.79	70.19	81.08	85.57	88.83	80
<b>Read shirt</b>	35.03	66.68	77.85	83.80	87.63	33
<b>Male</b>	43.13	73.86	84.56	89.37	92.37	140

### III.9. Comparaison des résultats entre LOMO et GOG

Les données présentées suggèrent que le descripteur LOMO surpasse GOG dans certaines catégories telles que les jeans, avec un taux de Re-ID de classification au premier Rank1 de 44,34% contre 39,97% pour GOG. Cependant, GOG semble obtenir de meilleures performances dans d'autres catégories telles que le Carrying et le Male, avec des taux de précision de classification au premier Rank de 40,79 % et 43,13 % respectivement, contre 40,22 % et 41,46 % pour LOMO.

Cette comparaison montre que les performances de LOMO et de GOG dépendent de la catégorie de classification spécifique, chacun pouvant surpasser l'autre dans une catégorie donnée en fonction du contexte et des données disponibles.



**Fig.III.6:** Comparaison des résultats entre LOMO et GOG.

### III.10. Conclusion

Dans ce chapitre, nous avons présenté une méthode efficace pour la ré-identification des personnes. Nous avons utilisé des descripteurs efficaces appelés LOMO et GOG, ainsi qu'une approche d'apprentissage métrique appelée XQDA pour améliorer les performances du système de ré-ID des personnes.

L'étude expérimentale a montré l'obtention de bons résultats de simulations et des taux de Re-ID dans les différentes techniques employées. Le meilleur taux de reconnaissance obtenu en utilisant LOMO+XQDA avec SB est Jeans pour un 44.34%(XQDA dims =100), et la méthode GOG+XQDA avec SB est Male pour un 43.13%(XQDA dims =140) pour Re-ID sur la base de données VIPeR.

Les systèmes biométriques douce existent déjà depuis quelques années mais ils sont plutôt adaptés à un niveau de sécurité élevé. En fait, plus le niveau de sécurité est élevé.

### Conclusion Générale

En conclusion, notre étude a examiné en détail les méthodes de Re-ID des personnes en utilisant les caractéristiques LOMO et GOG. Nous avons constaté que ces deux approches offrent des avantages distincts dans le domaine de la vision par ordinateur et de l'analyse d'images. LOMO se concentre sur l'extraction des caractéristiques saillantes des personnes dans une image, tandis que GOG vise à améliorer la qualité de l'image et à réduire le bruit. Nos expériences et analyses ont révélé que l'utilisation conjointe de ces deux techniques peut conduire à des résultats améliorés en termes de précision et de robustesse dans les systèmes de surveillance et de sécurité.

Cependant, malgré les progrès réalisés, il reste encore des défis à relever dans le domaine de la Re-ID des personnes. Des questions telles que la variabilité des conditions d'éclairage, les occlusions et les changements de pose continuent de poser des défis importants pour les systèmes de reconnaissance des personnes. Par conséquent, des recherches futures sont nécessaires pour améliorer la performance des algorithmes existants et développer de nouvelles approches innovantes pour surmonter ces obstacles.

En définitive, notre travail contribue à enrichir la compréhension des techniques de Re-ID des personnes et ouvre la voie à de nouvelles possibilités dans le domaine de la vision par ordinateur et de la sécurité. Il est essentiel de poursuivre les efforts de recherche et de développement pour faire progresser ces technologies et les rendre plus efficaces et plus fiables dans une variété d'applications pratiques.

Les **perspectives** de ce travail sont nombreuses:

- Il est aussi intéressant de constituer une large base de données, comprenant un grand nombre d'individus et davantage d'images par individu;
- La méthode CNN peut être développée d'améliorer la Re-ID;
- Fusion les deux méthodes LOMO et GOG.

### Bibliographie

- [1] Elaine R Berman, « Biochemistry of the Eye », Springer, 1991.
- [2] John G. Daugman, « High confidence visual recognition of persons by a test of statistical independence », IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 15, N° 11, pp. 1148-1160, November 1993.
- [4] Dalal, N. et Triggs, B. (2005). Histograms of oriented gradients for human detection. Dans Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, volume 1, pages 886–893.
- [5] Kalman, R. E. (1960). A new approach to linear filtering and prediction problems. Transactions of the ASME–Journal of Basic Engineering, 82(Series D):35–45.
- [6] G. Roethenbaugh. “An Introduction to Biometrics and General History”, Biometrics Explained, Section 1, 1998.
- [7] G .Olivier. Morphologie et types humains. Vigot Frères, 1961, 171 pages
- [8] A.K. Jain, P. Flynn, A. Ross, Handbook of Biometrics, Springer, New York, USA, 2008.
- [9] A.Ross, K.Nandakumar, and A.K. Jain, Handbook of Multibiometrics, Springer Heidelberg edition, New York, USA, 2006.
- [10] F. Karray, J. A. Saleh, M. N. Arab and M. Alemzadeh, Multi Modal Biometric Systems: A State of the Art Survey, Fourth International Conference on Computational Intelligence, Robotics and Autonomous Systems, New Zealand, Nov. 2007.
- [11] A.K. Jain, R. Bolle and S. Pankanti, BIOMETRICS: Personal Identification in Networked society, Kluwer Academic Publishers, 1999.
- [12] Biometric Technology Application Manual, Volume1, National Biometric Security Project, 2008.
- [13] J. Ashbourn, Guide to Biometrics for Large-Scale Systems, Technological, Operational, and User-Related Factors, Springer-Verlag, London, 2011.
- [14] Gong, S., The ré-identification challenge, in Person ré-identification. 2014, Springer. p. 1-20.
- [15] Zheng, L., Y. Yang, and A.G. Hauptmann, Person ré-identification: Past, present and future. arXiv preprint arXiv:1610.02984, 2016.
- [16] Gheissari, N., T.B. Sebastian, and R. Hartley. Person ré-identification using spatiotemporal appearance. in 2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). 2006. IEEE.
- [17] Gray, D. and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. in European conference on computer vision. 2008. Springer.
- [18] Zheng, L. Scalable person ré-identification: A benchmark. in Proceedings of the IEEE international conference on computer vision. 2015.
- [19] Wang, T. Person ré-identification by video ranking. in European conference on computer vision. 2014. Springer.
- [20] Zheng, L. Mars: A video benchmark for large-scale person ré-identification. in European Conference on Computer Vision. 2016. Springer.
- [21] Ye, M. Specific person retrieval via incomplete text description. in Proceedings of the 5th ACM on International Conference on Multimedia Retrieval. 2015.
- [22] Li, S. Identity-aware textual-visual matching with latent co-attention. in Proceedings of the IEEE International Conference on Computer Vision. 2017.
- [23] Ye, M., Deep learning for person ré-identification: A survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

- [24] M,IBN KHEDHER. ré-identification de personnes a partir des sequences video. .2014.
- [25] M, DEMRI. Multimodal Biometric Fusion Using Evolutionary Techniques. Juin 2012
- [26] L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and ré-identification. *Computer Vision and Image Understanding*, 117(2):130–144, 2013
- [27] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon. Information-theoretic metric learning. In *Proceedings of the 24th international conference on Machine learning*, pages 209–216. ACM, 2007.
- [29] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M.Cristani. Person ré-identification by symmetry-drive naccumulation of local features. In *CVPR*, pages 2360–2367, 2010.
- [30] N.Gheissari, T.B.Sebastian, and R. Hartley. Person ré-identification using spatiotemporal appearance. In *CVPR(2)*, pages 1528–1535, 2006.
- [31] S. Gong, M. Cristani, S. Yan, and C. C. Loy. *Person Ré-identification*. Springer, 2014.
- [32] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *European Conference on Computer Vision*, 2008.
- [34] M. Kostinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof. Large scale metric learning from equivalence constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012.
- [37] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person ré-identification. In *European Conference on Computer Vision Workshops*, 2012.
- [42] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *IEEE International workshop on performance evaluation of tracking and surveillance*, 2007.
- [43] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, 1997.
- [44] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. Properties and performance of a center/surround retinex. *Image Processing, IEEE Transactions on*, 6(3):451–462, 1997.
- [45] E. H. Land and J. McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971.
- [46] S. Liao, D. Yi, Z. Lei, R. Qin, and S. Z. Li. Heterogeneous face recognition from local structures of normalized appearance. In *International Conference on Biometrics*, 2009.
- [47] C. Liu, S. Gong, C. C. Loy, and X. Lin. Person ré-identification: what features are important? In *Computer Vision–ECCV 2012. Workshops and Demonstrations*, pages 391–401. Springer, 2012.
- [48] B. Moghaddam, T. Jebara, and A. Pentland. Bayesian face recognition. *Pattern Recognition*, 33(11):1771–1782, 2000.
- [49] S. Pedagadi, J. Orwell, S. Velastin, and B.Boghossian. Local fisher discriminant analysis for pedestrian ré-identification. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2013.
- [50] B. Prosser, W.-S. Zheng, S. Gong, T. Xiang, and Q. Mary. Person ré-identification by support vector ranking. In *BMVC*, 2010.
- [51] W.-S. Zheng, S.Gong, and T.Xiang. Person ré-identification by probabilistic relative distance comparison. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, pages 649–656. IEEE, 2011.

- [53] S. Bak, G. Charpiat, E. Corvée, F. Brémont, and M. Thonnat. Learning to match appearances by correlations in a covariance metric space. In *European Conference on Computer Vision (ECCV)*, volume 7574, pages 806–820, 2012.
- [54] S. Bak, E. Corvée, F. Brémont, and M. Thonnat. Boosted human ré-identification using Riemannian manifolds. *Image Vision Computing*, 30(6-7):443–452, 2012.
- [55] S. Gong, M. Cristani, S. Yan, and C. C. Loy. *Person Ré-identification*. *Advances in Computer Vision and Pattern Recognition*, Springer, 2014.
- [56] B. Ma, Q. Li, and H. Chang. Gaussian descriptor based on local features for person ré-identification. In *Asian Conference on Computer Vision (ACCV) Workshop*, pages 505–518, 2014.
- [57] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by Fisher vectors for person ré-identification. In *European Conference on Computer Vision (ECCV) Workshop*, pages 413–422, 2012.
- [58] B. Ma, Y. Su, and F. Jurie. Covariance descriptor based on bio-inspired features for person ré-identification and face verification. *Image and Vision Computing*, 32(6):379–390, 2014.
- [59] P. M. Roth, M. Hirzer, M. Köstinger, C. Belezni, and H. Bischof. Mahalanobis distance learning for person ré-identification. In *Person Ré-identification*, pages 247–267, 2014.
- [60] F. Xiong, M. Gou, O. Camps, and M. Szaier. Person ré-identification using kernel-based metric learning methods. In *European Conference on Computer Vision (ECCV)*, pages 1–16, 2014.
- [61] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person ré-identification. In *European Conference on Computer Vision (ECCV)*, pages 536–551, 2014.
- [62] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person ré-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3586–3593, 2013.
- [63] R. Zhao, W. Ouyang, and X. Wang. Learning mid-level filters for person ré-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 144–151, 2014.
- [64] W. Zheng, S. Gong, and T. Xiang. Ré-identification by relative distance comparison. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(3):653–668, 2013.
- [65] S. Amari and H. Nagaoka. *Methods of Information Geometry*. volume 191 of *Translations of mathematical monographs*. American Mathematical Society, 2001.
- [66] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM J. Matrix Analysis Applications*, 29(1):328–347, 2006.
- [67] L. Bazzani, M. Cristani, and V. Murino. Symmetry-driven accumulation of local features for human characterization and ré-identification. *Computer Vision and Image Understanding*, 117(2):130–144, 2013.
- [68] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Freeform region description with second-order pooling. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(6):1177–1189, 2015.
- [69] L. Gong, T. Wang, and F. Liu. Shape of Gaussians as feature descriptors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2366–2371, 2009.
- [70] T. Kobayashi and N. Otsu. Image feature extraction using gradient local auto-correlations. In *European Conference on Computer Vision (ECCV)*, pages 346–358, 2008.
- [71] R. Layne, T. M. Hospedales, and S. Gong. Person ré-identification by attributes. In *British Machine Vision Conference, (BMVC)*, pages 1–11, 2012.

- [72] P. Li and Q. Wang. Local log-Euclidean covariance matrix (L2ECM) for image representation and its applications. In *European Conference on Computer Vision (ECCV)*, pages 469–482, 2012.
- [73] P. Li, Q. Wang, and L. Zhang. A novel earth mover’s distance methodology for image matching with Gaussian mixture models. In *IEEE International Conference on Computer Vision (ICCV)*, pages 1689–1696, 2013.
- [74] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3610–3617, 2013.
- [75] S. Liao, Y. Hu, X. Zhu, and S. Z. Li. Person ré-identification by local maximal occurrence representation and metric learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2197–2206, 2015.
- [76] M. Lovrić, M. Min-Oo, and E. A. Ruh. Multivariate normal distributions parametrized as a Riemannian symmetric space. *Journal of Multivariate Analysis*, 74(1):36–48, 2000.
- [77] H. Nakayama, T. Harada, and Y. Kuniyoshi. Global Gaussian approach for scene categorization using information geometry. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2336–2343, 2010.
- [78] S. Pedagadi, J. Orwell, S. A. Velastin, and B. A. Boghossian. Local Fisher discriminant analysis for pedestrian ré-identification. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3318–3325, 2013.
- [79] J. Sánchez, F. Perronnin, T. Mensink, and J. J. Verbeek. Image classification with the Fisher vector: Theory and practice. *International Journal of Computer Vision*, 105(3):222–245, 2013.
- [80] R. Satta. Appearance descriptors for person ré-identification: a comprehensive review. *CoRR*, abs/1307.5748, 2013.
- [81] G. Serra, C. Grana, M. Manfredi, and R. Cucchiara. Covariance of covariance features for image classification. In *Proceedings of International Conference on Multimedia Retrieval (ICMR)*, page 411, 2014.
- [82] G. Serra, C. Grana, M. Manfredi, and R. Cucchiara. GOLD: Gaussians of local descriptors for image representation. *Computer Vision and Image Understanding*, 134:22–32, 2015.
- [83] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *European Conference on Computer Vision (ECCV)*, pages 589–600, 2006.
- [84] O. Tuzel, F. Porikli, and P. Meer. Pedestrian detection via classification on Riemannian manifolds. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(10):1713–1727, 2008.
- [85] M. Zeng, Z. Wu, C. Tian, L. Zhang, and L. Hu. Efficient person ré-identification by hybrid spatiogram and covariance descriptor. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 48–56, 2015.
- [86] T. Matsukawa, T. Okabe, E. Suzuki, and Y. Sato. Hierarchical Gaussian Descriptor for Person Ré-identification. *IEEE Xplore on Computer Vision and Pattern Recognition (CVPR)*, 2016
- [87] Chen, Y., *Person re-identification in images with deep learning*, 2018, Lyon.
- [88] Wang, X. and R. Zhao, *Person re-identification: System design and evaluation*
- [89] Nie, W., *Single/cross-camera multiple-person tracking by graph matching*. *Neurocomputing*, 2014. 139: p. 220-232.

- [90] Henriques, J.F. Exploiting the circulant structure of tracking-by-detection with kernels. in European conference on computer vision. 2012. Springer.
- [91] Suma, V., Computer vision for human-machine interaction-review. Journal of trends in Computer Science and Smart technology (TCSST), 2019. 1(02): p. 131-139.
- [92] S. Liao, Y.Hu, X. Zhu, and S.Z. Li. "Person Re-identification by Local Maximal Occurrence Representation and Metric Learning". IEEE Xplore on Computer Vision and Pattern Recognition (CVPR 2015), June 2015.
- [93] Gray, D., S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. in Proc. IEEE international workshop on performance evaluation for tracking and surveillance (PETS). 2007. Citeseer.