



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

Ré-identification du personne par descripteur
LOMO et GOG

Présenté par:

YOUSRA NAFTI

ROUMAÏSSA BEN NADJI

Supervisé par:

Dr : Fouaze MOUSSI

Devant le jury composé de :

Dr Fouaze MOUSSI

Superviseur

Dr.

Président

Dr.

Examineur

Année universitaire: 2023 /2024

Remerciement

Loué soit Dieu, qui nous a éclairés sur le chemin de la connaissance et de la connaissance, nous a aidés à remplir ce devoir, et nous a aidés à accomplir ce travail.

*J'adresse mes sincères remerciements, mon appréciation et ma grande gratitude au vertueux **MOUSSI FOUAZE** pour toutes les directions et recommandations, alors que Dieu le récompense en mon nom avec toute la bonté et le mette dans la balance de ses bonnes actions.*

*J'adresse également mes sincères remerciements et **TOUS** Les professeurs, membres du juré de discussion, pour l'honneur d'accepter l'appréciation de ce travail.*

Dédicace

Avec joie et fierté, je dédie ce travail à :

À Dieu qui m'a donné la patience pour accomplir ce travail.

À ma mère au cœur tendre, qui m'a aidée par ses prières.

À mon cher père, qui m'a toujours soutenue.

*À mon cher époux Moussa, qui a poursuivi avec moi mon
parcours académique.*

À toute ma famille, en particulier à ma sœur Anfal.

À ma nouvelle famille SOUFIA.

À mes chères amies.

Yousra

Dédicace

Nous n'avons entrepris nos débuts que par Sa guidance, nous n'avons atteint nos objectifs que par Sa grâce, et nous n'avons réalisé nos buts que par Sa faveur. Louange à Dieu qui nous a permis de franchir cette étape dans notre parcours académique.

Je dédie le fruit de mon succès à : ma mère, mon père et tous les membres de ma famille.

Roumaïssa

Résumé

La ré-identification des personnes (Re-ID) est une tâche importante dans le domaine de la sécurité intelligente. Il s'agit du problème d'identification des personnes sur des images capturées par différentes caméras de surveillance sans chevauchement des champs de vision. Dans un environnement réel, Re-ID est une tâche difficile en raison des variations des conditions d'éclairage, des angles de vision, de la pose et des occlusions. Deux problèmes fondamentaux sont essentiels pour la ré-identification des personnes, la représentation des caractéristiques et l'apprentissage métrique. Une représentation efficace des caractéristiques doit être robuste à l'éclairage et aux changements de point de vue, et une métrique discriminante doit être apprise pour correspondre à diverses images de personnes. Dans ce travail, Nous utiliserons une représentation efficace des caractéristiques appelée Occurrence Maximale Locale (LOMO) et le Descripteur Gaussien Hiérarchique (GOG), ainsi qu'une méthode d'apprentissage de sous-espace et d'apprentissage métrique appelée Analyse Discriminante Quadratique à Vues Croisées (XQDA) avec la distance de Mahalanobis pour améliorer les résultats de classification.

Les résultats expérimentaux obtenus (LOMO+XQDA=40,06%) et (GOG+XQDA=34,81%) sont efficaces pour les deux méthodes pour améliorer la capacité de ré-identification des personnes à l'aide de la base de données VIPeR.

Mots clés : Re-ID, LOMO, GOG, XQDA, Mahalanobis.

Abstract

Re-identification of individuals (Re-ID) is an important task in the field of intelligent security. It involves identifying individuals in images captured by different surveillance cameras with non-overlapping fields of view. In a real-world environment, Re-ID is challenging due to variations in lighting conditions, viewing angles, pose, and occlusions. Two fundamental problems are essential for person re-identification: feature representation and metric learning. An effective feature representation must be robust to lighting and viewpoint changes, and a discriminative metric must be learned to match various person images. In this work, we will use an efficient feature representation called Local Maximal Occurrence (LOMO) and the Hierarchical Gaussian Descriptor (GOG), along with a subspace learning and metric learning method called Cross-View Quadratic Discriminant Analysis (XQDA) with Mahalanobis distance to improve classification results.

The experimental results obtained (LOMO+XQDA=40.06%) and (GOG +XQDA =34.81%) are effective for the two methods in improving the ability to re-identify people using the VIPeR database.

Keywords: Re-ID, LOMO, GOG, XQDA, and Mahalanobis.

ملخص

إعادة تعريف الأفراد (Re-ID) ، هي مهمة مهمة في مجال الأمن الذكي. وتتضمن تحديد الأفراد في الصور التي التقطتها كاميرات المراقبة المختلفة مع حقول رؤية غير متداخلة. في بيئة العمل الحقيقية، تعتبر إعادة التعريف تحديًا بسبب التغيرات في ظروف الإضاءة وزوايا الرؤية والوضع والتعتيم. يعتبر مشكلتين أساسيتين ضروريتين لإعادة تعريف الأفراد: تمثيل الميزات وتعلم المقاييس. يجب أن يكون تمثيل الميزات فعالاً وقويًا ضد التغيرات في الإضاءة وزوايا الرؤية، ويجب أن يتم تعلم مقياس تمييزي لمطابقة مختلف صور الأفراد. في هذا العمل، سوف نستخدم تمثيلًا فعالًا للميزات يسمى الظهور المحلي الأقصى (LOMO) وواصف الهرم الغاوسي (GOG)، بالإضافة إلى طريقة تعلم الفضاء الفرعي وتعلم المقياس المسماة التحليل التمييزي التربيعة عبر العروض (XQDA) مع مسافة Mahalanobis لتحسين نتائج التصنيف.

النتائج التجريبية المتحصل عليها ($LOMO+XQDA=40.06\%$) و($GOG+XQDA=34.18\%$) فعالة للطريقتين في تحسين قدرة إعادة التعرف على هوية الأشخاص بإستعمال قاعدة البيانات VIPeR. الكلمات الرئيسية: إعادة التعرف (Re-ID) ، LOMO ، GOG ، XQDA ، Mahalanobis .

Liste des abréviations

- LOMO:** Occurrence maximale locale
- GOG:** Descripteur gaussien hiérarchique.
- XQDA:** Analyse Discriminante Quadratique entre Vues Croisées
- Re-ID:** Ré-Identification
- CMC:** Courbes des Caractéristiques de Correspondance Cumulative
- QDA :** Quadratic Discriminant Analysis
- SPD:** Symmetric Positive Definite
- LBP :** Motifs binaires locaux
- GMM :** Modèle de Mélange Gaussien
- SILTP :** Descripteur de modèle ternaire local invariant
- CNN :** Réseaux de neurones convolutionnels.
- HSV :** Histogrammes de couleurs.

Liste des figures

Chapitre 1: Biometrie et person ridentification

Figure 1.1: Capture de l'image d'un Reconnaisance faciale.....	5
Figure 1.2: Capture de l'image d'un iris	6
Figure 1.3: Capture de l'image d'une empreinte digitale	7
Figure 1.4: Capture d'une signature	7
Figure 1.5: Capture de Vérification vocal	8
Figure 1.6: Capture de la démarche.....	8
Figure 1.7: Capture d'une Oreille.....	9
Figure 1.8: Scan de la forme de la main	9
Figure 1.9: Pipeline d'un système de ré-identification de personne de bout en bout	12
Figure 1.10: Exemple de quelques défis de ré-identification de personnes. Chaque paire d'images montre la même personne sauf (g). (a) Variation du point de vue, (b) variation de la pose, (c) changements d'éclairage, (d) occlusion partielle, (e) détection de piétons inexacte, (f) changement d'accessoires, (g) basse résolution et (h) personnes différentes avec des vêtements similaires.....	15

Chapitre 2 : Les méthodes de Re-ID

Figure (2.1): Descripteur GOG	22
Figure 2.2: (a) Exemples de paires d'images de la base de données VIPeR; (b) Images traitées par Retinex. Images dans le même La colonne représente la même personne.....	28
Figure 2.3: Illustration de la méthode d'extraction de fonctionnalités LOMO.....	29
Figure 2.4. Courbes CMC comparant la fonction proposée avec et sans Retinex et l'opération d'occurrence maximale locale ((a) Cosinus et (b) XQDA). (c) Précision au rang 1 avec des dimensions de sous-espaces variables pour l'algorithme XQDA avec la fonction LOMO.	30
Figure 2.5: Distributions of ΩI and ΩE in one projected dimension.	32

Chapitre 3 : Résultat et discussion

Figure 3 .1: Paires d'images d'échantillon du jeu de données VIPeR.....	35
Figure 3.2 Exemple d'une courbe CMC.....	36
Figure3.3 : Person re-id consiste à rechercher dans la galerie une correspondance correcte avec la sonde donnée	37
Figure 3.4 courbe CMC de LOMO+XQDA	39
Figure 3.5 courbe CMC de GOG+XQDA.....	40
Figure 3.6 Diagramme des résultats de la comparaison entre "LOMO+ XQDA" et "GOG" pour Rank-1(%), sur la base de données VIPeR.	41

Liste de tableau :

Chapitre 1: Biometrie et person ridentification

Tableau 1.1: Exemple de différents ensembles de données et de leurs caractéristiques.....	16
--	----

Chapitre 3 : Résultat et discussion

Tableau 3.1: courbe CMC de LOMO+XQDA on $r=1, 5, 10, 15, 20$. des bases de données utilisées Viper	38
Tableau 3.2 : courbe CMC de GOG+XQDA on $r=1, 5, 10, 15, 20$. des bases de données utilisées Viper	39
Tableau 3.3 : comparaison de "LOMO+ XQDA" et "GOG" des bases de données utilisées Viper	40

Tables des matières

Remerciement	i
Dédicace.....	ii
Dédicace.....	iii
Résumé.....	iv
Abstract	v
ملخص	vi
Liste des abréviations	vii
Liste des figures	viii
Liste de tableau :.....	ix
Tables des matières.....	x
Introduction générale.....	1

Chapitre 1 : Biometrie et person re-identification

1.1. Introduction.....	4
1.2. Biometrie	4
1.3. Présentation des techniques biométriques	4
1.3.1 Reconnaissance faciale.....	4
1.3.2 Reconnaissance de l'iris	6
1.3.3 Empreinte digitale	6
1.3.4 Numérisation de la signature.....	7
1.3.5 . Vérification vocale	7
1.3.6 Reconnaissance de la démarche	8
1.3.7. Oreille.....	8
1.3.8 Géométrie de la main	9
1.4. Applications des systèmes biométriques	9
1.5. Ré-identification des personnes	10
1.6. Re-identification pipeline	11
1.7. Challenges.....	12
1.8. Bases des données de référence	15
1.9. Applications de la ré-identification des personnes	16
1.10. Conclusion	17

Chapitre 2 : Les méthodes de Re-ID

2.1. Introduction.....	20
------------------------	----

2.2 Représentations des caractéristiques pour la ré-identification des personnes	20
2.2.1 Apprentissage métrique pour la ré-identification des personnes	21
2.3. Descripteur GOG	21
2.3.1. Caractéristiques des pixels	22
2.3.2 Gaussiennes de région.....	23
2.3.3 Mappage dans l'espace tangent et demi-vectorisation	24
2.3.4 Gaussienne de région dans l'espace tangent.....	25
2.3.5 Descripteur de fusion de différents espaces colorimétriques.....	26
2.3.6 Normalisation de GOG	27
2.4. Descripteur LOMO	27
2.4.1 Gérer les variations d'éclairage	27
2.4.2 Gérer les changements de point de vue.....	29
2.5. Rôle de Retinex.....	29
2.6. Rôle de LOMO	30
2.7. XQDA.....	30
2.8. La métrique de distance de Mahalanobis.....	32
2.9. Conclusion	33

Chapitre 3 : Résultat et discussion

3.1. Introduction.....	35
3.3. Mesure de performance	36
3.4. Réalisation et résultats	37
3.4.1 Analyse des performances de LOMO+XQDA	37
3.4.2 Analyse des performances de GOG+XQDA	39
3.5 Conclusion	42
Conclusion générale.....	44
Références bibliographiques.....	47

Introduction générale

Introduction générale

La correspondance d'apparence des images de personnes capturées dans des vues de caméras disjointes, appelée réidentification des personnes, suscite un intérêt croissant en raison de sa large gamme d'applications. Dans cette tâche, les images des personnes sont capturées sous différents angles et éclairages, avec des résolutions variées, des poses humaines et des environnements de fond divers. Ces grandes variations intra-personnelles dans les images des personnes posent de sérieux défis. De plus, les vêtements similaires entre différentes personnes ajoutent des défis supplémentaires. Pour surmonter ces difficultés, les chercheurs travaillent activement sur des descripteurs d'apparence et des méthodes de correspondance. Les descripteurs d'apparence fournissent des informations sur l'apparence (couleur et texture) des vêtements humains. Un bon descripteur doit être robuste face aux variations intra-personnelles tout en ayant une forte capacité discriminative pour distinguer différentes personnes.

Les images de personnes sont de faible résolution et présentent de grandes variations de pose ; par conséquent, il a été prouvé que l'indicateur le plus important pour la réidentification des personnes est l'information de couleur, comme les histogrammes de couleur et les descripteurs de nom de couleur [1]. Étant donné qu'ils ne peuvent pas suffisamment différencier les personnes de couleurs similaires, des descripteurs de texture comme le Local Binary Pattern (LBP) et les réponses des banques de filtres sont souvent combinés avec des descripteurs de couleur.

Un descripteur de covariance [2] décrit une région d'intérêt comme une covariance de caractéristiques de pixels. Il fournit un moyen naturel de fusionner différentes modalités, telles que la couleur et la texture, des caractéristiques de pixels en un seul méta-descripteur. Étant donné que le descripteur de covariance est obtenu par la moyenne des caractéristiques à l'intérieur de la région, il atténue les effets du bruit et des désalignements spatiaux. En conséquence, il a été appliqué avec succès à la réidentification des personnes.

Le processus de réidentification des personnes suit généralement les deux étapes suivantes :

- ✓ Extraction des descripteurs de personnes à partir de plusieurs caméras, comprenant la détection, le suivi, l'extraction de caractéristiques et la génération de descripteurs.
- ✓ Établir la correspondance (classement de similarité des descripteurs basé sur les attributions d'identité).

Les principales contributions de notre travail pour améliorer l'efficacité du système de réidentification des personnes sont :

1.Utilisation de la fonctionnalité LOMO (Local Maximal Occurrence), qui repose sur l'extraction des caractéristiques locales des images, permettant de se concentrer sur les informations locales importantes dans l'image. L'objectif est d'extraire des caractéristiques qui révèlent les détails fins des personnes et permettent de les distinguer précisément dans les images.

2.En plus de la fonctionnalité GOG (Hierarchical Gaussian Descriptor), qui repose sur l'utilisation de distributions gaussiennes hiérarchiques pour représenter l'image. L'image est divisée en régions selon les distributions gaussiennes, permettant de représenter l'image de manière hiérarchique, reflétant les détails locaux et globaux de l'image.

3.Pour améliorer la capacité de discrimination dans la réidentification des personnes, nous avons intégré la méthode d'apprentissage métrique connue sous le nom d'Analyse Discriminante Quadratique Croisée (XQDA) avec les fonctionnalités LOMO et GOG individuellement, et avons enregistré les résultats obtenus. Nous avons utilisé la fonction de distance de Mahalanobis pour mesurer les distances et obtenir les meilleurs résultats possibles.

4.Enfin, nous avons évalué et comparé les fonctionnalités LOMO et GOG pour déterminer laquelle est la meilleure pour atteindre rapidement l'identification des personnes.

La structure de notre travail est organisée comme suit :

- Le **premier chapitre**, nous présentons la biométrie et la ré-identification des personnes et leurs applications.
- Le deuxième **chapitre**, traite de la reconnaissance des fonctionnalités LOMO et GOG et de leur rôle dans l'identification des personnes.
- Dans le troisième **chapitre**, la partie pratique, qui contient la ré-identification des personnes en utilisant les deux fonctionnalités précédentes en les combinant avec XQDA et en le comparant.

Notre mémoire se termine par une conclusion générale et des perspectives pour les travaux futurs déduites de notre étude et résultats obtenus.

Chapitre 1: Biometrie et person re-identification

1.1. Introduction

Actuellement, les systèmes biométriques sont devenus les plus populaires en raison de leur fiabilité. Ils constituent une branche de l'informatique qui s'occupe de l'identification et de la vérification des individus. Le développement de ce type de systèmes vise à renforcer la sécurité et la confidentialité des informations. Dans ce chapitre, nous examinerons les différentes modalités utilisées en biométrie et leurs applications, ainsi que la réidentification des personnes.

1.2. Biometrie

Qu'est-ce que la Biométrie ?

La biométrie est un domaine qui traite de la mesure et de l'analyse des caractéristiques biologiques ou comportementales individuelles afin de déterminer ou de vérifier l'identité des individus. L'objectif de l'utilisation des technologies biométriques est de fournir un moyen efficace et sécurisé pour vérifier l'identité ou effectuer des opérations de reconnaissance électronique. la biométrie se distingue des méthodes traditionnelles d'authentification de l'identité, telles que les mots de passe ou les cartes intelligentes, car elle repose sur des caractéristiques individuelles uniques pour chaque personne, réduisant ainsi les risques de contrefaçon ou d'intrusion.

La biométrie consiste en l'analyse mathématique des caractéristiques biologiques d'une personne et a pour objectif de déterminer son identité de manière irréfutable. Contrairement à ce que l'on sait ou ce que l'on possède, la biométrie est basée sur ce que l'on est et permet ainsi d'éviter la duplication, le vol, l'oubli ou la perte. les caractéristiques utilisées doivent être universelles (c'est-à-dire communes à tous les individus), uniques (pour pouvoir différencier deux individus) et permanentes (c'est-à-dire invariantes dans le temps pour chaque individu [13]). La biométrie a révolutionné la manière dont l'identification est réalisée. elle devient un élément essentiel de tout système de sécurité, notamment dans le contrôle d'accès, les applications gouvernementales et la criminalistique. plusieurs caractéristiques biométriques sont utilisées dans l'identification des individus, parmi lesquelles: le visage, l'iris, la voix, l'empreinte digitale, la signature, la géométrie de la main, l'oreille, etc [14].

1.3. Présentation des techniques biométriques

1.3.1 Reconnaissance faciale

Tout au long de l'histoire de l'humanité, les gens ont utilisé le visage pour distinguer une personne d'une autre. La reconnaissance faciale est une application informatique qui identifie

ou vérifie automatiquement une personne à l'aide d'une image numérique ou d'une trame vidéo provenant d'une source vidéo. Une des façons de le faire est de comparer l'exemple donné avec les exemples dans la base de données [10]. En biométrie faciale, principalement des images visibles sont utilisées, qui ne sont pas affectées par la température et les lunettes [9].

Avec la modalité de reconnaissance faciale, une image numérique est utilisée pour analyser les caractéristiques faciales telles que les distances entre les yeux, la bouche et le nez. Ces mesures sont stockées dans une base de données et utilisées pour la comparaison avec une personne se tenant devant une caméra. La modalité de balayage facial est basée sur la séquence biométrique standard d'acquisition d'image, de traitement d'image, de localisation des caractéristiques distinctives, de création de modèle et de correspondance. Une image optimale est capturée à l'aide d'une caméra haute résolution, avec un éclairage modéré et les utilisateurs font face directement à la caméra. Les images d'inscription définissent les caractéristiques faciales à utiliser dans toutes les futures vérifications, donc une inscription de haute qualité est essentielle. Les défis qui surviennent dans le processus d'acquisition d'image comprennent la distance par rapport à l'utilisateur, l'acquisition inclinée et l'éclairage. La distance par rapport à la caméra réduit la taille du visage et donc la résolution de l'image. Les utilisateurs ne regardant pas directement la caméra, positionnés verticalement ou horizontalement à plus de 15 degrés de la position idéale sont moins susceptibles d'avoir des images acquises. Les conditions d'éclairage, qui peuvent rendre une image sous-exposée ou éblouissante, peuvent poser des défis. De plus, la reconnaissance bidimensionnelle est affectée par les changements de cheveux, d'âge et de lunettes de la personne. Cette modalité nécessite un équipement approprié de caméra pour l'identification de l'utilisateur. D'autre part, cette modalité présente l'avantage d'être non intrusive et relativement moins chère [11].



Figure 1.1: Capture de l'image d'un Reconnaissance faciale

1.3.2 Reconnaissance de l'iris

L'iris est un diaphragme circulaire mince, situé entre la cornée et le cristallin de l'œil humain. L'iris est perforé près de son centre par une ouverture circulaire appelée la pupille. La fonction de l'iris est de contrôler la quantité de lumière entrant par la pupille. La biométrie de l'iris est plus fiable et précise par rapport à d'autres traits biométriques tels que l'empreinte digitale.

La localisation de l'iris est une étape importante dans la reconnaissance de l'iris car, si elle est mal faite, le bruit résultant (comme les cils, les reflets, les pupilles et les paupières) dans l'image peut entraîner de mauvaises performances. L'imagerie de l'iris nécessite l'utilisation d'une caméra numérique de haute qualité. Les caméras d'iris commerciales d'aujourd'hui utilisent généralement la lumière infrarouge pour illuminer l'iris sans causer de dommages ou d'inconfort à la personne [9].



Figure 1.2: Capture de l'image d'un iris

1.3.3 Empreinte digitale

L'empreinte digitale est l'une des plus anciennes et des plus populaires biométries qui a prouvé son efficacité et sa fiabilité dans l'identification des utilisateurs. L'empreinte digitale est composée de crêtes et de vallées trouvées à la surface du doigt [4]. Un inconvénient majeur des systèmes d'identification par empreinte digitale est que de petites blessures et brûlures affectent fortement l'empreinte digitale [5]. Les 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 médico-légales du monde entier pour les enquêtes criminelles [6].



Figure 1.3: Capture de l'image d'une empreinte digitale

1.3.4 Numérisation de la signature

Une autre biométrie comportementale est la signature, par laquelle les données peuvent être extraites par la signature de cette personne particulière [10]. Il s'agit d'une analyse comportementale où différents éléments (mesure de la vitesse, ordre d'écriture, pression exercée, accélération...) sont Mesurés lors de la signature. La falsification est possible en passant par une phase d'apprentissage, la signature peut varier selon le stress de l'utilisateur [7].



Figure 1.4: Capture d'une signature

1.3.5 . 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 de la parole [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 synthèse du son.

Les avantages de cette modalité incluent sa nature non intrusive, une acceptabilité sociale élevée, un temps de vérification relativement court et une technologie requise relativement peu coûteuse. D'autre part, il présente certaines limites. La voix d'une personne peut être facilement enregistrée et utilisée pour un accès non autorisé. Il a une faible précision, où une maladie ou l'environnement environnant peut changer la voix d'une personne, rendant l'identification absolue difficile voire impossible [9].



Figure 1.5: Capture de Vérification vocal

1.3.6 Reconnaissance de la démarche

La démarche est la manière dont une personne marche et est l'un des rares traits biométriques qui peut être utilisé pour reconnaître les gens à distance. La plupart des algorithmes de reconnaissance de la démarche tentent d'extraire la silhouette humaine afin de décrire la démarche d'un individu [11].

Cependant, la démarche d'un individu est affectée par plusieurs facteurs, dont le choix de chaussures, la nature des vêtements, l'affliction des jambes, la surface de marche ... etc

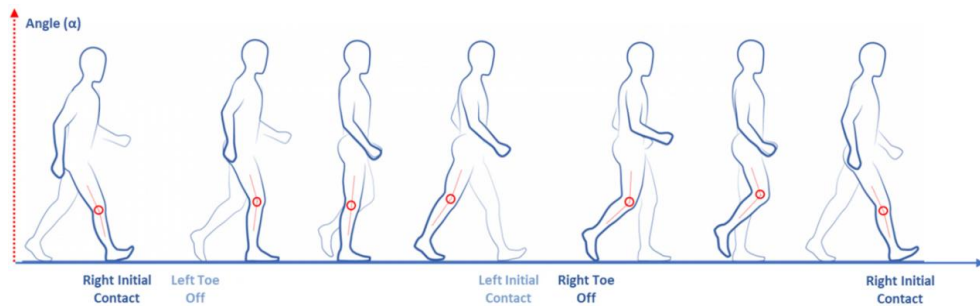


Figure 1.6: Capture de la démarche

1.3.7. Oreille

Cette modalité est plus stable que les expressions faciales, ce qui la rend attrayante pour l'identification biométrique. La reconnaissance de l'oreille peut se faire à l'aide d'une image numérique, d'une image thermographique ou d'une empreinte auriculaire. Une empreinte auriculaire peut être obtenue en pressant l'oreille contre du papier plat. Cette modalité présente quelques avantages puissants par rapport à de nombreuses autres modalités, tels que sa précision, permettant des taux d'inscription et de vérification élevés, ne nécessitant pas un expert pour interpréter le résultat de la comparaison, et pouvant être utilisée avec des caméras existantes et des dispositifs de capture d'images [9].



Figure 1.7: Capture d'une Oreille

1.3.8 Géométrie de la main

Jusqu'à 90 caractéristiques de la main sont mesurées (forme de la main et des articulations, longueur et largeur des doigts, longueur interarticulations, etc.). Le taux d'erreurs dans la reconnaissance est assez élevé, en particulier pour des personnes appartenant à une même famille en raison d'une forte ressemblance. De plus, la forme de la main évolue beaucoup avec l'âge [12].

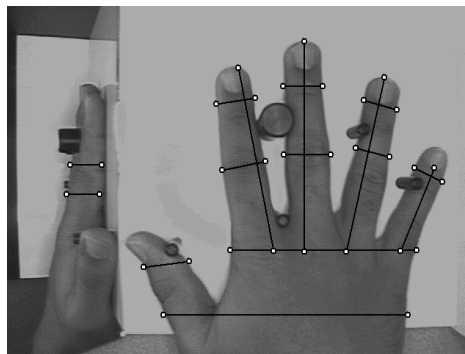


Figure 1.8: Scan de la forme de la main

1.4. Applications des systèmes biométriques

Il existe de nombreuses applications des systèmes de reconnaissance biométrique utilisés pour la surveillance, l'application de la loi, la gestion du temps et des présences, le contrôle d'accès logique et le contrôle d'accès physique. D'autres applications comprennent :

- Les systèmes biométriques sont utilisés pour sécuriser les systèmes de surveillance et également pour les bâtiments physiques pour diverses applications telles que le contrôle d'accès, les réseaux informatiques, la vérification, les enquêtes, l'authentification, la détermination de la parenté, la banque en ligne, le contrôle des frontières, le commerce électronique, la gestion des dossiers médicaux, la surveillance de la sécurité et la distribution des aides sociales.

- Pour améliorer le niveau de confiance après la vérification de l'individu authentifié avec tous ses droits et privilèges.

- Pour identifier un individu et contrôler l'accès aux informations, aux espaces physiques et aux services.
- Améliorer la sécurité, réduire la fraude et renforcer la sécurité nationale.
- Fournir une unicité, une forte immunité et une stabilité à la contrefaçon [13].

Les applications des technologies biométriques peuvent être divisées en principaux groupes tels que les applications commerciales, gouvernementales, judiciaires, de gestion des frontières, civiles, de vérification des clients au point de vente. Les applications clés comprennent la gestion des frontières, l'interface pour les applications criminelles et civiles, et la vérification des premiers intervenants. Les solutions d'entreprise nécessitent la supervision des personnes, des processus et des technologies. L'infrastructure réseau est devenue essentielle aux fonctions des entreprises, du gouvernement et des modèles commerciaux basés sur le web. Par conséquent, sécuriser l'accès à ces systèmes et s'assurer de l'identité de chacun est essentiel. Cela est utilisé pour prévenir le vol de biens marquant l'identité de la personne autorisée, tels que les badges de sécurité, les licences, ou les propriétés, et pour prévenir les actes frauduleux tels que la contrefaçon de badges d'identification ou de licences, afin de garantir la sécurité et la sûreté, et ainsi de réduire les taux de criminalité [14].

1.5. Ré-identification des personnes

La ré-identification des personnes (Re-ID) a été largement étudiée en tant que problème spécifique de récupération de personnes à travers des caméras non chevauchantes [15]. À partir de la personne d'intérêt de la requête, l'objectif de la Re-ID est de déterminer si cette personne est apparue à un autre endroit à un moment distinct capturé par une caméra différente [16]. La personne de la requête peut être représentée par une image [17], une séquence vidéo [18], voire une description textuelle [19]. En raison de la demande urgente de sécurité publique et du nombre croissant de caméras de surveillance sur les campus universitaires, dans les parcs d'attractions, dans les rues, etc., la ré-identification des personnes est indispensable dans la conception des systèmes de surveillance vidéo intelligents. En raison de son impact de recherche et de son importance pratique, la Re-ID est une communauté de vision en pleine croissance.

La ré-identification des personnes repose sur la tâche de détection des piétons. Tout d'abord, un ensemble de galeries 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 requête. Enfin, l'image la mieux adaptée en fonction de la similarité mesurée est considérée [20].

1.6. Re-identification pipeline

Un mécanisme automatisé de réidentification prend en entrée soit des pistes, soit des boîtes englobantes contenant des images segmentées de personnes individuelles, telles que générées par un processus de suivi ou de détection localisé d'un système de surveillance visuelle. Pour faire correspondre automatiquement des personnes à différents endroits dans le temps capturé par différentes vues de caméra [16], un processus de réidentification suit généralement les étapes suivantes :

- Extraction de caractéristiques d'imagerie plus fiables, robustes et concises que les données brutes des pixels.
- Construction d'un descripteur ou d'une représentation, par exemple un histogramme de caractéristiques, capable de décrire et de discriminer les individus.
- Appariement des images ou des pistes spécifiées avec une galerie de personnes dans une autre vue de caméra en mesurant la similarité entre les images, ou en utilisant une procédure d'appariement basée sur un modèle.

Une étape d'entraînement pour optimiser les paramètres d'appariement peut être nécessaire ou non en fonction de la stratégie d'appariement. Ces étapes de traitement imposent certaines exigences en matière de conception d'algorithmes et de systèmes. Cela a conduit à la fois au développement de nouvelles techniques de vision par ordinateur et à l'exploitation des techniques existantes pour résoudre les problèmes de représentation des caractéristiques, d'appariement de modèles et d'inférence dans le contexte afin d'améliorer l'efficacité des systèmes de ré-identification [16]. Les approches contemporaines de représentation de la ré-identification exploitent généralement des caractéristiques de bas niveau telles que la couleur, la texture, la structure spatiale [1], ou des combinaisons de celles-ci [18]. Cela est dû au fait que ces caractéristiques peuvent être mesurées relativement facilement et de manière fiable, et fournissent un niveau raisonnable de discrimination inter-personnelle ainsi qu'une invariance inter-caméra. Ces caractéristiques sont ensuite encodées en des descripteurs de personne de longueur fixe, par exemple sous forme d'histogrammes, de covariances ou de vecteurs de Fisher [16]. Prenons par exemple l'appariement ; une fois qu'une représentation appropriée a été obtenue, des algorithmes d'appariement basés sur le plus proche voisin ou des algorithmes d'appariement basés sur un modèle tels que le classement par vecteurs de support peuvent être utilisés pour la ré-identification. Dans chaque cas, une métrique de distance (par exemple Euclidienne ou Bhattacharyya) doit être choisie pour mesurer la similarité entre deux

échantillons. Les approches d'appariement basées sur un modèle [19] et les métriques de distance basées sur le plus proche voisin [20] peuvent toutes deux être optimisées de manière discriminative pour maximiser les performances de ré-identification en utilisant des données d'entraînement annotées d'images de personnes. Reliant ces deux étapes, certaines études ont également tenté d'apprendre directement à partir des données des caractéristiques de bas niveau discriminatives [16].

Des aspects complémentaires du problème de ré-identification ont également été explorés pour améliorer les performances, tels que l'amélioration de la robustesse en combinant plusieurs cadres de caractéristiques le long d'une trajectoire de piste [18], l'analyse basée sur des ensembles [21], la prise en compte du contexte externe tel que les groupes de personnes [22], et l'apprentissage de la topologie des réseaux de caméras [23] afin de réduire l'espace de recherche d'appariement et donc de réduire les faux positifs [16]. Le pipeline général d'un système de ré-identification de personne de bout en bout est présenté dans la figure (1.9).

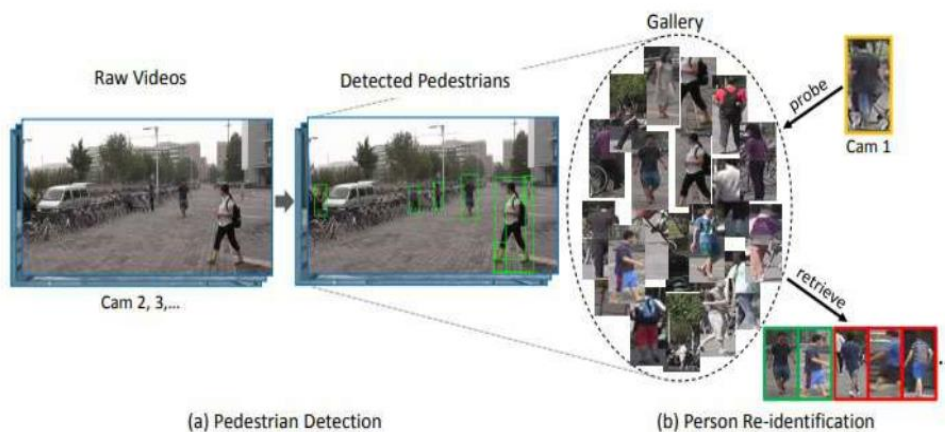


Figure 1.9: Pipeline d'un système de ré-identification de personne de bout en bout

1.7. Challenges

La Re-ID est une tâche difficile en raison de nombreux facteurs tels que la présence de différents points de vue, des résolutions d'image variables, des changements d'éclairage, des poses non contraintes, des occultations, des modalités hétérogènes, etc. Quelques exemples de défis sont montrés dans la figure 1.7. Les facteurs de défis et leurs effets sont expliqués ci-dessous:

- **Variation de l'éclairage:** Les conditions d'éclairage peuvent varier dans différentes scènes de caméra ou au cours de la journée. La même personne observée sous différentes

conditions d'éclairage peut avoir une couleur différente dans son apparence. Cela augmente la variation intra-classe [20].

- **Variation du point de vue de la caméra:** Étant donné que la hauteur des caméras, la distance entre la personne et la caméra, et la direction dans laquelle les personnes font face varient, différentes formes ou tailles de piétons peuvent être observées sous différents angles de vue [20], ce qui peut entraîner l'échec des systèmes de Re-ID entraînés sur le domaine source uniquement lors des tests directement sur un domaine cible, car le modèle ne peut pas apprendre des caractéristiques invariantes à la caméra cible [24].

- **Variation des poses:** Les déformations importantes des poses sur le corps humain sont généralement causées par les mouvements relatifs entre les parties rigides du corps. Pendant le mouvement, certaines parties du corps comme les bras, les jambes et le tronc peuvent être considérées comme des régions rigides, et leurs mouvements peuvent être modélisés par des transformations affines [25]. Dans un scénario de surveillance typique, la pose d'une personne par rapport à une caméra varie en raison du point de vue de la caméra et de la direction de marche de la personne. En conséquence, les informations extraites des images à la fois des indices corporels statiques et dynamiques peuvent changer radicalement à la fois entre les caméras et sur une longue période. Ainsi, la direction de la pose/marche du sujet est un critère très important à considérer dans le schéma de re-ID basé sur la marche [26].

- **Basse résolution:** Dans la plupart des environnements réalistes, le coût du nombre requis de caméras dans toutes les zones peut être très élevé, rendant la couverture plutôt clairsemée, laissant des "zones mortes". Les caméras sont donc généralement installées en hauteur sur les murs, et les piétons sont ainsi généralement loin de la caméra. Même pour les caméras hautes résolution, pour une personne donnée, l'image pourrait toujours être de résolution relativement basse [20].

- **Détection inexacte des piétons:** Dans un contexte d'analyse vidéo automatique, les méthodes de ré-identification des personnes fonctionnent généralement sur des images de piétons recadrées renvoyées par un détecteur de personnes. Cependant, les performances des algorithmes de détection de piétons existants ne sont pas suffisamment précises pour le but de la ré-identification, c'est-à-dire que les détections incluent trop de fond ou contiennent seulement une partie de la personne. Les régions du corps humain ne sont donc pas bien

alignées entre les images, ce qui a un impact sérieux sur les performances de ré-identification de la plupart des méthodes existantes [20].

- **Grand nombre de candidats dans l'ensemble de la galerie:** Un réseau de caméras peut couvrir un grand espace public, comme une gare ou un campus. Il peut donc y avoir un grand nombre de candidats pour une requête de ré-identification donnée, et le nombre de candidats augmente avec le temps. La computation pour la correspondance avec un grand ensemble de galeries devient coûteuse. Pour atténuer ce problème, certains raisonnements temporels et la disposition spatiale des différentes caméras peuvent être utilisés pour réduire l'ensemble des correspondances candidates [20].

- **Occlusion partielle:** Les types d'occlusion faciale qui se produisent peuvent varier considérablement. En général, il existe deux types majeurs d'occlusion partielle : systématique et temporaire. Les occlusions systématiques sont causées par l'existence de composants faciaux individuels (par exemple, cheveux, moustache, ou une cicatrice), ou par les gens portant des ornements (par exemple, lunettes, vêtements, un chapeau ou un masque chirurgical, ou des marques). Les occlusions temporaires surviennent lorsqu'une partie du visage est temporairement obscurcie par d'autres objets (par exemple, des personnes se déplaçant devant le visage ou des mains couvrant le visage), ou une auto-occlusion due à des changements de pose de la tête ou plaçant temporairement des objets devant le visage [27].

- **Contrainte en temps réel:** Dans certaines situations d'urgence, nous devons trouver immédiatement la localisation d'un suspect. Il est important d'avoir une implémentation en temps réel avec une faible latence pour traiter de nombreux flux vidéo entrants, et retourner les résultats de la requête rapidement. L'espace de recherche pour la correspondance des personnes peut être extrêmement large avec de nombreux candidats potentiels à discriminer. Ainsi, le temps de recherche est crucial [20].

- **Taille d'échantillon réduite:** En général, un module de ré-identification peut être requis pour faire correspondre des images uniques de la sonde avec des images uniques de la galerie. Cela signifie que, d'un point de vue de classification conventionnelle, il y a probablement insuffisamment de données pour apprendre un bon modèle de la variabilité intra-classe de chaque personne. Un apprentissage "one-shot" peut être nécessaire sous lequel seule une paire d'exemples est disponible pour l'apprentissage du modèle. Pour cette raison, de nombreux cadres traitent la re-ID comme un problème de classification binaire par paires

(même contre différent) [19] au lieu d'un problème de classification multi-classe conventionnel [15].

- **Vêtements similaires:** Dans un grand ensemble de galeries, il y a une forte probabilité que les gens portent des vêtements similaires. La plupart des gens dans les espaces publics portent des vêtements sombres en hiver ; beaucoup de gens portent presque les mêmes jeans bleus. Cela augmente l'ambiguïté et l'incertitude dans le processus de correspondance. Dans ce cas, il est plus difficile de trouver une signature distinctive de l'apparence visuelle des différentes personnes [20].

- **Étiquetage des données:** Si l'on exploite une stratégie d'apprentissage supervisé pour entraîner un bon modèle robuste aux variations inter-caméras, des personnes de chaque vue annotées avec des identités ou des étiquettes binaires décrivant même contre différent sont nécessaires. Par conséquent, les modèles qui peuvent être appris avec moins de données d'entraînement sont préférés, car pour un grand réseau de caméras, collecter des données étiquetées extensives de chaque caméra serait prohibitif en termes de coût [15].



Figure 1.10: Exemple de quelques défis de ré-identification de personnes. Chaque paire d'images montre la même personne sauf (g). (a) Variation du point de vue, (b) variation de la pose, (c) changements d'éclairage, (d) occlusion partielle, (e) détection de piétons inexacte, (f) changement d'accessoires, (g) basse résolution et (h) personnes différentes avec des vêtements similaires.

1.8. Bases des données de référence

Plusieurs bases de données de référence pour la ré-identification de personnes ont été publiées ces dernières années. Plusieurs facteurs doivent être pris en compte lors de la création d'un jeu de données de référence [28], tels que:

- Nombre de piétons.

- Nombre d'images par personne dans une vue de caméra.
- Variations des résolutions, des éclairages, des poses, des occultations et de l'arrière-plan dans une même vue de caméra et entre différentes vues de caméra.
- Nombre de vues de caméra.

Quelques jeux de données publics qui ont été testés pour les algorithmes de ré-identification de personnes et leurs caractéristiques sont illustrés dans le Tableau (1.1).

Tableau 1.1: Exemple de différents ensembles de données et de leurs caractéristiques

Ensembles de données	Temps de sortie	# Identités	# Image	# caméras	Taille de la culture
VIPeR	2007	632	1264	2	128*48
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	731,2	592,438	3,4,4	—
ThermalWo-rld	2018	516	15,118	20	—

1.9. Applications de la ré-identification des personnes

Les stratégies de ré-identification des personnes offrent un grand potentiel dans divers domaines, allant de la défense et de la surveillance au commerce de détail et aux soins de santé.

- **Suivi de personne entre les caméras :** La vision par ordinateur nécessite la capacité de surveiller les personnes à travers plusieurs caméras, d'analyser les mouvements de foule et de reconnaître les déplacements afin de comprendre une scène. En général, le suivi entre les caméras se compose de deux étapes : le suivi d'objet dans une seule caméra et l'association des données entre plusieurs caméras. Par conséquent, outre la difficulté du suivi dans une seule caméra, le principal défi du suivi entre les caméras est d'associer les trajectoires des mouvements humains dans différentes vues [16].

- **Suivi par détection :** La ré-identification 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 occlusions fréquentes et des interactions individuelles. L'idée principale est de détecter les personnes, d'estimer leurs modèles de mouvement et d'associer les détections dans différents cadres. Cette étape de liaison, appelée association de données, est en fait une sorte de ré-identification. 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 [16].

- **Recherche de personne:** Dans ce cas, la ré-identification est liée à une tâche de reconnaissance. Une requête spécifique avec un individu cible est donnée, et toutes les instances correspondantes dans une vaste base de données sont recherchées. Pour la recherche d'images, la tâche de ré-identification est utilisée, et elle renvoie généralement des listes classées, des objets similaires et ainsi de suite [16].

- **Interaction homme-machine:** Au cours des dernières décennies, les scientifiques se sont davantage concentrés sur le développement de technologies qui contribuent 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 [29].

- **Analyse à long terme du comportement humain et des activités:** Analyser les habitudes de shopping des consommateurs, par exemple, en les observant toucher, interroger et essayer des articles dans les magasins sous diverses caméras de surveillance. Un autre exemple est la recherche en santé gériatrique, qui examine les actions à long terme des personnes âgées pour aider les médecins à établir des diagnostics plus précis [16].

1.10. Conclusion

La ré-identification des personnes est un domaine de recherche récent qui combine la biométrie et la vision par ordinateur, mais qui donne d'excellents résultats. Dans ce chapitre, nous avons discuté du domaine de la biométrie ainsi que de ses différentes modalités et de leurs applications.

Nous avons également abordé la ré-identification des personnes. Des études en cours développent constamment de nouvelles méthodes efficaces pour relever les défis de la Re-ID individuelle et augmenter la précision de la reconnaissance en incorporant des méthodes puissantes d'extraction de caractéristiques, de réduction de la dimensionnalité, de fusion et de classification.

Chapitre 2 : Les méthodes de Re-ID

2.1. Introduction

La Re-ID des personnes est un problème qui consiste à retrouver une personne dans une galerie d'images ayant la même identité que celle de la sonde. C'est une problématique complexe en raison des grandes variations intra-catégories dans l'éclairage, la pose ou l'angle de vue, ainsi que l'occultation. De nombreuses approches ont été proposées pour la réidentification des personnes, ce qui a considérablement fait avancer ce domaine. Dans ce chapitre, nous présentons une représentation efficace des caractéristiques appelée "Occurrence Maximale Locale" (LOMO), ainsi qu'un descripteur gaussien hiérarchique nommé "GOG". Nous abordons également l'approche d'apprentissage métrique (XQDA) ainsi que l'utilisation de la distance de Mahalanobis pour mesurer la similarité.

2.2 Représentations des caractéristiques pour la ré-identification des personnes

De nombreuses approches tentent de construire des représentations distinctives des caractéristiques pour décrire l'apparence des piétons dans différents environnements. Afin de parvenir à une discrimination et une robustesse contre les variations, ces caractéristiques sont généralement extraites de bandes horizontales ou de blocs denses. À partir des bandes chevauchantes, Lisanti et al. [30] ont calculé les caractéristiques locales pondérées de l'histogramme des couleurs et des motifs binaires locaux (LBP). En général, ces descripteurs basés sur des bandes sont robustes face au problème de désalignement des corps entre différentes vues et peuvent bien capturer l'apparence globale.

Zhao et al. ont essayé d'apprendre des filtres de niveau intermédiaire à partir de clusters de petits blocs denses. Récemment, Liao et al. [31] ont proposé le descripteur LOMO qui a montré une impressionnante robustesse contre les changements d'angle de vue. En décrivant chaque image de piéton comme un ensemble de distributions gaussiennes hiérarchiques représentées par les moyennes et les covariances, Matsukawa et al. ont conçu le descripteur GOG qui est également calculé à partir de blocs denses. Cependant, comme une pièce de monnaie à deux faces, un inconvénient de ces descripteurs basés sur des blocs denses est qu'ils ne sont pas bons pour décrire l'apparence globale, bien que certains aient envisagé de calculer des caractéristiques à partir d'espaces multi-échelles comme LOMO et GOG.

Pour combiner les avantages des caractéristiques basées sur des bandes et des blocs denses, nous soutenons que les représentations de caractéristiques doivent être calculées simultanément à partir de bandes (régions plus grandes) et de blocs denses (petites régions). Dans ce travail, nous fusionnons le descripteur LOMO avec les caractéristiques extraites d'un

espace pyramidal à deux niveaux de bandes chevauchantes. Ainsi, les détails fins des blocs denses et l'apparence grossière des régions plus grandes sont bien intégrés pour améliorer la discrimination [32].

2.2.1 Apprentissage métrique pour la ré-identification des personnes

L'apprentissage métrique a été largement appliqué au problème de la ré-identification des personnes, et les métriques sont généralement apprises soit du point de vue de la distance, soit du point de vue de la similarité bilinéaire. Du point de vue de la distance, Mignon et al. ont proposé une méthode pour traiter le problème des données d'entraînement déséquilibrées. Liao et Li [33] ont proposé d'apprendre la métrique à partir d'une formulation logistique en appliquant une stratégie de pondération asymétrique. En utilisant des contraintes par triplets, Zheng et al. ont proposé d'apprendre la métrique à partir de la comparaison des distances relatives. Pour aborder le problème du désalignement des vues, Sun et al. ont intégré la métrique avec certaines variables latentes pendant l'apprentissage. En exploitant des échantillons négatifs facilement disponibles, Zhou et al. ont proposé d'apprendre une métrique globale avec certaines métriques locales fines. Sans optimisation lourde, l'algorithme KISSME [34] a dérivé une solution de forme fermée de la métrique qui peut être calculée très efficacement. En tant qu'extension de KISSME, l'analyse discriminante quadratique entre vues (XQDA) a appris une métrique plus efficace accompagnée d'un sous-espace discriminant [35].

2.3. Descripteur GOG

Il s'agit d'un descripteur récemment proposé basé sur la distribution gaussienne des caractéristiques des pixels. L'image est divisée en 7 bandes horizontales superposées qui sont à leur tour divisées en petits patches de 5×5 se chevauchant. Les caractéristiques des pixels sont extraites de chaque patch et résumées en utilisant une distribution gaussienne. Enfin, les patches dans la même bande sont modélisés en utilisant une autre distribution gaussienne unique [36].

Nous présentons le descripteur gaussien hiérarchique proposé nommé GOG dans la figure 2.1. Pour obtenir la représentation des caractéristiques d'une image de personne, nous adoptons un modèle basé sur les parties [37]. Nous supposons que G régions d'une image de personne sont données à l'avance, ce qui correspond généralement à 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.

The GOG descriptor est présenté dans la figure 4. Pour obtenir une représentation des caractéristiques d'une personne, nous utilisons un modèle par parties, comme illustré dans la figure (2.1) [38]. Nous supposons que les régions G d'une image de personne sont données à l'avance, généralement les bandes horizontales de l'image. Ces régions sont représentées par un vecteur de caractéristiques retourné par le GOG.

Le reste de cette section entre dans les détails du descripteur. Dans la section (a), nous extrayons des patches topiques de chaque région. Dans la section (b), nous utilisons ensuite une distribution gaussienne des caractéristiques des pixels pour décrire chacun des patches locaux ; ceux-ci sont appelés Gaussiennes de patch. Dans la section (c), en tenant compte de la géométrie gaussienne sous-jacente, chaque patch gaussien est aplati et vectorisé. Dans la section (d), les patches gaussiens à l'intérieur d'une région sont ensuite additionnés pour former une région gaussienne. Dans la section (e), nous créons un vecteur de caractéristiques après avoir aplati la région gaussienne. Dans la section (f), les vecteurs de caractéristiques de chaque région sont combinés en un seul vecteur.

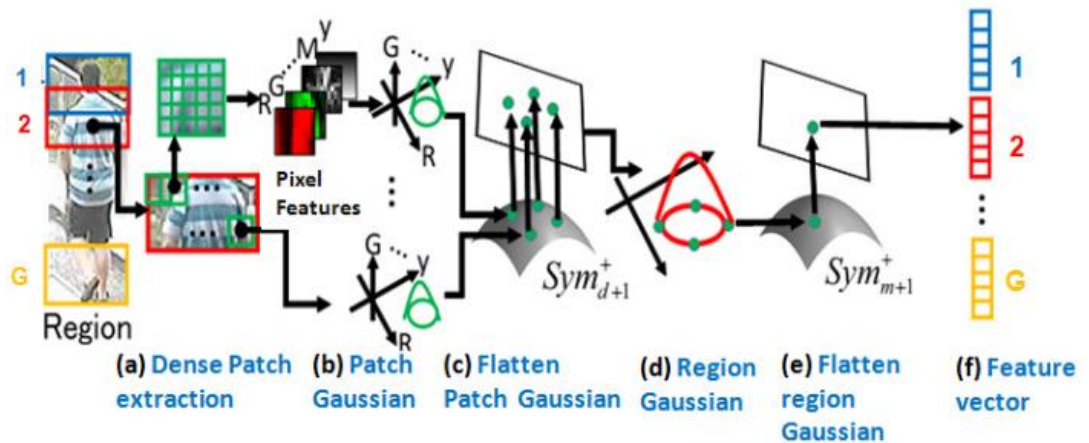


Figure (2.1): Descripteur GOG

2.3.1. Caractéristiques des pixels

Concentrons-nous sur l'une des régions G 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 pixels p (voir Figure 2.1 (a)). Afin de caractériser chaque pixel dans le patch, nous extrayons un vecteur de caractéristiques de dimension d 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. Comme 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 de 8 dimensions définies comme suit:

$$F_I = [y, M_{0^\circ}, M_{90^\circ}, M_{180^\circ}, M_{270^\circ}, R, G, B]^T \quad (2.1)$$

Où y est l'emplacement du pixel dans la direction verticale, $M_\theta \in \{0^\circ, \dots, 270^\circ\}$ sont les magnitudes du gradient d'intensité du pixel 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 pour égaliser les échelles des différentes valeurs des caractéristiques. L'emplacement du pixel est introduit pour exploiter les informations spatiales dans chaque région [35]. L'utilisation uniquement de l'emplacement vertical de l'image découle de l'analyse dans, les images de personnes tendent à être bien alignées dans la direction verticale tandis que le changement de pose / point de vue provoque un grand désalignement dans la direction horizontale. Notez que l'on aimerait définir y_i à partir du haut (ou du centre) de la région actuelle comme dans. 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 à partir du haut de l'image [39]. Les informations de gradient sont introduites pour décrire les informations texturales des vêtements. L'orientation du gradient $O = \arctan(I_Y/I_X)$ est calculée à partir des dérivées 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'informations due à la quantification, nous utilisons un vote doux dans les deux bins d'orientation proches. Les poids de vote sont déterminés linéairement à partir des distances des orientations quantifiées comme dans le vecteur GOG dans. Pour se concentrer sur les bords de gradient élevés, nous multiplions la magnitude de gradient $M = \sqrt{I_Y^2 + I_X^2}$ par l'orientation quantifiée O_θ et obtenons la magnitude de gradient orientée; $M_\theta = M O_\theta$. Les informations de couleur sont l'indice le plus important pour la réidentification des personnes. Nous utilisons les valeurs des canaux de couleur de l'espace colorimétrique le plus basique: RGB. D'autres espaces colorimétriques, tels que Lab, HSV et YCbCr.

2.3.2 Gaussiennes de région

Après avoir extrait les caractéristiques des pixels à l'intérieur d'un patch, nous les résumons ensuite via la distribution paramétrique la plus classique qui a pour paramètres la moyenne et la covariance : la distribution gaussienne (Figure 2.1 (b)). Pour chaque patchs, nous modélisons les vecteurs de caractéristiques sous forme de la gaussienne du patch $N(f; \mu, \Sigma)$ définie comme suit,

$$N(F; U_S, \Sigma_S) = \frac{\exp(-\frac{1}{2}(F-U_S)^T \Sigma_S^{-1}(F-U_S))}{(2\pi)^{d/2} |\Sigma_S|} \quad (2.2)$$

Où $|\cdot|$ est le déterminant d'une matrice, μ_s est le vecteur moyen et Σ_s est la matrice de covariance du patch échantillonnés. Le vecteur moyen et la matrice de covariance sont respectivement estimés par $u_s = \frac{1}{n_s} \sum_i \in L_s f_i$ et $\Sigma_S = \frac{1}{n_s-1} \sum_i \in l_s (f_i - u_s) (f_i - U_S)^T$.
Où L_s est la zone du patch échantillonné s et n_s désigne le nombre de pixels dans L_s .

Il convient de noter que les vecteurs moyens et les matrices de covariance échantillonnés de manière dense peuvent être calculés efficacement à l'aide d'images intégrales. Comme les régions peuvent se chevaucher, nous construisons, les images intégrales des caractéristiques des pixels pour une image globale de 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. Comme on s'attend à ce qu'un patch local se compose d'un petit nombre de couleurs/textures, nous supposons que la gaussienne unimodale est suffisante pour décrire la distribution de ses caractéristiques de pixels [38].

2.3.3 Mappage dans l'espace tangent et demi-vectorisation

Comme nous l'expliquerons dans la sous-section suivante, notre descripteur est une représentation résumée des gaussiennes de patch dans une région. Pour cette résumé, 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 [40].

Une variété riemannienne peut être aplatie localement dans un espace euclidien en le projetant dans un espace tangent doté d'une métrique riemannienne. L'espace des matrices symétriques définies positives (SPD) est également considéré comme une variété riemannienne et cet espace est récemment bien compris. La métrique log-euclidienne pour les matrices SPD fournit un moyen solide de mapper un point sur la variété dans un espace tangent euclidien via un logarithme matriciel. Pour tirer parti des avantages de la métrique log-euclidienne, nous incorporons les gaussiennes de patch dans la matrice SPD de manière similaire au travail précédent. À partir d'une analyse de la littérature sur la géométrie de l'information, l'espace des gaussiennes multivariées d -dimensionnelles peut être incorporé dans l'espace des matrices SPD $d + 1$ dimensions, désigné par $S_{ym_{d+1}}^+$. Nous représentons la gaussienne de patch d -dimensionnelle $N(u_s, \Sigma_s)$ dans $S_{ym_{d+1}}^+$ comme P_s

$$N(F; U_S, \Sigma_S) \sim P_S = |\Sigma_S|^{-\frac{1}{d+1}} \begin{bmatrix} \Sigma_S + U_S U_S^T & U_S \\ U_S^T & 1 \end{bmatrix} \quad (2.3)$$

Pour une théorie plus détaillée de cet encastrement, on peut se référer à la littérature. La matrice de covariance du patch local devient souvent singulière en raison du manque de nombre suffisant de pixels à l'intérieur du patch. Nous évitons ce problème en ajoutant la matrice identité Id à Σ_S avec une petite valeur constante positive, ϵ_S comme $\Sigma_S \leftarrow \Sigma_S + \epsilon_S \text{Id}$. Afin de décrire la distribution de la région dans une opération euclidienne, nous projetons ensuite chacune des gaussiennes de patch P_s dans un espace tangent via un logarithme matriciel (Figure (2.1 (c))). Nous stockons ensuite la partie triangulaire supérieure de la matrice projetée sous forme de vecteur puisque la matrice est symétrique. En considérant les entrées hors diagonale comme étant comptées deux fois lors du calcul de la norme, la matrice des gaussiennes de patch p_s devient un vecteur de dimension $m = (d^2 + 3d)/2 + 1$, défini comme suit : $\mathbf{g}_s = \text{vec}(\log(p_s)) = [\mathbf{b}_{s(1,1)}, \sqrt{2} \mathbf{b}_{s(1,2)}, \dots, \sqrt{2} \mathbf{b}_{s(1,d+1)}, \mathbf{b}_{s(2,2)}, \sqrt{2} \mathbf{b}_{s(2,3)}, \dots, \mathbf{b}_{s(d+1,d+1)}]^T$, Où $\log(\cdot)$ est l'opérateur de logarithme matriciel et $b_{s(i,j)}$ est l'élément (i, j) de $B_s = \log(p_s)$.

2.3.4 Gaussienne de région dans l'espace tangent

En raison de la variation de 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 dans une distribution de région (Figure (.2.1.d)). Pour cette sommation, nous utilisons également une distribution gaussienne qui peut décrire non seulement la covariance mais aussi la moyenne. Encore une fois, GMM pourrait être utilisé pour décrire des distributions plus précises. Cependant, faire correspondre des GMM n'est pas un problème trivial et entraînera une complexité pour faire correspondre les descripteurs de région [41]. La sommation avec une distribution gaussienne est réalisée en tenant compte d'une propriété spatiale des patches comme suit. Une image de personne contient souvent des régions d'arrière-plan qui diffèrent significativement d'un endroit à l'autre. Pour supprimer l'effet des régions d'arrière-plan, nous introduisons un poids pour chaque patch de manière similaire aux histogrammes [41] de couleur pondérés. 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 y central d'une image : $W_s = \exp(-(\mathbf{x}_s - \mathbf{x}_c)^2 / 2\sigma^2)$ où $\mathbf{x}_c = W/2$, $\sigma = W/4$.

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

$$u^G = \frac{1}{\sum_{S \in G} w_s} \sum_{S \in G} w_s g_s \quad (2.4)$$

$$\Sigma^G = \frac{1}{\sum_{S \in G} w_s} \sum_{S \in G} w_s (g_s - u^G)(g_s - u^G)^T \quad (2.5)$$

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 la forme de la gaussienne de région $N(g; \mu_G, \Sigma_G)$. Pour la mise en correspondance entre des descripteurs de région, il est pratique d'aplatir la gaussienne de région dans l'espace euclidien, car la plupart des méthodes de mise en correspondance telles que l'apprentissage de métrique sont conçues dans un espace euclidien. À cette fin, nous incorporons la gaussienne de région de dimension m dans des matrices SPD de dimension $m + 1$ de la même manière que l'équation (3) : $N(g; \mu_G, \Sigma_G) \sim \mathbf{Q}$ où \mathbf{Q} est une matrice SPD $(m + 1) \times (m + 1)$. Ici, la matrice de covariance Σ_G est régularisée comme $\Sigma_G \leftarrow \Sigma_G + \epsilon^G \mathbf{I}_m$, Nous cartographions ensuite \mathbf{Q} dans l'espace tangent de Sym^{+m+1} en utilisant le logarithme matriciel et le demi-vectoriser pour former un vecteur de caractéristiques de dimension $(m^2 + 3m)/2 + 1$, que nous désignons z (Fig.3 (e)). En extrayant la gaussienne de région pour chacune des régions G , nous obtenons des vecteurs de caractéristiques $\{z_g\}_{g=1}^G$. Afin de maintenir 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 $Z = [z_1^T, \dots, z_G^T]^T$ [1].

2.3.5 Descripteur de fusion de différents espaces colorimétriques

Il a été prouvé que les descripteurs extraits de différents espaces colorimétriques ont des propriétés complémentaires, et leur fusion améliore les précisions de Ré-ID [42]. Pour extraire plus d'informations colorimétriques dans les descripteurs GOG, nous remplaçons les valeurs des canaux RVB dans la caractéristique des pixels dans l'équation (1) par trois valeurs de canaux de couleur alternatives $\{\text{Lab}, \text{HSV}, \text{nRGB}\}$ et fusionnons leurs descripteurs GOG. Ici, nRGB est l'espace colorimétrique normalisé (par exemple, $\text{nR} = \text{R}/(\text{R}+\text{G}+\text{B})$). Comme il y a une redondance dans cet espace, nous utilisons uniquement $\{\text{nR}, \text{nG}\}$ dans cet espace colorimétrique. Ainsi, la dimension de la caractéristique des pixels de chaque espace colorimétrique $\{\text{RVB}, \text{Lab}, \text{HSV}, \text{nRnG}\}$ 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 z extrait de l'équation (1) comme GOG_{RGB} , et les descripteurs extraits des canaux de couleur alternatifs

comme GOG_{Lab} , GOG_{HSV} et GOG_{nRnG} , respectivement. La fusion est simplement effectuée en concaténant les descripteurs GOG sur les différentes caractéristiques des pixels comme

$GOG_{Fusion} = [GOG_{RGB}^T, GOG_{Lab}^T, GOG_{HSV}^T, GOG_{nRnG}^T]^T$. Par conséquent, la dimension du descripteur de fusion est de 3 (espaces colorimétriques) $\times 1081$ ($= (452 + 3 \times 45)/2 + 1$) $\times G$ (régions) $+ 1$ (espace colorimétrique) $\times 703$ ($= (362 + 3 \times 36)/2 + 1$) $\times G$ (régions) [1].

2.3.6 Normalisation de GOG

Pour les caractéristiques de haute dimension, la normalisation est un facteur important pour améliorer leurs performances. Étant donné que le descripteur GOG est de haute dimension, nous normalisons le descripteur en utilisant la normalisation de la norme L2, qui est la plus largement adoptée [43]. Nous avons observé qu'il existe des dimensions qui ont généralement des valeurs élevées/faibles communes entre différentes images dans le descripteur GOG. Cela est dû au fait que nous utilisons des caractéristiques de pixels qui ont différentes propriétés de leurs distributions, par exemple, la magnitude du gradient se distribue de manière dispersée dans les images, et l'intensité de couleur se distribue plus uniformément. Dans un tel 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 comme suit:

$$Z = (z - \bar{z}) / \|z - \bar{z}\|_2 \quad (2.6)$$

Où \bar{z} est la moyenne des échantillons de descripteurs GOG. Pour le descripteur de fusion, nous normalisons chacun des descripteurs GOG extraits sur les quatre espaces colorimétriques avant de les concaténer [44]. Pour la représentation en sacs de mots, une normalisation similaire est proposée pour refléter les mots co-manquants pour la similarité cosinus. En revanche, nous l'employons pour remédier à l'effet des dimensions biaisées [38].

2.4. Descripteur LOMO

2.4.1 Gérer les variations d'éclairage

La couleur est une caractéristique importante pour décrire les images d'une personne. Cependant, les conditions d'éclairage des caméras peuvent être très différentes, et les paramètres de l'appareil photo peuvent également être différents d'une caméra à l'autre [45].

Par conséquent, la perception les couleurs d'une même personne peuvent varier considérablement de différentes vues de la caméra. Par exemple, la figure 2.2 (a) montre un

échantillon images de la base de données VIPeR on peut voir que les images de la même personne dans les deux vues couleur de la caméra ont une grande différence d'éclairage et d'apparence de couleur. Dans cet article, nous proposons d'appliquer l'algorithme Retinex pour prétraiter les images de personnes. Retinex considère légèreté humaine et perception des couleurs. Il vise à produire une image couleur qui correspond à l'observation humaine de la scène. L'image restaurée contient généralement des informations sur les couleurs vives, en particulier des détails améliorés dans les zones sombres.

Nous implémentons l'algorithme Retinex multi-échelle selon, qui combine le Retinex à petite échelle pour la compression de la plage dynamique et le Retinex à grande échelle pour rendu tonal simultanément. En conséquence, l'algorithme gère automatiquement à la fois la constance des Couleurs et la compression de la plage dynamique, obtenant ainsi une bonne approximation de perception visuelle humaine. Plus précisément, nous utilisons deux échelles de centre/environnement Retinex, avec $\sigma=5$ et $\sigma=20$. De plus nous calculons automatiquement les paramètres gain/offset afin que les intensités résultantes s'étendent linéairement dans $\{0.255\}$. La figure (2.1(b)) montre quelques exemples d'images traitées par notre implémentation de Retinex. Par rapport à la figure (2.2 (a))



Figure 2.2: (a) Exemples de paires d'images de la base de données VIPeR; (b) Images traitées par Retinex. Images dans le même La colonne représente la même personne.

On peut observer que les images Retinex de la même personne à travers les caméras ont une meilleure cohérence en termes d'éclairage et couleur. Cela rend la réidentification des personnes plus facile que l'utilisation des images originales. Avec les images Retinex, nous appliquons l'histogramme de couleur HSV pour extraire les caractéristiques de couleur.

En plus de la description des couleurs, nous utilisons également l'échelle Descripteur de modèle ternaire local invariant (SILTP) pour description de la texture invariante de l'éclairage. SILTP est un opérateur amélioré par rapport au modèle binaire local bien connu (LBP) [46].

En fait, LBP a une belle propriété invariante sous des transformations monotones en Echelle de gris, mais elle n'est pas robuste à bruits d'images. SILTP améliore la LBP en introduisant une échelle tolérance de comparaison locale invariante, réalisation de l'invariance aux changements d'échelle d'intensité et à la robustesse aux bruits d'image.

2.4.2 Gérer les changements de point de vue

Nous avons appliqué la fonctionnalité Local Maximal Occurrence (LOMO), où l'algorithme Retinex est utilisé pour générer une image couleur perceptible par l'homme. Les caractéristiques LOMO sont construites par deux échelles de modèles ternaires Locaux invariants d'échelle (SILTP) [47], et les histogrammes de couleurs HSV sont dans des fenêtres coulissantes. Les caractéristiques sont extraites à trois échelles différentes en utilisant les mêmes procédures. Seule la valeur maximale de l'occurrence locale de chaque motif entre ces sous-fenêtres est conservée pour toutes les sous-fenêtres de taille 10×10 sur une même ligne d'image. L'histogramme atteint un niveau élevé d'invariance pour visualiser les changements tout en capturant les caractéristiques locales d'une personne. La figure 5 montre la procédure d'extraction des fonctionnalités LOMO [48].

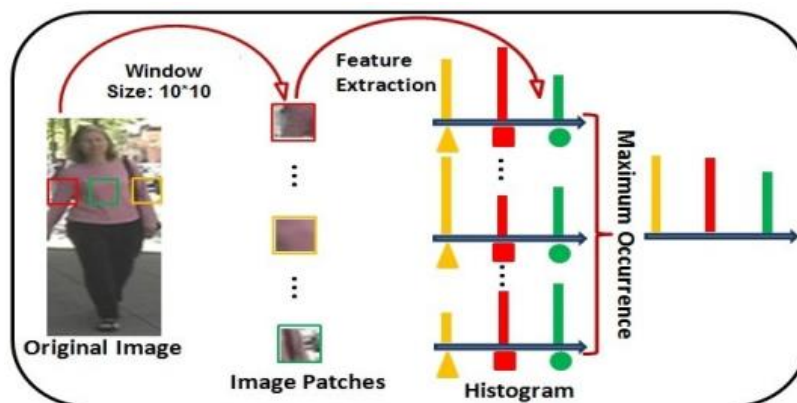


Figure 2.3: Illustration de la méthode d'extraction de fonctionnalités LOMO.

2.5. Rôle de Retinex

Nous avons comparé la fonctionnalité LOMO proposée avec et sans le prétraitement Retinex, avec les résultats présentés sur la figure 2.4(a) et (b). Cette comparaison a été réalisée en utilisant la mesure directe de similarité cosinus et l'algorithme XQDA respectivement. De la Figure (2.4(a)), nous pouvons voir que, pour une correspondance, les performances peuvent évidemment être améliorées par en appliquant la transformation Retinex, avec une précision de Rang1 étant sans Retinex et 20,25 % avec Retinex % 12,97%. Le résultat indique que Retinex aide à dériver une couleur cohérente.

Le résultat indique que Retinex aide à dériver une couleur cohérente ou une représentation à travers différentes vues de caméra, comme cela peut également être observé sur la figure 2.4 (b).

Cependant, d'après la Figure 2.4 (b), il On peut voir que les deux fonctionnalités sont améliorées par XQDA à une aire. Cela peut indiquer que XQDA est capable pour apprendre une métrique robuste contre les variations d'éclairage.

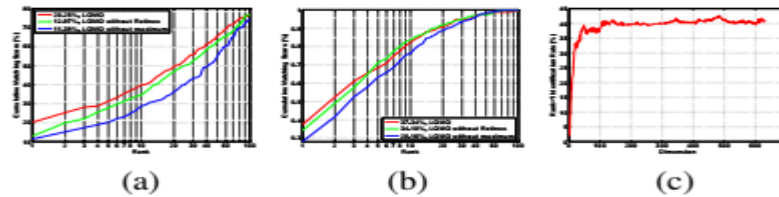


Figure 2.4. Courbes CMC comparant la fonction proposée avec et sans Retinex et l'opération d'occurrence maximale locale ((a) Cosinus et (b) XQDA). (c) Précision au rang 1 avec des dimensions de sous-espaces variables pour l'algorithme XQDA avec la fonction LOMO.

2.6. Rôle de LOMO

Les performances de réidentification des personnes sont largement affectées par les changements de point de vue, qui doivent être pris en compte dans conception de fonctionnalités ou apprentissage de classificateurs. Le local proposé L'extraction de caractéristiques d'occurrence maximale est une stratégie visant à L'extraction de caractéristiques d'occurrence maximale est une stratégie visant à représentation de caractéristiques robustes en pose ou en point de vue. En comparant la fonctionnalité proposée avec et sans l'opération d'occurrence maximale locale, nous constatons que cette opération ne améliorent considérablement les performances de réidentification des personnes à vue croisée, comme le montrent les figures (2.4 (a et b)). Sans l'opération d'occurrence maximale locale, la précision de rang 1 en appliquant la mesure de similarité cosinus (Figure 2.4 (a)) est en appliquant cette stratégie, la précision de Rang1 est 11,39 %, amélioré à 20,25%. Lors de l'application ultérieure de XQDA (Figure 2.4(b)), l'écart de performances est réduit, mais la fonctionnalité proposée fonctionne toujours bien mieux avec l'opération d'occurrence maximale locale.

2.7. Algorithme XQDA

Habituellement, les dimensions d de l'entité d'origine sont grandes et un L'espace de faible dimension R_r ($r < d$) est préféré pour la classification. Ont suggéré de décomposer Σ_I et Σ_E séparément. Pour réduire les dimensions. Puis Σ_I et Σ_E ont été estimés dans le sous-espace PCA. Cependant les deux méthodes ne sont pas optimales car la réduction de dimension ne prend pas en compte l'apprentissage des métriques à distance. Dans cet article, nous étendons la face bayésienne et KISSME approches d'apprentissage métrique à vues croisées, où nous envisageons d'apprendre un sous-espace $W = (w_1, w_2, \dots, w_r)$ avec des données de vue croisée, et en même temps apprendre une distance fonction dans le sous-espace dimensionnel r pour la vue croisée mesure de similarité. Supposons que nous ayons une formation multi-vues ensemble $\{X, Z\}$ de classes c , où $X = (x_1, x_2, \dots, x_r)$ contient n échantillons dans un espace à d dimensions de une vue, $Z = (z_1, z_2, \dots, z_r) \in R^{d \times m}$ contient m échantillons dans le même espace à d - dimensions mais provenant de l'autre voir. Le problème de la correspondance entre vues découle de nombreuse application, comme la reconnaissance faciale hétérogène et réidentification de personne invariante du point de vue. Noter que Z est identique à X dans le scénario de correspondance à vue unique. Considérant un sous-espace W . Dans le sous-espace dimensionnel r est calculé comme

$$d_w(X, Z) = (X, Z)^T W \left(\sum_I^{l-1} - \sum_E^{l-1} \right) W^T (X - Z) \quad (2.7)$$

où $\Sigma'_I = W^T \Sigma_I W$ et $\Sigma'_E = W^T \Sigma_E W$, Donc $M(W) = W (\Sigma'_I - \Sigma'_E) W^T$. Nous devons apprendre une matrice de noyau Cependant, optimiser directement d_W est difficile car W est contenu dans deux matrices inverses. Rappelons que Ω_I et Ω_E ont une moyenne nulle, alors étant donné sur une base w , les échantillons projetés des deux classes seront toujours centré sur zéro, mais peut avoir des écarts différents, comme illustré sur la figure 3. Dans ce cas, le critère traditionnel de Fisher utilisé pour dériver la LDA n'est plus adapté car les deux classes ont la même moyenne. Cependant, les variances σ_I et σ_E peuvent encore être utilisées pour distinguer les deux des classes.

Par conséquent, nous pouvons optimiser la direction de projection w telle que. Cependant, optimiser directement d_W est difficile car W est contenu dans deux matrices inverses Rappelons que Ω_I et Ω_E ont une moyenne nulle, alors étant donné sur une base w , les échantillons projetés des deux classes seront toujours centré sur zéro, mais peut avoir des écarts différents, comme illustré sur la figure(3). Dans ce cas, le critère traditionnel de Fisher utilisé pour dériver la LDA n'est plus adapté car les deux classes ont la même moyenne [45]. Cependant, les variances σ_I et σ_E peuvent encore être utilisées pour distinguer les deux des classes. Par conséquent, nous pouvons optimiser la direction de projection W telle que $\sigma_E(W)/\sigma_I(W)$ est maximisée. Remarquez que $\sigma_I(W) = W^T \Sigma_I W$ et $\sigma_E(W) =$

$W^T \sum_E W$, Donc l'objectif $\sigma_E(W)/\sigma_I(W)$ correspond au Généralisé Quotient de Rayleigh

$$J(W) = \frac{W^T \sum_E W}{W^T \sum_I W} \quad (2.8)$$

La maximisation de $J(W)$ est équivalente à

$$\max_w W^T \sum_E W, \text{ s. t. } W^T \sum_I W = 1 \quad (2.9)$$

Propres qui peuvent être résolu par le problème de décomposition généralisée des valeurs similaire à LDA. C'est-à-dire le plus grand LDA. C'est-à-dire la plus grande valeur propre de $\sum_I^{-1} \sum E$ est la valeur maximale de $J(W)$ Et le vecteur propre correspondant W_1 est la solution. De plus, la solution Orthogonale à W_1 et correspondant à la deuxième plus grande valeur de $J(W)$ est le vecteur propre de deuxième plus grande valeur propre de $\sum_I^{-1} \sum E$ et ainsi de suite par conséquent, avec $W = (w_1, w_2, \dots, w_r)$ on apprend un sous-espace discriminant, ainsi qu'une fonction de distance dans le sous-espace appris. Nous appelons l'algorithme dérivé Crossview Quadratic Discriminant Analysis (XQDA) pour refléter sa connexion à QDA et sortie d'une métrique à vues croisées

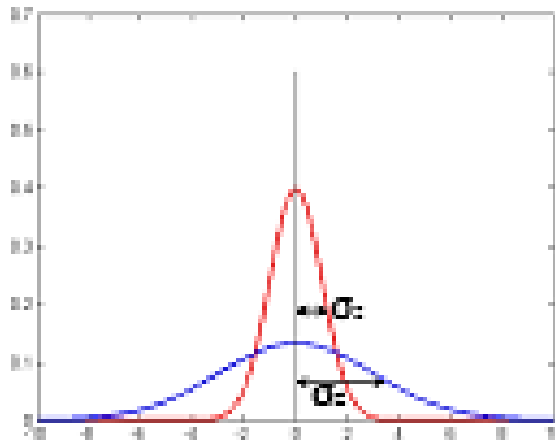


Figure 2.5: Distributions of Ω_I and Ω_E in one projected dimension.

2.8. La métrique de distance de Mahalanobis

La distance de Mahalanobis est une méthode classique pour mesurer la distance ou la similarité dans la réidentification de personnes. En ne supposant que le vecteur $v = [x_1, x_2, \dots, x_n]$

$$d_M(x_i, x_j) = \sqrt{(x_i - x_j)^T M (x_i - x_j)} \quad (2.10)$$

Où la matrice semi-définie positive $M \in R^{d \times d}$ est la matrice de métrique de Mahalanobis à résoudre. La métrique de Mahalanobis classique est facile à sur-ajuster lorsque les échantillons d'entraînement sont limités. Pour surmonter ce problème de sur-ajustement, Qi et al. [49] ont proposé une matrice de métrique régularisée indépendante. Contrairement à l'algorithme de régularisation lisse [50], la méthode proposée a appris la matrice de métrique séparément dans quatre espaces colorimétriques, puis la matrice de métrique de Mahalanobis indépendante régularisée a été obtenue par régularisation de quatre matrices de métrique respectivement. Zhou et al. [51] ont construit des ensembles de données de référence basés sur trois principes différents et ont défini une nouvelle contrainte de référence. Sur la base de cette contrainte, une matrice de métrique de distance de Mahalanobis optimisée améliorée a été définie. Outre les méthodes classiques de métrique de Mahalanobis et les méthodes de métrique de Mahalanobis améliorées, selon les difficultés pratiques de la réidentification de personnes, certaines nouvelles méthodes de métrique ou fonctions de similarité sont proposées, les quatre sections suivantes les décrivent en détail.

2.9. Conclusion

Dans ce chapitre, nous avons examiné les approches proposées pour la réidentification 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 : Résultat & discussion

3.1. Introduction

La plupart des méthodes utilisées pour la ré-identification des personnes reposent sur l'apprentissage métrique et la représentation des caractéristiques. De nombreux efforts ont été déployés dans les deux directions pour relever le défi de la ré-identification des personnes. Une représentation des caractéristiques efficace doit être robuste aux variations d'éclairage et de points de vue, et une métrique discriminante doit être apprise pour correspondre aux différentes images de personnes.

Dans ce chapitre, nous discuterons, évaluerons et comparerons les performances des systèmes de réidentification des personnes utilisant les caractéristiques "Local Maximal Occurrence" (LOMO) et "Gaussian Of Gaussian" (GOG), en les combinant avec la méthode efficace d'apprentissage métrique XQDA. Nous avons mené ces expériences en utilisant différentes configurations et critères sur la base de données VIPeR pour évaluer leur efficacité.

3.2. Bases de données VIPeR

La base de données VIPeR est une base de données de réidentification de personnes particulièrement difficile, largement utilisée pour l'évaluation de référence. Il a été capturé à partir de deux vues de caméra. Il contient 1264 images de 632 identités. Chaque paire contient deux images du même individu vues sous des angles différents. Chaque paire d'images a été prise à partir d'un point de vue arbitraire sous des conditions d'éclairage variables. Les données ont été collectées dans un cadre académique sur plusieurs mois. Chaque image a été redimensionnée à 128x48 pixels.

La Figure (3.1) montre des paires d'images d'échantillon du jeu de données VIPeR. Le protocole expérimental largement adopté pour cette base de données consiste à diviser aléatoirement les 632 paires d'images en deux moitiés, l'une pour l'entraînement et l'autre pour les tests, et à répéter la procédure 10 fois pour obtenir une performance moyenne [52].



Figure 3 .1: Paires d'images d'échantillon du jeu de données VIPeR.

3.3. Mesure de performance

La performance d'une approche est mesurée par la courbe CMC (pour «Cumulative Match Characteristic»). Comme on compte ré-identifier une personne requête parmi N personnes référence, la courbe CMC représente la probabilité de trouver la bonne personne référence parmi les r meilleurs appariements. R est appelé le rang de ré-identification. Au rang 1, la valeur de la courbe CMC est appelée CCR (pour «Correct Classification Rate»). Autrement dit, la courbe CMC donne le pourcentage des personnes ré-identifiées en fonction du rang de ré-identification. La courbe CMC est croissante et converge vers la valeur 100% de taux de ré-identification (figure 3.2).

L'importance du CMC réside dans :

- Les courbes CMC sont faciles à comprendre et offrent une représentation visuelle claire de la performance de l'algorithme.
- Elles fournissent une vue d'ensemble des performances du système à travers différents rangs, aidant à évaluer les points forts et faibles d'un système de ré-identification.
- Elles sont utilisées comme référence pour évaluer et comparer de manière fiable les performances de différents systèmes de ré-identification.

Grâce au CMC, les développeurs et les chercheurs peuvent améliorer leurs algorithmes pour offrir de meilleures performances en ré-identification des personnes, ce qui renforce la sécurité et la sûreté dans diverses applications.

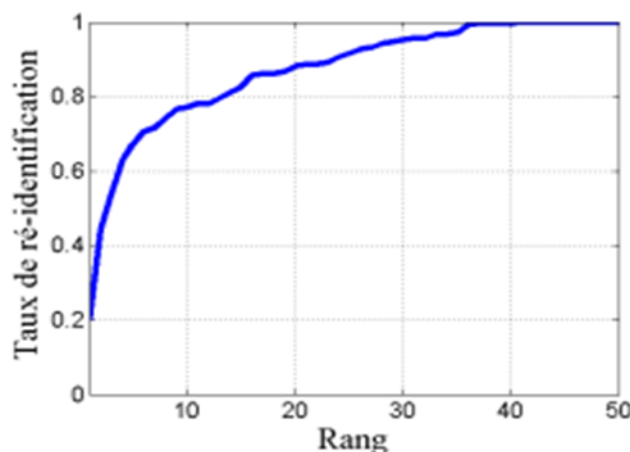


Figure 3.2 Exemple d'une courbe CMC.

3.4. Réalisation et résultats

Notre objectif est d'évaluer les performances entre les méthodes LOMO et GOG dans la réidentification des personnes et de déterminer laquelle est la meilleure dans ce domaine. Pour atteindre notre objectif, nous allons intégrer ces deux méthodes avec l'approche d'apprentissage de métrique XQDA et utiliser la distance de Mahalanobis comme mesure pour atteindre notre objectif. Les résultats sont indiqués ci-dessous.

Rank : Dans la Re-ID humaine, un humain compare un utilisateur de recherche à un ensemble de personnes dans une galerie pour produire une liste classée selon leur similarité de correspondance, en supposant généralement que la correspondance correcte est attribuée au rang le plus élevé, idéalement le Rank 1.

La réidentification de la personne consiste à rechercher dans la galerie une identification correcte correspondre à la sonde donnée (Figure3.3).



Figure3.3 : Person re-id consiste à rechercher dans la galerie une correspondance correcte avec la sonde donnée

3.4.1 Analyse des performances de LOMO+XQDA

Le tableau (3.1) montre les performances de la combinaison de la fonctionnalité LOMO avec l'algorithme XQDA à travers différentes dimensions, en affichant les résultats pour différents Ranks.

Tableau 3.1 : Analyse des performances de LOMO+XQDA. (en %)

Rank1%	Rank5%	Rank10%	Rank15%	Rank20%	XQDA Dims
36,93	67,47	80,16	87,09	91,33	50
37,82	67,53	80,38	87,31	91,71	60
37,59	67,88	80,82	87,28	91,46	70
37,97	67,94	80,47	86,96	91,74	80
38,64	68,10	80,44	87,18	91,61	90
38,58	68,10	80,41	87,22	91,58	100
39,15	67,88	80,28	87,56	91,58	110
39,21	67,47	80,16	87,31	91,49	120
39,27	68,07	80,38	87,28	91,52	130
39,37	67,78	80,44	87,41	91,49	140
39,27	68,10	80,47	87,09	91,39	170
39,53	68,07	80,32	87,22	91,20	190
39,53	67,69	80,44	86,90	91,36	210
39,78	67,94	80,41	87,15	91,30	230
39,78	67,65	80,44	87,22	91,20	250
39,81	67,88	80,32	87,31	91,20	270
39,91	68,10	80,41	87,41	91,23	290
40,00	68,04	80,54	87,31	91,14	300
40.06	68.01	80.47	87.47	91.11	310
40,00	68,13	80,51	87,37	91,08	320
40,00	68,13	80,51	87,37	91,08	330

Discussions : Le système LOMO+XQDA montre une performance remarquable et stable à travers une large gamme de dimensions. Il est notable que l'utilisation des dimensions comprises entre 50 et 120 offre un bon équilibre entre précision et complexité computationnelle, en atteignant des résultats remarquables pour les rangs de 1% et 5%. Au-delà de ce point, les améliorations deviennent minimales, ce qui indique que l'augmentation importante des dimensions peut ne pas être nécessaire pour atteindre une performance plus élevée.

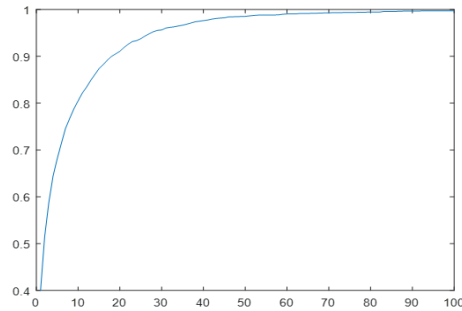


Figure 3.4 courbe CMC de LOMO+XQDA .

3.4.2 Analyse des performances de GOG+XQDA

Le tableau (3.2) présente les performances de l'intégration de la fonctionnalité GOG avec l'algorithme XQDA à travers différentes dimensions, en affichant les résultats pour différents ranks.

Tableau 3.2 : Analyse des performances de GOG+XQDA(en %)

Rank1%	Rank5%	Rank10%	Rank15%	Rank20%	XQDA Dims
33.51	63.64	75.76	82.22	86.80	50
34.08	63.16	75.82	82.12	86.55	60
33.51	62.91	75.32	81.77	86.42	70
34.05	63.04	75.38	81.65	86.01	80
34.81	63.35	75.32	81.68	85.82	90
34.08	63.16	75.22	81.55	85.79	100
33.39	63.10	75.25	81.30	85.32	110
33.20	63.01	75.13	81.08	85.35	120
33.62	62.94	74.84	80.89	85.41	130
33.83	62.91	74.59	80.79	85.28	140
33.39	62.41	74.65	80.47	85.00	150
33.80	62.18	74.34	80.38	84.56	170
32.88	62.15	74.21	80.28	84.53	190
32.82	62.25	74.05	80.03	84.59	200
33.07	61.74	73.86	79.94	84.24	220
32.75	61.71	73.58	79.87	84.27	240
32.16	61.49	73.07	79.56	83.89	260
33.16	61.58	73.07	79.40	83.80	280
32.85	61.84	72.88	79.40	83.92	300
32.63	61.61	72.97	79.05	83.96	320

Discussions : L'intégration de la fonctionnalité GOG avec l'algorithme XQDA montre des performances solides mais pas aussi élevées que l'utilisation de XQDA seul.

Il est crucial de sélectionner soigneusement le nombre de dimensions pour optimiser les performances sans introduire une complexité computationnelle excessive. Les dimensions entre 50 et 100 semblent offrir un bon compromis entre précision et efficacité computationnelle pour le système GOG+XQDA.

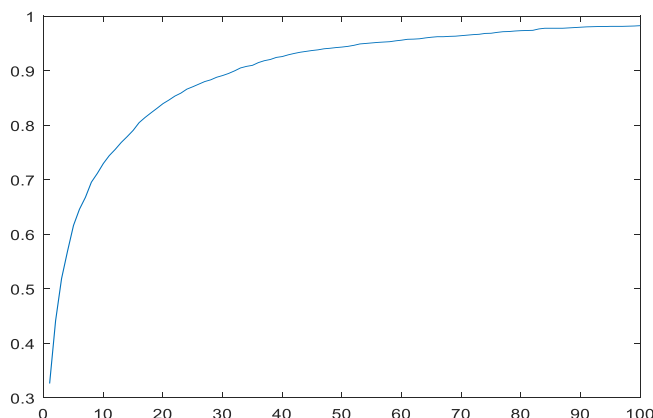


Figure 3.5 courbe CMC de GOG+XQDA

Après avoir identifié les deux méthodes LOMO+XQDA et GOG+XQDA, nous avons consigné le meilleur résultat obtenu dans chaque cas dans le tableau (3.3).

Tableau 3.3: comparaison de "LOMO+ XQDA" et "GOG+ XQDA"

	Rank1%	Rank5%	Rank10%	Rank15%	Rank20%	XQDA Dims
GOG+ XQDA	34.81	63.61	75.32	81.68	85.82	90
LOMO+ XQDA	40.06	68.01	80.47	87.47	91.11	310

Discussions : Lors de la comparaison des méthodes LOMO+XQDA et GOG+XQDA sur la base de données VIPeR, nous constatons les points suivants : au Rang 1, LOMO+XQDA montre une performance significativement meilleure avec un taux de 40.06% comparé à 34.81% pour GOG+XQDA. Au Rang 5, LOMO+XQDA continue de surpasser GOG+XQDA avec un taux de 68.01% contre 63.61%. Pour le Rang10, LOMO+XQDA atteint un taux de 80.47% tandis que GOG+XQDA obtient 75.32%. Au rang 15%, LOMO+XQDA obtient un taux de 87.47% comparé à 81.68% pour GOG+XQDA. Enfin, au rang 20%, LOMO+XQDA obtient le meilleur résultat avec un taux de 91.11% contre 85.82% pour GOG+XQDA.

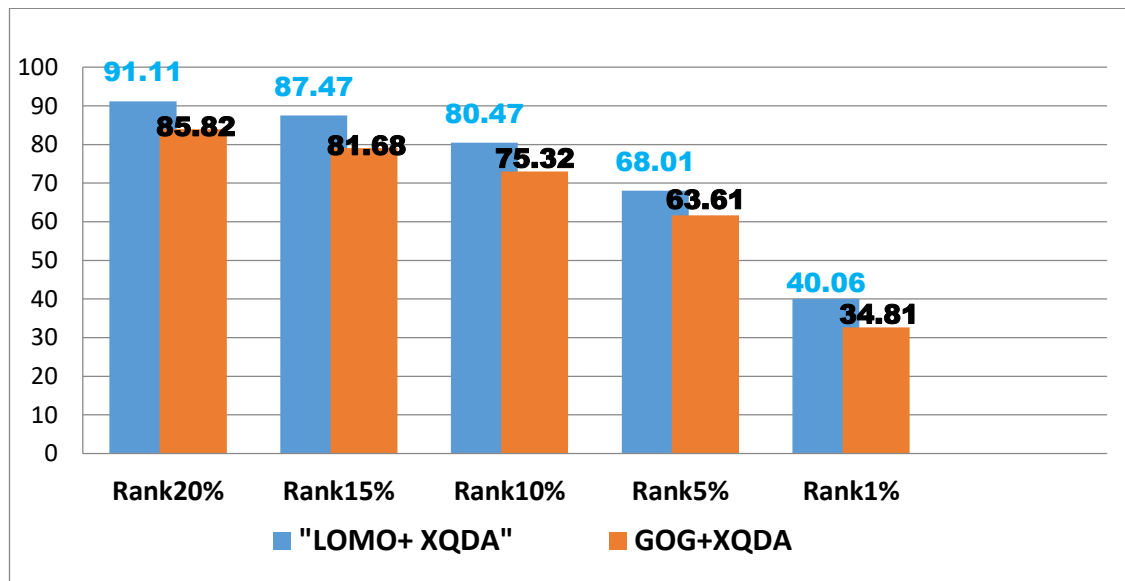


Figure 3.6 Diagramme des résultats de la comparaison entre "LOMO+ XQDA" et "GOG" pour Rank1(%), sur la base de données VIPeR.

En se basant sur les résultats présentés dans le tableau (3.3), on peut conclure que la méthode LOMO+XQDA (Rank1 =40.06%, pour XQDA Dims=310) offre généralement de meilleures performances que GOG+XQDA (Rank1=34.81, pour XQDA Dims=90) à travers tous les rangs évalués. Cela est particulièrement notable aux rangs inférieurs, où les performances de LOMO+XQDA sont significativement meilleures, indiquant une capacité supérieure de cette méthode à identifier correctement l'identité avec une précision accrue.

En explorant davantage les dimensions et les paramètres, il pourrait être possible d'améliorer les performances de chaque méthode. Par exemple, ajuster les paramètres de XQDA pourrait conduire à des gains supplémentaires en termes de précision. Une approche future prometteuse pourrait être de fusionner les méthodes LOMO et GOG pour tirer parti de leurs points forts respectifs. En fusionnant les caractéristiques des deux descripteurs avant de les passer à XQDA, il pourrait être envisagé de créer un modèle hybride offrant globalement de meilleures performances.

Performances de LOMO+XQDA sont significativement meilleures, indiquant une capacité supérieure de cette méthode à identifier correctement l'identité avec une précision accrue.

3.5 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-identification des personnes.

Enfin, pour évaluer nos résultats, nous avons présenté les courbes de caractéristiques cumulatives de correspondance et les avons comparées avec les recherches les plus récentes dans ce domaine.

Conclusion générale

Conclusion générale

La réidentification des personnes est un nouveau domaine développé pour garantir la sécurité et la sûreté publiques. Nous pouvons suivre le parcours d'une personne et nous assurer qu'aucune activité illégale ou inappropriée n'est effectuée en utilisant des modèles de réidentification et de suivi. La réidentification est un problème difficile car les personnes sont souvent capturées par des caméras de faible résolution, dans des conditions d'occultation, avec une mauvaise luminosité et dans diverses postures. Dans ce cas, une modélisation robuste de l'apparence complète du corps de la personne est nécessaire, surtout lorsque d'autres indicateurs biométriques traditionnels ne sont pas disponibles ou impossibles à capturer en raison des limitations de résolution des capteurs ou de fréquence d'images. Les chercheurs essaient d'améliorer les performances du système de réidentification des personnes en développant une représentation de caractéristiques robuste et un apprentissage métrique discriminant.

L'objectif de notre travail est d'évaluer les performances des méthodes Local Maximal Occurrence (LOMO) et Hierarchical Gaussian Descriptor (GOG) en termes de rapidité et de précision pour les systèmes de réidentification des personnes en les combinant chacune avec la méthode d'apprentissage métrique XQDA. Notre projet peut être résumé comme suit :

- ✓ Une description générale des biométries, des méthodes biométriques et de leurs applications.
- ✓ Un examen de la littérature sur la réidentification des personnes avec son pipeline, les défis et les applications.
- ✓ La représentation des caractéristiques pour la réidentification des personnes et l'introduction de deux méthodes : Local Maximal Occurrence (LOMO) et Hierarchical Gaussian Descriptor (GOG), en explorant leurs concepts, leurs principes de fonctionnement et leurs contributions à l'amélioration de la tâche de réidentification des personnes.
- ✓ L'utilisation de l'approche d'apprentissage métrique XQDA pour la combiner avec les caractéristiques LOMO et GOG individuellement afin d'obtenir les meilleurs résultats pour chaque descripteur pour la réidentification des personnes.

L'utilisation de la fonction de normalisation des scores sur la distance de Mahalanobis qui a montré un grand impact sur le processus de réidentification sur l'ensemble de

Données VIPeR. En raison des performances prometteuses des caractéristiques Local Maximal Occurrence (LOMO) et Hierarchical Gaussian Descriptor (GOG).

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 ;
- Utiliser la soft biométrie douce pour améliorer la Re-ID ;
- Fusion les deux méthodes LOMO et GOG.

Références bibliographiques

Références bibliographiques

- [1] J. Li, “Pose-guided representation learning for person re-identification”. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019.
- [2] O. Tuzel, F. Porikli, and P. Meer, “Pedestrian detection via classification on Riemannian manifolds”, IEEE Trans. Pattern Anal. Mach. Intell.,p:1713–1727, 2008.
- [3] N. HAMITOUCHE, Z. SALMI, “Système d’identification biométrique de personnes par reconnaissance de l’iris”, 2009.
- [4] F.Belhadj, Biometric “system for identification and authentication”, Ecole nationale Supérieure en Informatique Alger, 2017.
- [5] S. Masood Khan and M. Shah. “A multiview approach to tracking people in crowded scenes using a planar homography constraint”, *In European Conference on Computer Vision*, 2006.
- [6] D.R.Ibrahim, A. Tamimi, and A.M. Abdalla, “Performance analysis of biometric recognition modalities”, IEEE international conference on information technology (ICIT), 2017.
- [7] Z. Megyesi, “*Dense Matching Methods for 3D Scene Reconstruction from Wide Baseline Images*”. PhD thesis, Eotvos Lorand University, 2009.
- [8] Hassanien, A.E, “Proceedings of the International Conference on Advanced Intelligent Systems and Informatics”, 2016.
- [9] F. Karray, J. A. Saleh, M. N. Arab and M. Alemzadeh, “**Multi Modal Biometric Systems: A State of the Art Survey**”, International Conference on Computational Intelligence, New Zealand, Nov 2007.
- [10] A.K. Jain, R. Bolle and S. Pankanti, “**BIOMETRICS: Personal Identification in Networked society**”, Kluwer Academic Publishers, 1999.
- [11] Li. Ma, Y. Wang, T.Tieniu, “Personal *identification based on iris texture analysis*”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 25, (N° 12), December 2003,p: 1519-1533.
- [12] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C.Yan, Y. Yang, “Improving person reidentification by attribute and identity learning”, Science Direct, Pattern Recognition95, 2019.
- [13] Li. S, “Identity-aware textual-visual matching with latent co-attention”,2019.
- [14] M.Ye, “Specific person retrieval via incomplete text description”, InProceedings of the fifth ACM on International Conference on Multimedia Retrieval, 2015.

- [15] M.Ye, “Deep learning for person re-identification: A survey and outlook”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021.
- [16] S.Gong, T. Xiang, “Person re-identification, in *Visual Analysis of Behaviour*”, 2011, p: 301-313.
- [17] M.Farenzena .“Person re-identification by symmetry-driven accumulation of local features”, *IEEE computer society conference on computer vision and pattern recognition*, 2010.
- [18] L .Bazzani, “Multiple-shot person re-identification by chromatic and epitomic analyses”, *Pattern Recognition Letters*, 2012, p: 898-903.
- [19] T. Avraham, “Learning implicit transfer for person re-identification”, in *European Conference on Computer Vision*, 2012.
- [20] W .Zheng, S. Gong, T. Xiang, “Reidentification by relative distance comparison”. *IEEE transactions on pattern analysis and machine intelligence*, 2012, **35**(3): p. 653-668.
- [21] W.Li, R. Zhao, X. Wang, “Human reidentification with transferred metric learning”, in *Asian conference on computer vision*, 2012.
- [22] W, Zheng, S, Gong, and T. Xiang, “Associating Groups of People”, in *BMVC*, 2009.
- [23] C.Loy, T. Xiang, S. Gong, “Incremental activity modeling in multiple disjoint cameras”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, p: 1799-1813.
- [24] C.Ren, “Domain adaptive person re-identification via camera style generation and label propagation”, *IEEE Transactions on Information Forensics and Security*, 2019, p: 1290-1302.
- [25] J. Li, “Pose-guided representation learning for person re-identification”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.
- [26] A. Nambiar, A. Bernardino, and J.C. Nascimento, “Gait-based person reidentification: A survey”, *ACM Computing Surveys (CSUR)*, 2019, p: 1-34.
- [27] L.Zhang, “Facial expression analysis under partial occlusion”, a survey. *ACM Computing Surveys (CSUR)*, 2018, p: 1-49.
- [28] X. Wang and R .Zhao, “Person re-identification: System design and evaluation overview, in *Person Re-Identification*”, 2014, p: 351-370.
- [29] V.Suma, “Computer vision for human-machine interaction-review”, *Journal of trends in Computer Science and Smart technology (TCSST)*, 2019, **1**(02): p. 131-139.

- [30] G. Lisanti, I. Masi, A.D. Bagdanov, A.D. Bimbo, “Person re-identification by iterative re-weighted sparse ranking”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, P: 1629–1642.
- [31] S. Liao, “Person re-identification by local maximal occurrence representation and metric learning”, in *IEEE conference on computer vision and pattern recognition*, 2015.
- [32] Y.C. Chen, X. Zhu, W.S. Zheng, J.H. Lai, “Person re-identification by camera correlation aware feature augmentation”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2017.
- [33] S. Liao, S.Z. Li, “Efficient PSD constrained asymmetric metric learning for person re-identification”, in *IEEE International Conference on Computer Vision*, 2015, p:3685–3693.
- [34] M. Köestinger, M. Hirzer, P. Wohlhart, P.M. Roth, H. Bischof, “Large scale metric learning from equivalence constraints”, in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, p: 2288–2295.
- [35] B. Ma, Q. Li, and H. Chang, “Gaussian descriptor based on local features for person re-identification”, In *Asian Conference on Computer Vision (ACCV) Workshop*, 2014, pages 505–518.
- [36] H. Košice, “Person Re-Identification: Attribute-Based Feature Evaluation”, *IEEE 16th World Symposium on Applied Machine Intelligence and Informatics*, 2018.
- [37] R. Satta, “Appearance descriptors for person re-identification: a comprehensive review”, *CoRR*, 2013, p: 1307.5748.
- [38] T. Matsukawa, T. Okabe, E. Suzuki and Y. Sato, “Hierarchical gaussian descriptor for person re-identification”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [39] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu, “Freeform region description with second-order pooling”, *IEEE Trans. Pattern Anal. Mach. Intell.*, 2015, p: 1177–1189.
- [40] S. Amari and H. Nagaoka, “Methods of Information Geometry. Volume 191 of Translations of mathematical monographs”, *American Mathematical Society*, 2001.
- [41] P. Li, Q. Wang, and L. Zhang, “A novel earth mover’s distance methodology for image matching with Gaussian mixture models”, *IEEE International Conference on Computer Vision (ICCV)*, 2013, p: 1689–1696.
- [42] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li, “Salient color names for person re-identification”, In *European Conference on Computer Vision (ECCV)*, 2014, p: 536–551.
- [43] J. Sanchez, F. Perronnin, T. Mensink, and J. J. Verbeek, “Image classification with the Fisher vector: Theory and practice”, *International Journal of Computer Vision*, 2013, p: 222–245.

- [44] H. Jegou, O. Chum, “Negative evidences and co- occurrences in image retrieval: The benefit of PCA and whitening”, In European Conference on Computer Vision (ECCV), 2012, p: 774–787.
- [45] Z. Henriques, J. F., “*Exploiting the circulant structure of tracking-by-detection with kernels*”, In European conference on computer vision. Springer, 2012.
- [46] D. J. Jobson, Z.-U. Rahman, G. A. Woodell, “A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing*”, *IEEE Transactions on*, 1997, p: 965–976.
- [47] S. Liao, G. Zhao, V. Kellokumpu, M. Pietikainen, S. Z., “Li. Modeling pixel process with scale invariant local patterns for background subtraction in complex scenes”, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, USA, June 2010.
- [48] Moussi, ouamane, ouafi, “person re-identification using soft biometrics”, 2023.
- [49] Q. Meibin, W. Yunxia, T. Shengshun, “Person re-identification based on regularization of independent measure matrix”. *Pattern Recognition and Artificial Intelligence*, 2016, p: 511–518.
- [50] T. DP, J. LW, Y. Wang, “Person re-identification by regularized smoothing KISS metric learning”, *IEEE Trans on Circuits and Systems for Video Technology* , 2013, p: 1675–1685.
- [51] J. Zhou, Su. B., “multi-shot person re-identification from reference constraints”, *IEEE Conference on Computer Vision*, 2018, p: 5373–5381.
- [52] S. Liao, “Person re-identification by local maximal occurrence representation and metric learning”, *IEEE conference on computer vision and pattern recognition*, 2015.