PEOPLE'S DEMOCRATIC REPUBLIC OF ALGERIA

MINISTRY OF HIGHER EDUCATION AND SCIENTIFIC RESEARCH

YAHIA FARES UNIVERSITY OF MEDEA



Faculty of **T**echnology

Department of **E**lectrical **E**ngineering

Master's End of Study Project

Branch: Telecommunications.

Option: Telecommunications Systems.

Multi-dimensional features for Person Re-identification based tensor analyses

Presented by:

RABAI Hiba

ABDELLATIF Marwa

Proposed and supervised by:

Dr. CHOUCHANE Ammar

Members of the committee	
Dr. ESSAID Bilel	**President**
Dr. KHEDDAR Hamza	**Examiner**
Dr. CHOUCHANE Ammar	**Supervise**

Promotion 2022

Dedication

I dedicate this modest work:

To my family, who gave me a dignified education, her love has made me what I am today:

To my dear mother HADDA Whatever I do or say, I cannot thank you as it should be your affection covers me, and your presence at my side has always been my source of strength to face the various obstacles.

To my dear father BENYOUCEFyou have always been by my side to support and encourage me during these years of study.

To my husband, for his patience, love, support and encouragement.

To my lovely sisters IMANE and RACHA.

To my brother BADR EDINE. To my dear partner MARWA.

To all the teachers who help me from near or far.

HIBA RABAI

I dedicate this modest work

To my dear parents, for all their sacrifices, love, tenderness, support and prayers throughout my studies, To my dear sisters, for their constant encouragement and moral support, To my dear brothers, for their support and encouragement, To all my family for their support throughout my academic career.To my partner HIBA. Thank you for always being there for me.

MARWA ABDELLATIF

Acknowledgments

We thank God for giving us the courage, strength and will to make this work.

We would like to thank Mr. CHOUCHANE Ammar for his assistance, encouragement and guidance.

And we would like to thank the team ofLI3C laboratory at BISKRA University.

Many Thanks go to the teachers of Electrical Engineering department of Medea University.

We hope this brief will be helpful for all future members.

Abstract

Person Re-identification (PRe-ID) is among the main constituents of an automated visual surveillance system. It is defined as the process of determining whether a target person has already appeared on a camera network Re-ID is a challenging task due to variations in lighting and viewpoint, pose and occlusions.

In this work, we are going to use two effective feature representation called convolutional neural network (CNN) and Local Maximal Occurrence (LOMO). Our principal contribution is how to fuse CNN and LOMO features using high order tensor representation in order to improve the accuracy of our approachand metric learning method called Tensor Cross-view Quadratic Analysis (TXQDA) with the mahalanobis distance. Then we will apply a three challenging person re-identification datasets VIPeR, GRID, and PRID450S.

Keywords: Person Re-identification, Tensor representation of data, CNN, LOMO, TXQDA andMahalanobis distance.

Résumé

La ré-identification des personnes (PRe-ID) fait partie des principaux éléments d'un système de surveillance visuelle automatisé. L'identification de la personne est définie comme le processus consistant à déterminer si une personne cible est déjà apparu sur un réseau de caméras. L'identification de la personne est une tâche difficile en raison des variations d'éclairage et de point de vue, de pose et d'occlusions.

Dans ce travail, nous allons utiliser deux représentations efficaces appelées réseau de neurones convolutionnels (CNN) et occurrence maximale locale (LOMO). Notre principale contribution est de savoir comment fusionner les fonctions CNN et LOMO en utilisant la représentation d'un tenseur d'ordre élevé afin d'améliorer la précision de notre approche et de notre méthode d'apprentissage métrique appelée Tensor Cross-view Quadratic Analysis (TXQDA) avec la distance mahalanobis. Ensuite, nous appliquerons trois ensembles de données de ré-identification de personnes difficiles, VIPeR, GRID et PRID450S.

Les mots-clés :Ré-identification de personne, Tensorreprésentation des données, CNN, LOMO, TXQDA et Mahalanobis distance.

ملخص

إعادة تحديد هوية الأشخاص (Re-Id) هي من بين المكونات الرئيسية لنظام المراقبة البصرية الآلي. يتم تعريف إعادة تعريف الشخص على أنه عملية تحديد ما إذا كان الشخص المستهدف قد ظهر بالفعل على شبكة الكاميرا Re-ID هي مهمة صعبة بسبب الاختلافات في الإضاءة ووجهة النظر والوضع والانسداد.
في هذا العمل، سنستخدم ميزتين فعالتين تسمى الشبكة العصبية التلافيفية (CNN) والحدوث المحلي الأقصى (LOMO). تتمثل مساهمتنا الرئيسية في كيفية دمج ميزات CNN و LOMO باستخدام تمثيل موتر عالي المستوى من أجل تحسين دقة نهجنا وطريقة التعلم المتري المسماة تحليل تربيعي لعرض الموتر (TXQDA) مع مسافة mahalanobis. ثم سنطبق ثلاث مجموعات بيانات صعبة لإعادة تحديد هوية الشخص VIPeR و GRID و PRID450S.

الكلمات المفتاحية: إعادة تحديدهويةالشخص, , مسافة mahalanobis , تمثيل بيانات الموترTXQDA, الشبكة العصبية التلافيفيةCNN و الحدوث المحلي الأقصىLOMO.

Table of content

Dedication.....I
Acknowledgments.....II
Abstract.....III
Résumé.....IV
ملخص.....V
Table of content.....VI
List of figures.....IX
List of tables.....X
List of abbreviations.....X
General introduction.....1

Chapter I Generalities of Biometrics and person Re-identification

I.1. Introduction.....4
I.2. Biometrics.....4
I.3. Biometric systems' modalities.....4
I.4. Steps of biometric system.....6
I.5. Applications of biometric systems.....6
I.5.1. Person Re-identification system.....7
I.5.2. Applications of person Re-identification.....8
I.5.3. Challenges of person Re-identification.....9
I.6. Crosse validation protocol-based CMC.....11
I.7. Conclusion.....12

Chapter II Person Re-identification: Algorithms and Methods

II.1. Introduction.....14
II.2. Artificial intelligence.....14
II.2.1. Application of AI.....14
II.2.2. Challenges of AI.....16

II.3. Machine learning 18
II.3.1. Classification of machine learning methods 18
II.3.1.1. Supervised learning 18
II.3.1.2. Unsupervised learning 20
II.3.1.3. Reinforcement learning 22
II.4. Transfer learning based deep learning 22
II.5. Feature extraction in Person Re-identification 23
II.5.1. Convolutional neural network CNN based Alexnet architecture 23
II.5.2. LOMO Descriptor 25
II.6. Conclusion 26

Chapter III Implementation and results

III.1. Introduction 28
III.2. Person Re-identification datasets 28
III.3. Overview of the proposed approach 30
III.4. TXQDA multilinear subspace projection 32
III.5. Results analysis 33
III.5.1. Protocol and parameter settings 33
III.5.2. Discussions 36
III.5.3. Comparison with the stat of the art 41
III.6. Conclusion 43
General conclusion 44
References 45

List of figures

Figure I. 1: Various types of biometrics. 4
Figure I. 2: Person Re-identification system. 8
Figure I. 3: Example of some person Re-identification challenges. 9
Figure I. 4: Example of cross validation model. 11
Figure I. 5: 10-fold cross-validation. 12

Figure II. 1 : Computer vision and its sub-concepts 15
Figure II. 2 : Supervised learning workflow 19
Figure II. 3 : Decision tree 19
Figure II. 4 : K-mean clustering algorithm 21
Figure II. 5 : Simple neural network exemple 23
Figure II. 6 : The convolution operation 24
Figure II. 7 : Example of 2x2 max pooling 24
Figure II. 8 : The architecture of AlexNet 25
Figure II. 9 : Feature extraction for LOMO descriptor at image-level 26

Figure III. 1: Part of VIPeR dataset. 29
Figure III. 2: Random images from GRID. 29
Figure III. 3: Sample image pairs from the PRID450S dataset. 30
Figure III. 4: Person Re-identification system. 31
Figure III. 5: CNN 3rd order tensor design. 33
Figure III. 6: LOMO 3rd order tensor design. 34
Figure III. 7: The proposed features fusion. 35
Figure III. 8: CMC curve of the best features on VIPeR. 37
Figure III. 9: CMC curve of the best features on GRID. 38
Figure III. 10: CMC curve of the best features on PRID450s. 40
Figure III. 11: The Rank-1 performance on the three datasets VIPeR, GRID and PRID450s. 40

List of tables

Table III. 1: Datasets and their characteristics……..28
Table III. 2 :CMC scores (%) of the used databases VIPeR in CNN and LOMO……36
Table III. 3 : CMC scores (%) of the used databases VIPeR in CNN+LOMO……37
Table III. 4 : CMC scores (%) of the used databases GRID in CNN and LOMO……38
Table III. 5 : CMC scores (%) of the used databases GRID in CNN+LOMO…….38
Table III. 6 : CMC scores (%) of the used databases PRID450s in CNN and LOMO……39
Table III. 7 : CMC scores (%) of the used databases PRID450s in CNN+LOMO…….39
Table III. 8: Comparison with the state-of-the-art of Rank-1 and Rank-20 identification rates (%) on the PRID450s, GRID and VIPeR databases…….42

List of abbreviations

AI: Artificial Intelligence.

CMC: Cumulative Match Characteristic Curve.

CNN: Convolutional Neural Network.

DL: Deep Learning.

IoT: Internet of Things.

LOMO: Local Maximal Occurrence.

LRN: Local Response Normalization.

ML: Machine Learning.

NLP: Natural Language Processing.

PCA: Principal Component Analysis**.**

PRe-ID: Person re-identification.

ReLUs: Rectified Linear Units.

SILTP: Scale Invariant Local Ternary Patter.

SVM: Support Vector Machine.

TXQDA: Tensor Cross-view Quadratic Analysis

General introduction

In the last decade, the scientific community of computer vision, machine learning (ML) and artificial intelligence (AI) have become increasing interested in person re-identification (PRe-ID).Surveillance cameras installed at public places is mainly used for monitoring the behavior and identification of suspects in those areas. Malicious events such as criminal activities have increased in recent years, there is need for video network systems to ensure the safety of people. In public transports like railway station, airports, inside planes and trains etc[1]. PRe-ID refers to the task of matching pedestrian observed from non-overlapping views of surveillance camera networks [2]. It is mostly done by extracting some kind of features captured in the image. However, the difficulty level of this task is increasedbecause the views captured by the surveillance cameras are recorded under unconstrained conditions, and thus, the obtained images may contain large variations due to changes in pose, viewpoint, and illumination, as well and occlusion etc [2]. In order to tackle the challenge of person re-identification (e.g. Occlusion, pose variation and illumination change etc.) many efforts have been made on discriminative features robust to these challenges and learning distance models for correct matching [3].

The main contributions of our work in order to enhance the efficiency of person re-identification system are:

1. A comparative study of two types of effective features, CNN and LOMO is carried out in order to improve the accuracy of our person Re-identification system.
2. A new features model based on fusion CNN and LOMO descriptors is done using multidimensional data representation based high order tensor.
3. To improve discriminative power for person Re-identification we applied a new metric learning method called Tensor cross-view quadratic analysis (TXQDA) Motivated by its success in kinship verification. In which, third order tensor data that combined several feature parts is used.
4. Finally, we evaluated and compared our results against the state of-the-art methods using the three challenging person Re-identification databases: VIPeR, GRID and PRID450s.

The structure of our work is organized as follows:

in chapter 1 biometrics and PRe-ID systems are described. Deep learning and artificial intelligence are conducted in chapter 2. Chapter 3 presents the experimental results on the three mentioned datasets.

Chapter I Generalities of Biometrics and person Re-identification

I.1. Introduction 4
I.2. Biometrics 4
I.3. Biometric systems' modalities 4
I.4. Steps of biometric system 6
I.5. Applications of biometric systems 6
I.5.1. Person Re-identification system 7
I.5.2. Applications of person Re-identification 8
I.5.3. Challenges of person Re-identification 9
I.6. Crosse validation protocol-based CMC 11
I.7. Conclusion 12

I.1. Introduction

Biometrics is the most suitable means of identifying and authenticating individuals in a reliable and fast way through unique biological characteristics. The objective of the development of this type of systems is to improve the security and protection of information.

With increasing concern and demands on security, biometrics has become more widespread and more efficient. It stands for "biological measure ".Biometric technology is now becoming the basis for a wide range of highly secure personal identifications and verifications.

In this chapter, we introduce a general description of biometric systems, presenting the different modalities of biometrics, its steps and applications, with focusing on our application which is PRe-ID, its applications and challenges.

I.2. Biometrics

Biometrics refers to the identification of individuals based on their unique characteristics. More specifically, biometrics is the science of identifying or verifying an individual's identity based on biological or behavioral characteristics. Physiological biometrics, such as fingerprints or hand geometry. behavioral biometrics such as signature or voice[4].

I.3. Biometric systems' modalities

Biometric systems exploit a variety of biometric characteristics (or modalities) including fingerprint, face, hand geometry, iris, DNA, vein, voice, signature and gait. The most leading biometric modalities are described and illustrated in Figure I.1 below:

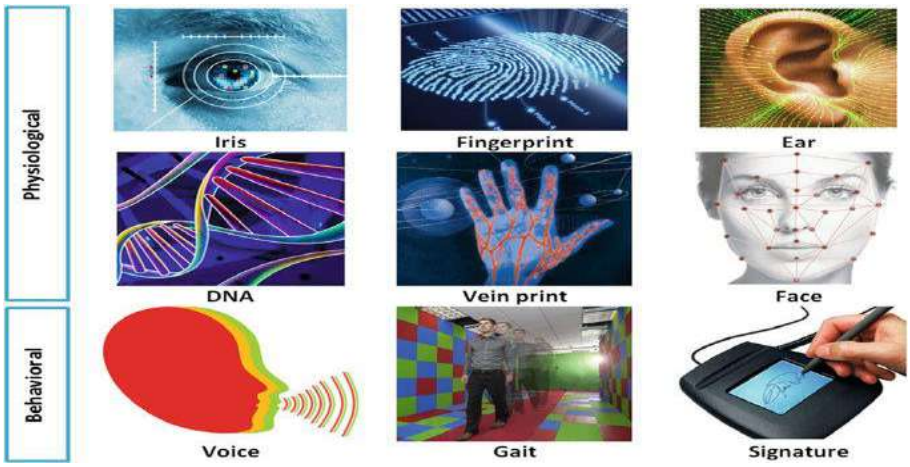


Figure I. 1: Various types of biometrics.

- **Face**: Humans use faces to distinguish one person from another. Face recognition is a computer application that automatically identifies or authenticates a person using digital images or video frames from a video source. One way is to compare the examples given with the examples in the database [5].
- **Fingerprint**: Is the most widely used biometric technology. Fingerprint (optical, silicon, ultrasound, touch less) uniqueness can be defined by analyzing the trivia of a human being. Trivia include sweat pores, distance stuck between ridges, bifurcation [6].
- **Hand Geometry**: Is the use of the geometric shape of the hand for recognition purposes. This method is based on the fact that one person's hand shape is different from another's hand shape and does not change after a certain age [5].
- **Iris**: Is getting more and more attention due to its high reliability for personal identification. The human iris, the annular part between the black pupil and the white sclera, has an unusually rich structure and contains many intertwined microscopic features such as freckles, crowns, streaks, furrows, and crypts. This visible characteristic, commonly referred to as the iris texture, is unique to each subject. Individual differences in the development of the anatomy of the body lead to this uniqueness, which leads to high reliability of individual identification [7].
- **DNA**: It's a genetic code, which mainly used in medico-legal applications and is not useful in real-time automatic recognition applications [8].
- **Signature**: A process used to recognize a person's handwriting or signature. Dynamic signature verification technology uses behavioral biometrics of handwritten signatures to verify the identity of a client computer. Analyzing speed, shape, stroke, pressure and time information while signing is a natural thing to do [6].
- **Voice**: Is defined as "the numerical representation of the sounds, patterns, and rhythms of the human voice". Voice biometrics is a type of signal-based biometric technology, where each person has a unique voice fingerprint, and voice biometrics can be used to identify and authenticate person. The human voice has many features that can be used to identify a person. Various pitches, tone of voice according to the shape of the mouth, movement of the lips, and structure of the vocal cords are unique characteristics that can identify a person. Collect voice biometric data and find it quickly [9].
- **Gait**: Is an emerging biometric technique, in which person are only identified by analyzing their gait. Although research is still ongoing, there has been interest in the gait as a method of identification, as it does not require the individual's cooperation. However, walking can

be affected by: a woman's pregnancy, after an accident, leg illness and significant weight gain etc. [8].

I.4. Steps of biometric system

The basic two modules of any biometric system are, enrollment (training) and matching (testing). In the following enrollment and testing are described [10]:

- **Enrollment:** The acquired raw data is first preprocessed to enhance its quality. After that, some relevant discriminatory features are extracted, in order to generate a compact representation called “template” that efficiently resumes the biometric characteristics. The generated template is then sent to the storage system. Generally, the enrollment step allows the biometrics recognition system to learn the identities of the authentic persons in working environment.
- **Matching module:** During the operating phase, the system is requested to identify a person. It proceeds to extract his discriminatory features using the extractor sub-module in the same manner that it has been done in the enrollment step. These extracted features are called query features. After that, the stored template is revoked to be compared with the query. The comparison aims to confirm that both the query and the template features originate from the same biometric subject (person). Generally, the comparison result is a degree of similarity ranging between 0 (total mismatch) and 1 (perfect match) that allows the system to take the suitable decision about the identity of the user or the target.

  On the other hand, the biometric system can operate either in verification or identification mode. In verification mode, the comparison is made only against one template in the system by conducting 1 to 1 comparison. This is possible when we want to confirm the identity claimed by a user. In the identification mode, the comparison is achieved against all records in the database by conducting 1 to many comparisons. This is the case when we want to know if the individual already exists in the database. So, the system tries to answer the question “who is the user?”.

I.5. Applications of biometric systems

The applications of biometric recognition systems are numerous. Basically, these are used for surveillance, law enforcement, time and attendance, logical access control, and physical access control. Other applications include [11]:

- Biometric systems are used to secure the systems for the surveillance and also for the physical buildings for various applications such as access control, computer networks, verification, investigations, authentication, parenthood determination, online banking, border control, e-commerce, medical records management and security monitoring.
- For improving confidence level after the verification of authenticated individual with all rights and privileges.
- For the safe, secure, efficient and convenient working of transactions using reliable methods.
- To identify an individual and control access to information, physical spaces and services.
- Improving safety, reducing fraud and national security.
- Provides uniqueness, strong immunity and stability to forgery.

I.5.1. Person Re-identification system

PRe-ID is defined as the process of determining whether a target person has already appeared on a camera network. Depending on the informational prompts currently available in the video analytics system, the task may be considered at various levels. For example, biometric data such as face, iris, or gait can be used to identify a person. However, in most video surveillance scenarios, these details are not available due to low video resolution or complex segmentation (congested places such as airports and subway stations). Therefore, reliable modeling of the appearance (clothes) of a person is required for re-identification. This is a particularly difficult problem, as changes in angle of view, lighting conditions, and various human positions can significantly change appearance [12].

PRe-ID is an important component of video surveillance systems. It refers to the process of identifying probes in galleries taken with different cameras, typically used in scenarios such as: If there is a probe image or video sequence with a specific person under a specific camera, this person will be requested from the other camera [13]. Figure I.2 illustrates an example for persons re-identification of system.

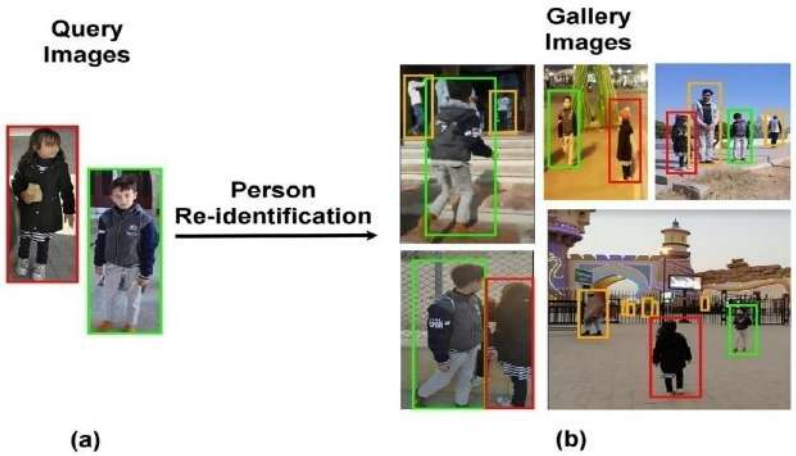


Figure I. 2: Person Re-identification system.

I.5.2. Applications of person Re-identification

PRe-ID strategies have a lot of potential in a lot of different fields, from defense and surveillance to retail and health care.

- **Cross-camera person tracking:** Computer vision requires the ability to monitor person through several cameras, conduct crowd movement analysis, and recognize movements in order to comprehend a scene [14]. Typically, cross-camera tracking consists of two steps, object tracking in single camera and data association between multiple cameras. Consequently, besides the difficulty in single camera tracking, the main challenge of cross-camera tracking is to associate the trajectories of human movement in different views [15].
- **Tracking by detection:** Individual re-identification may be useful even in a single-camera tracking setup. Multiple persons tracking is a difficult job, particularly in complex and crowded scenarios with frequent occlusions and individual interaction. The main idea is to detect the persons, estimate their motion patterns and associate detections in different frames. This linking step, called data association, is in fact a kind of re-identification [14]. This stems directly from the development of powerful discriminative methods in machine learning, and their application to detection with offline training. Many of these algorithms can be adapted for online testing, where each successful detection provides more information about the target [16].
- **Person retrieval:** Re-identification is linked to a recognition task in this case. A specific query with a target individual is given, and all of the corresponding instances in a broad

database are searched. For image retrieval, the re-identification task is used, and it typically returns ranked lists, similarly related objects, and so on [14].

- **Human-machine interaction:** In the recent decades the scientist have concentrated more in the developing technologies that help in the improving the interactions between the humans and the machines by integrating novel resourceful procedures with the computers. This caused a significant breakthrough in the computer-vision and enabled the computers to operate or control the machine by monitoring the gestures and the expression of the humans [17].
- **Long-term human behavior and activity analysis:** Analyzing consumer shopping habits, for example, by watching them touch, interview, and try items in stores under various surveillance cameras. Another example is geriatric health care research, which looks at the long-term actions of elderly person to help physicians make more precise diagnosis [14].

I.5.3. Challenges of person Re-identification

In the following, we present some challenging and difficult problems associated with re-identification of persons to solve. The most important are illustrated in Figure I.3.

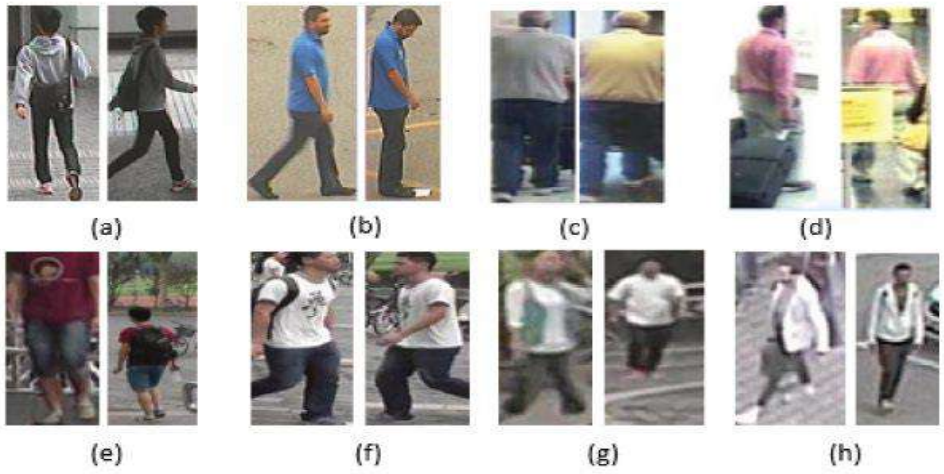


Figure I. 3: Example of some person Re-identification challenges.

- **Low resolution**: In most realistic conditions, the cost of the required number of cameras in any area can be very high, leaving a "clogged gap" with very sparse coverage. Therefore, the camera is usually mounted high on a wall, so pedestrians are usually far away from the camera. Even with a high-resolution camera for a specific person, the image is still relatively low resolution [14]. A camera view change occurs when a camera with a different field of view tries to re-identify a person.

- **Illumination change**: For PRe-ID, this issue is a critical one, especially in the case of cameras without overlapping field of view. The same person observed with two cameras under different illumination conditions can have two different appearances [18].Caused by the environment can also significantly change the camera's field of view.
- **Camera viewpoint variation**: Viewpoint is most important challenge in person re-id problem because Different heights of the camera, the distance between the person and the camera, and the direction in which person look, allow you to observe pedestrians of different shapes or sizes from different viewing angles. It supports the learning algorithm specifically in identifying the pedestrian. Finding the particular angles are important for learning models to identify the person [14, 19].
- **Large number of candidates in gallery set**: The camera network can cover large public places such as train stations or university campuses. Thus, the number of candidates for a given re-identification request can be enormous, and the number of candidates increases over Time. Computing to match a large set of galleries is expensive. To alleviate this problem, we can reduce the set of candidate matches using some temporal considerations and the spatial arrangement of different cameras [14].
- **Similar clothing**: In a large gallery set, persons are more likely to wear the same clothes. Most person wear dark clothes in winter in public places. Many persons wear almost identical jeans. This increases ambiguity and uncertainty in the matching process. In this case, it is more difficult to find distinctive features in the appearance of others [14].
- **Real-time constraint**: Some emergencies require immediate finding of a suspect. A low-latency real-time implementation is critical to processing multiple input video streams and returning query results quickly. The search space for a match can be very large and many potential candidates will be highlighted. So lookup time is important [14].
- **Pose variation**: The change of pose caused by the variation of the angles of view and by the person's walking direction leads to significantly different appearances of the same person seen by different cameras. A model trained in a standing position may not detect person running, crouching, or sitting. It is difficult to predict in the resulting image with relatively low resolution and quality [14, 20].
- **Partial occlusion:** Sometimes persons are partially or completely occluded by objects they are carrying, by overlaps with other people/objects, or by structures in the environment. If some important features of the visual signature are not visible, the re-identification fails [18].

I.6. Crosse validation protocol-based CMC

Cross validation may be any of various model validation techniques that are used to assess how well a predictive model will generalize to an independent set of data that the model has not seen before. Hence, it is typically employed in situations where we predict something and we want to gain a rough estimate of how well our predictive model will perform in practice [21]. Figure I.4 presents a simple explication.

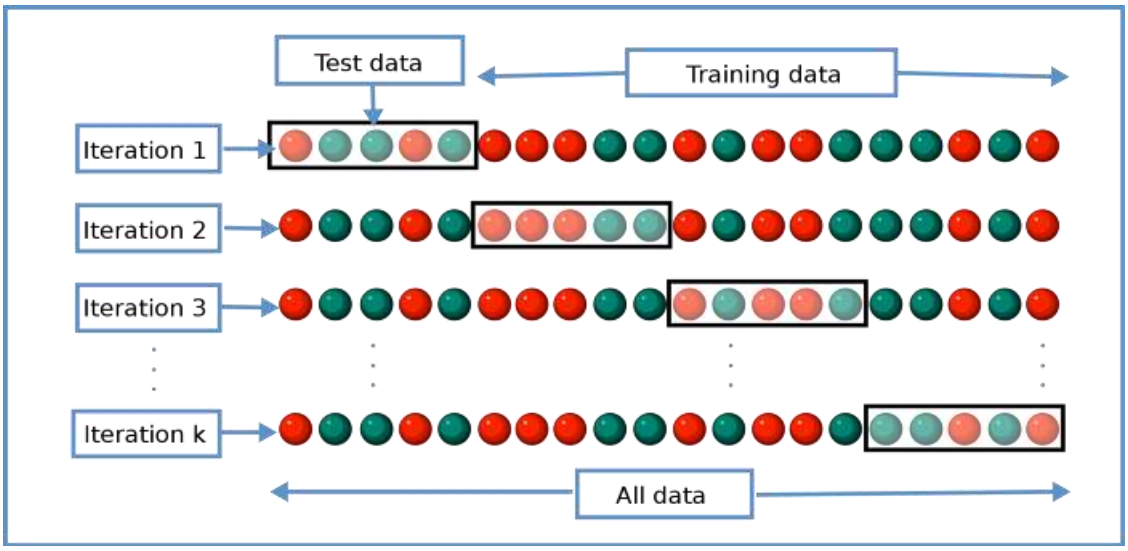


Figure I. 4: Example of cross validation model.

Cross-validation is similar to the repeated random subsampling method, but the sampling is done in such a way that no two test sets overlap. In k-fold cross-validation, the available learning set is partitioned intok disjoint subsets of approximately equal size. Here, “fold” refers to the number of resulting subsets. This partitioning is performed by randomly sampling cases from the learning set without replacement. The model is trained using k − 1 subsets, which, together, represent the training set. Then, the model is applied to the remaining subset, which is denoted as the validation set, and the performance is measured. This procedure is repeated until each of the k subsets has served as validation set. The average of the k performance measurements on the k validation sets is the cross-validated performance [22].

Figure I.5 illustrates this process for k = 10, i.e., 10-fold cross-validation. In the first fold, the first subset serves as validation set Dval,1 and the remaining nine subsets serve as training set Dtrain,1. In the second fold, the second subset is the validation set and the remaining subsets are the training set, and so on [22].

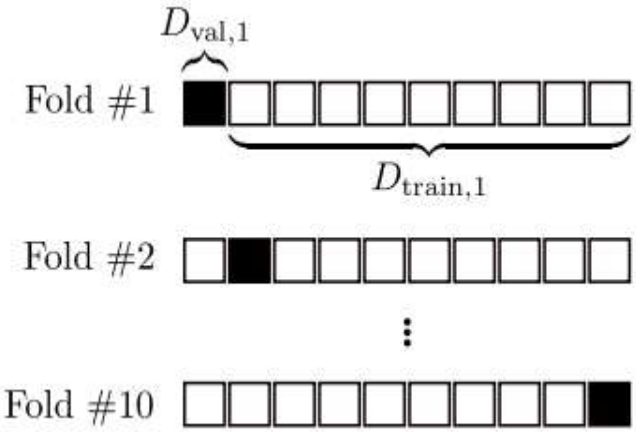


Figure I. 5: 10-fold cross-validation.

- **Cumulative Match Characteristic Curve (CMC):** CMC is the graphical representation of identification results, plotting Rank value on the x-axis, and the probability of correct identification on the y-axis.

I.7. Conclusion

In this chapter we have described in a general way the field of biometrics and different modalities used to recognize person. We presented the process of the PRe-ID system, the different applications and challenges encountered in the real world, such as the variation of illuminations, Camera viewpoint variation and the variation of poses. We have introduced the most popular and widely used databases in the literature for the evaluation of different algorithms of PRe-ID system.

Chapter II Person Re-identification: Algorithms and Methods

II.1. Introduction 14
II.2. Artificial intelligence 14
II.2.1. Application of AI 14
II.2.2. Challenges of AI 16
II.3. Machine learning 18
II.3.1. Classification of machine learning methods 18
II.3.1.1. Supervised learning 18
II.3.1.2. Unsupervised learning 20
II.3.1.3. Reinforcement learning 22
II.4. Transfer learning based deep learning 22
II.5. Feature extraction in Person Re-identification 23
II.5.1. Convolutional neural network CNN based Alexnet architecture 23
II.5.2. LOMO Descriptor 25
II.6. Conclusion 26

II.1.Introduction

Nowadays AI is already deeply embedded in our everyday life, it is the simulation of natural intelligence in machines that are programmed to learn and mimic the actions of humans. These machines are able to learn with experience and perform human-like tasks. It has various sub-domains such as ML, deep learning (DL) and neural networks, the objective of this chapter is to provide an overview of the basic concepts in ML literature, including their types and implementations, as well as their various algorithms. We also go over the CNN.

II.2.Artificial intelligence

AI is a field of computer knowledge by which we can produce intelligent machines which can comport like a human, figure like humans, and qualified to make decisions on their own. It is composed of two words artificial which define object that's made by humans and intelligence which refers to the capability to suppose on its own, and therefore this makes artificial intelligence " thinking power made by human" [23].

AI is a branch of science which aims to help machines find solutions to complex problems in a more human manner. It is usually a question of borrowing features of human intelligence and applying them like algorithms in a computer [24].

It is associated with informatics but it has many important links win other fields such as maths, psychology, cognition, biology and philosophy among many others.

II.2.1. Application of AI

There are various applications of AI, The most important are the following:

- **Computer vision:** Unlike AI techniques in which neural nets-which were considerably in use since 1940's, computer vision deals with concepts such as mimicking human visual system through various understanding of how camera schematics, projections and photogrammetry works. Within few decades, computer vision made its combined existence in all possible fields such as pattern recognition, machine learning, computer graphics, 3D reconstructions, virtual reality, and augmented reality. In 2010, computer vision was able to pull off high-end tasks such as object recognition, autonomous vehicle navigation, face detection, fingerprint recognition, fast computing image processing, and robotic navigation. Techniques such as line detection, feature extractions, segmentations, feature matching and tracking, optimization and reconstruction of 3D reality opened doors to

ample of astounding inventions such as visual simultaneous localization and mapping (SLAM), object tracking etc. The below shown Figure II.1 clearly illustrates the taxonomy of computer vision field and its inter-related branches such as science and technology, mathematics and geometry, physics and probability etc. Practically, AI techniques such as machine learning and deep learning needs huge chunks of training data which can be retrieved and processed using computer vision techniques [25].

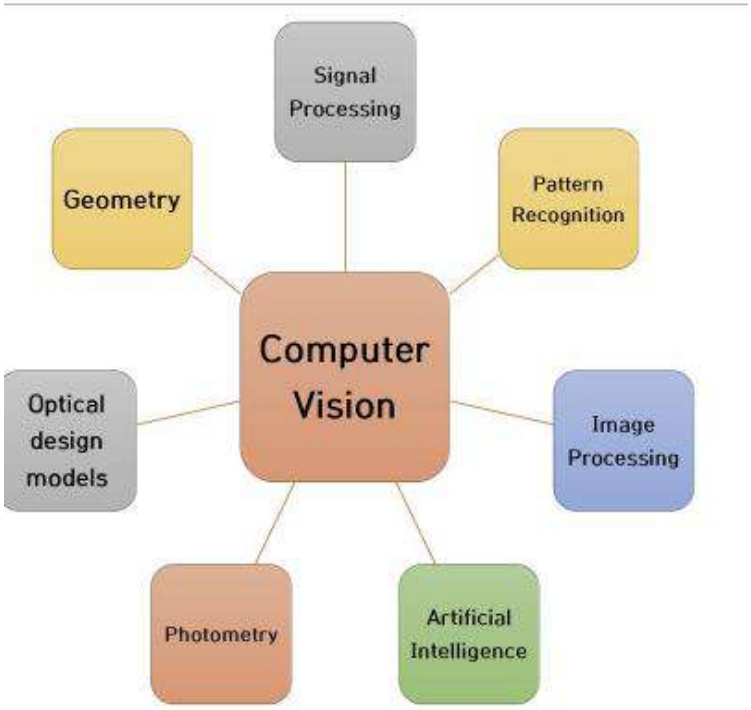


Figure II. 1 : Computer vision and its sub-concepts.

- **Internet of things (IoT)**: The research of AI applied to the IoT is no more novel. In the past, there were many proposed ideas about AI applied to the IoT. One of the ideas proposed is making all devices communicate to each other. This means that from the transport, a user can control home appliances. From the smart phone, the user not only making a call but also can control the home appliances [26].
- **Natural language processing**: Natural Language processing (NLP) is a branch of computer science, AI and linguistics concerned with the interactions between computers and human (natural) language. Natural language is the language spoken by humans. And any language that person learns from the environment and use to communicate with each other. The field of natural language processing is deep and diverse NLP is a collection of techniques used to extract grammatical structure and meaning from input in order to

perform a useful task as a result, natural language generation builds output based on the rules of the target language and the task at hand. NLP is useful in the fields of tutoring systems, duplicate detection, computer-aided instruction, and database interfaces as it provides a way to increase interactivity and productivity [27].

- **Robotics**: AI has a remarkable role in Robotics. Usually, general robots are programmed such that they can perform some repetitive task, but with the help of AI, we can create intelligent robots which can perform tasks with their own experiences without pre-programmed [28]. AI in robotics helps robots performs the crucial tasks with a human-like vision to detect or recognize the various objects. Nowadays, robots are developed through machine learning training. A huge amount of datasets is used to train the computer vision model, so that robotics can recognize the various objects and carry out the actions accordingly with right results [29].
- **Crowd sourcing and human computation**: It is focused on the creation of innovative ways to exploit human intelligence [30].
- **Neuromorphic computing**: Traditional computers use von Neumann's architecture model. With the success of the deep neural networks, alternative models are being developed, many of which are inspired by neural biological networks [30].
- **Collaborative systems**: Research on collaborative systems investigates models and algorithms to support the development of autonomous systems that can collaborate with each other and with human beings [30].

II.2.2. Challenges of AI

AI is poised to be one of the biggest things to hit the technology industry (and many other industries) in the coming years. But just because AI holds enormous potential does not mean that it does not also have its challenges.

- **Building trust**: AI is hard to trust because it's based on science, technology, and algorithms that most person aren't aware of [28].
- **AI human interface**: Because this is a new technology, there is a significant shortage of skilled person in data analysis and data science. You can then assign them to take full advantage of artificial intelligence. As AI advances, businesses are running out of skilled professionals who can meet their needs and work with this technology. Employers must train their employees to use this skill [30].

- **Investment**: AI is an expensive technology that not every business owner or manager can invest money into as large amount of computing power will be necessary and sometimes hardware acceleration with GPU, FPGA, or ASIC must be in place to run machine learning models effectively. Though adoptability of AI is surging high, it has not been integrated fully in business's value chain at the scale which it should have. Moreover, enterprises of those who have incorporated are still in nascent stage which have resulted in the slowdown in the lifting of the AI technology at scale and thus been deprived of cost benefit of scale. After decades of speculation and justifiable anxiety about the social implications of intensifying & potentially de-stabilizing AI technology for humankind and Black box problem, AI investors are bit skeptical from parking their money in potential startups [30].
- **Software malfunction**: With machines and algorithms controlling AI, decision-making ability is automatically ceded to code-driven Black Box tools. Automation makes it difficult to identify the cause of mistakes and malfunctions. Moreover, due to the lack of ability of human beings to learn and understand how these tools work, they have little or no control over the system which is further complicated as automated systems become more prevalent and complex [30].
- **Non-invincible**: Like any other technology, AI also has its own limitations; it simply cannot replace all tasks. However, it will result in emerging new job domain with different quality job profile [30].
- **High expectations:** Research in artificial intelligence is conducted by large pool of technologist and scientists with varying objectives, motivation perspectives, and interests. Main focus of research is confined in understanding the underlying basis of cognition and intelligence with heavy emphasis on unraveling the mysteries of human intelligence and thought process. Not everyone understands the functioning of AI and might also have very high expectation of functioning [30].
- **Data security**: ML and decision-making capability of AI and its application are based on huge volumes of classified data, often sensitive and personal in nature. This makes it vulnerable to serious issues like data breach and identity theft. Mostly, companies and government striving for profits and power, respectively, exploit the AI-based tools which are generally globally networked which make them difficult to regulate or rein in [30].
- **Algorithm bias**: The good or bad nature of an AI system really depends on the amount of data they are trained on. Hence, the ability to gain good data is the solution to good AI

systems in the future. But, in reality, the everyday data the organizations collect is poor and holds no significance of its own [31].

- **Data scarcity**: The main factor on which all the deep and machine learning models are based on is the availability of data and resources to train them. but as this data is generated from millions of users around the globe, there are chances this data can be used for bad purposes [31]. Power and capabilities of AI and its applications depend directly on the accuracy and relevancy of supervised and labeled datasets being used for training and learning. There is scarcity of quality-labeled data. Though efforts are underway by means of transfer learning, active learning, deep learning, and unsupervised learning, to devise methodologies to make AI models learn despite the scarcity of quality-labeled data, it will only aggravate the problem [30].

II.3.Machine learning

MLis an evolving branch of computational algorithms that are designed to emulate human intelligence by learning from the surrounding environment. They are considered the working horse in the new era of the so-called big data. Techniques based on machine learning have been applied successfully in diverse fields ranging from pattern recognition, computer vision, spacecraft engineering, finance, entertainment, and computational biology to biomedical and medical applications [32].

ML is a subset of AI that studies algorithms and automation techniques to solve complex problems [33]. With the help of computer science, we develop systems that allow us to learn the data in a way that we train. systems can learn and improve with experience and time, develop a template that can be used to predetermine the results of questions based on previous learning [34].

Many basic problems in ML (such as supervised and unsupervised learning) are well known. A central focus in current studies concerns the chance of increasing the ability of algorithms to work on extremely large datasets[30].

II.3.1. Classification of machine learning methods

Based on the methods and way of learning, ML is divided into mainly three types, which are: supervised learning, unsupervised learning and reinforcement learning.

II.3.1.1.Supervised learning

Supervised learning is a type of machine learning that uses labeled information to form ML models. The output of the labeled data is already known. The template should simply match the

inputs with the respective outputs [35]. Supervised machine learning algorithms are these algorithms that require external support. The input dataset is divided into train and test dataset. The input dataset is split into train and test dataset. The train dataset has output variable which needs to be predicted or classified. All algorithms learn some kind of designs from the training dataset and utilize them to the test dataset for prediction or classification [36]. Figure II.2 present supervised learning workflow.

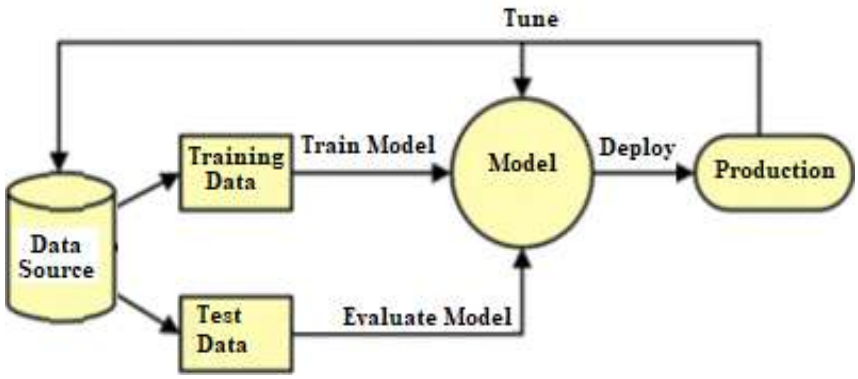


Figure II. 2 : Supervised learning workflow.

Most famous supervised machine learning algorithms are:

- **Decision Trees**: are a supervised learning technique used for classification. This type of algorithm builds trees to predict the outcome of an input vector based on decision rules derived from features present in the data. Decision trees are useful because they are easy to visualize so you can understand the factors that led to the outcome [37]. Figure II.3 shows a simple decision tree.

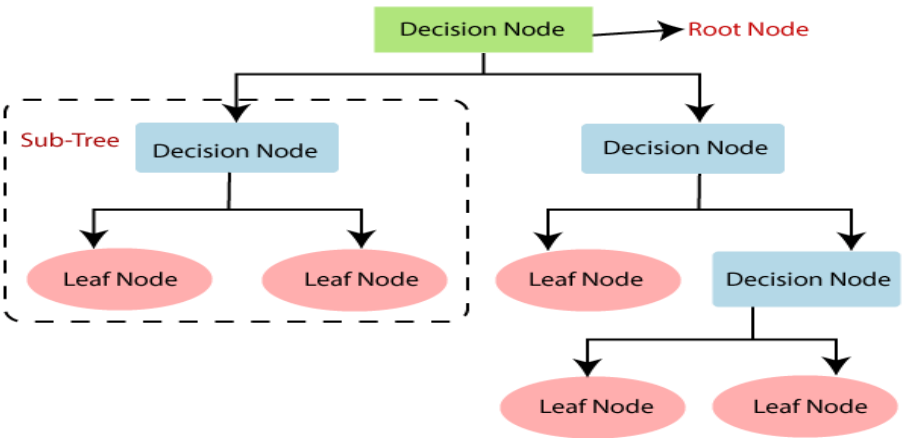


Figure II. 3 : Decision tree.

- **Support Vector Machine**: A support vector machine (SVM) is a supervised learning model with associated learning algorithms that analyze data used for classification and regression analysis [36].For instance, an SVM can learn to recognize fraudulent credit card activity by examining hundreds or thousands of fraudulent and nonfraudulent credit card activity reports [38]. SVM can learn to recognize handwritten digits by examining a large collection of scanned images of handwritten zeroes, ones and so forth.
- **NavieBayes**: Isa famous classifier which consists on conditional probabilities. Indeed, it employs Bayes theorem and it suppose that features have a solid independence between each other. Furthermore, it has many advantages such as simplicity of use, quick convergence, and high scalability. Finally, Naïve Bayes needs less training data for building a model [35].

II.3.1.2.Unsupervised learning

Unsupervised learning is a type of machine learning that uses unlabeled information to form ML models [35]. The unsupervised learning algorithms learn few features from the data. When new data is introduced, it uses the previously learned features to recognize the class of the data. It is mainly used for clustering and feature reduction [36].

The unsupervised learning attribute to the following algorithm types:

- **K-Means clustering:** The main goal of the K-Means clustering algorithm is to minimize the sum of distances between the points and their respective cluster centroid. Also, to assign each data point to its closet k-centre. Points that are near to a particular K centre, create a cluster [39]. And also the algorithm is to divide the examples in the dataset into k clusters. Each example is a numeric vector that allows the distance between vectors to be calculated as Euclidean distance [37]. The algorithm is presented in Figure II.3 bellow.

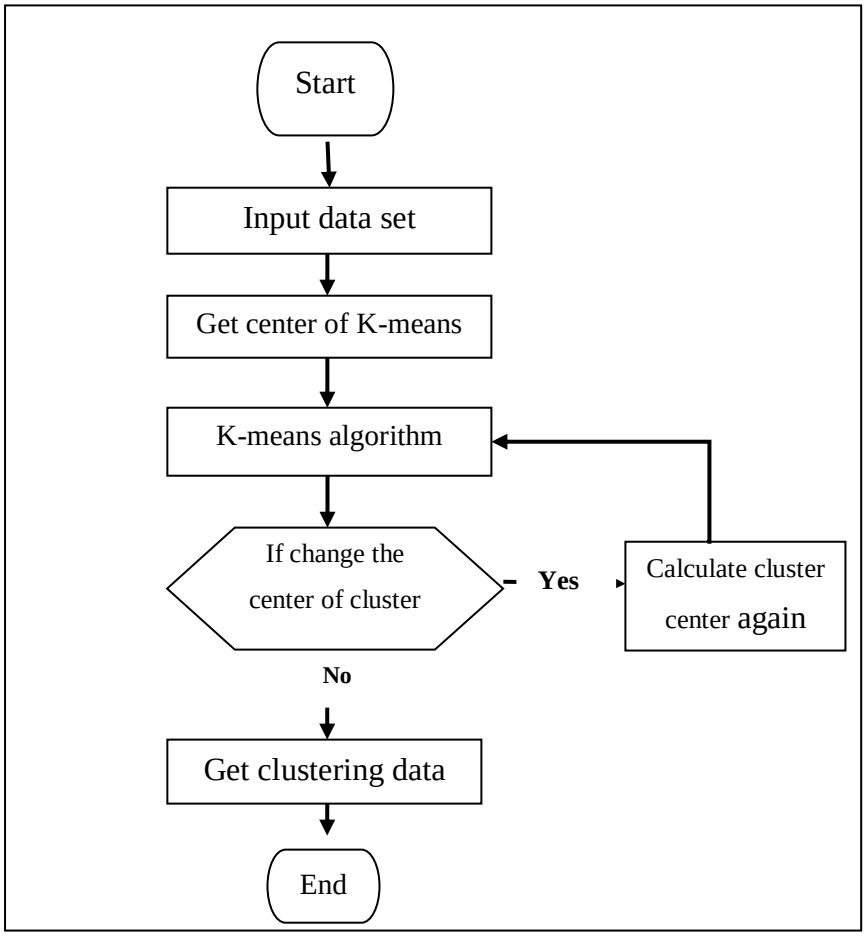


Figure II. 4 : K-mean clustering algorithm.

- **Principal Component Analysis (PCA)**

Component Analysis is an unsupervised learning algorithm that is used for the dimensionality reduction in machine learning. It is a statistical process that converts the observations of correlated features into a set of linearly uncorrelated features with the help of orthogonal transformation. These new transformed features are called the Principal Components. It is one of the popular tools that are used for exploratory data analysis and predictive modeling. It is a technique to draw strong patterns from the given dataset by reducing the variances. PCA generally tries to find the lower-dimensional surface to project the high-dimensional data.

PCA works by considering the variance of each attribute because the high attribute shows the good split between the classes, and hence it reduces the dimensionality. Some real-world applications of it are image processing, movie recommendation system, optimizing the power allocation in various communication channels. It is a feature extraction technique, so it contains the important variables and drops the least important variable [40].

II.3.1.3.Reinforcement learning

Reinforcing learning is a model of behavior learning. The algorithm receives comments from the data analysis so that the user is guided towards the best outcome. Reinforcement learning is different than other types of supervised learning because the system is not trained with the sample dataset. It's more of a trial and error learning system. Therefore, a sequence of successful decisions will lead to the process being "strengthened" because it best solves the issue at hand [41].

Through rule extraction, pattern matching, and mining, machine learning become one of the important tools which is further complimented by motivational decision-making capability implemented via reinforcement learning. Advent of reinforcement learning sharpens the ability of AI to address the real-world dynamic problem of complex nature [30].

II.4.Transfer learning based deep learning

DL is a subfield of artificial intelligence that focuses on building large neural network models capable of making accurate data-driven decisions, simple neural network example is shown in Figure II.5. DL is especially useful in environments where the data is complex and large datasets are available [42].

Dl is a form of machine learning that allows computers to learn from experience and understand the world based on a conceptual level. Because computers accumulate knowledge from experience, human computer operators do not need to formally specify all the knowledge a computer needs. Hierarchies of concepts allow computers to learn complex concepts by building them from simpler concepts [43].

Transfer learning solves the basic problem of insufficient training data. It tries to transfer the knowledge from the source domain to the target domain by relaxing the assumption that the training data and the test data must be. This will leads to a great positive effect on many domains that are difficult to improve because of insufficient training data [44].

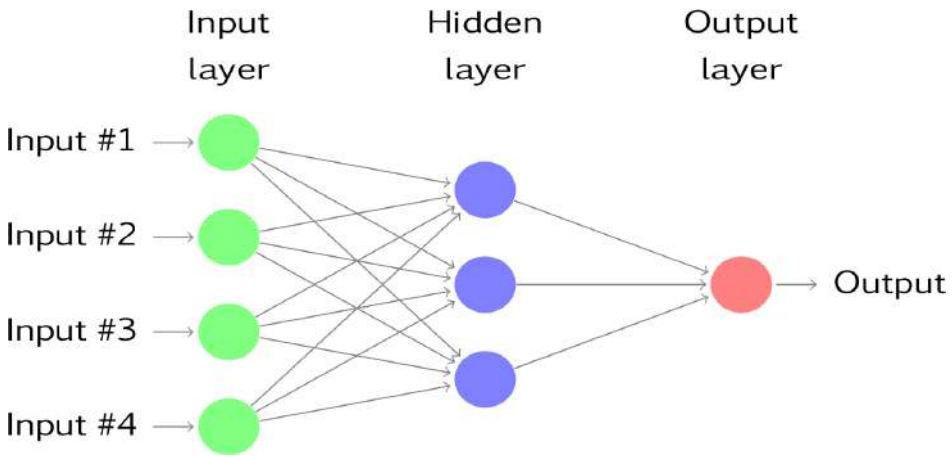


Figure II. 5 : Simple neural network exemple.

II.5.Feature extraction in Person Re-identification

Feature extraction is crucial components in PRe-ID. The large pose deformations and the complex view variations exhibited by the captured person images significantly increase the difficulty of learning and matching of the features from person images [13].

II.5.1. Convolutional neural network CNN based Alexnet architecture

Convolutional neural networks (CNNs) are a class of deep neural networks designed mainly for image analysis. CNN is a mathematical construct that is typically composed of three types of layers (or building blocks): convolution, pooling, and fully connected layers. The first two, convolution and pooling layers, perform feature extraction, whereas the third, a fully connected layer, maps the extracted features into final output.

- **Convolution layer**: Is the most important component of any CNN architecture. It contains a set of convolutional kernels (also called filters), which gets convolved with the input image (N-dimensional metrics) to generate an output feature map[45]. the convolution operator is defined as:

$$(W * X)(i,j) = \sum_m \sum_n X(m,n)W(i-m, j-n) \quad \textbf{(II.1)}$$

Figure II.6 shows the convolution operation.

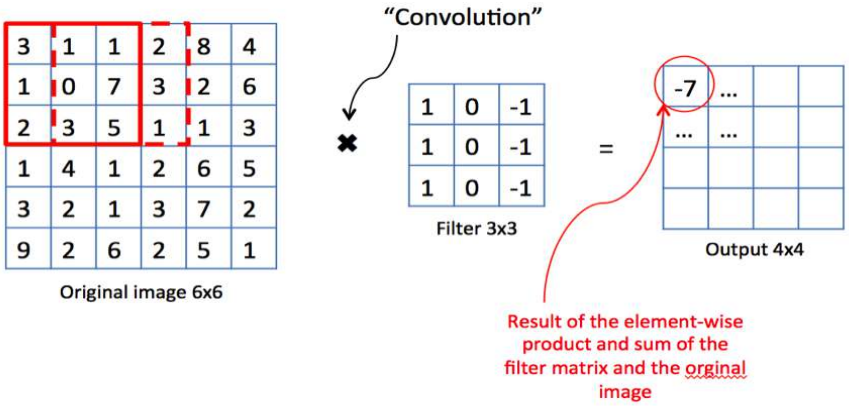


Figure II. 6 : The convolution operation.

- Stride and Padding: Used to determine the behavior in convolving the filters over the image and specifying the output dimension of the given layer [46].

  The stride determines the size of sliding over, while the padding refers to expanding the dimension with zero pixels to avoid losing the corner features during the convolution process [46].

- Activation function: The convolved results are sent through a nonlinear activation function. Nonlinear activation functions allow the extraction of nonlinear features. Traditionally, the sigmoid and hyperbolic tangent functions were used. Recently, rectified linear units (ReLUs) have become popular [46].

  The ReLU is a piecewise linear function, has the simplified form F (x) = max(x, 0). The ReLU retains only the positive part of the activation, by reducing the negative part to zero, while the integrated maximum operator promotes faster computation [46].

- **Pooling layer**: This is used for dimension reduction after one or more convolutional layers. In theory, there are different pooling techniques, including max, average and sum pooling. Max-pooling is the most popular method adopted as it significantly removes the outliers upon selecting the features. It basically refers to selecting the largest value in each receptive field of the feature map [46]. Figure II.7 shows an example of 2x2 max pooling.

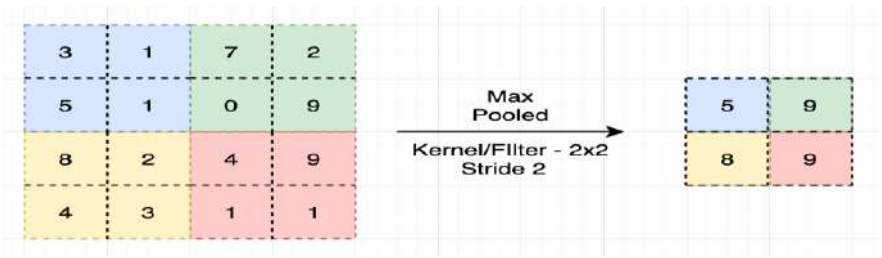


Figure II. 7 : Example of 2x2 max pooling.

- **Fully connected layer**: This is the layer preceding the final classification output of the network. It has two parameters: the number of nodes and the selected activation function. It is a vector in which all the pixels from the prior layer are flattened into one [46].

- **AlexNet architecture**

Krizhevsky et al [47],designed first large-scale CNN model, called AlexNet in 2012, which is designed to classify ImageNet data. It consists of eight weighted (learnable) layers among which the first five layers are convolutional layers, and afterward, the last three layers are fully connected layers. Since it was designed for ImageNet data, so the last output layer classifies the input images into one of the thousand classes of the ImageNet dataset with the help of 1,000 units. The architecture of AlexNet is shown in Figure II.8.

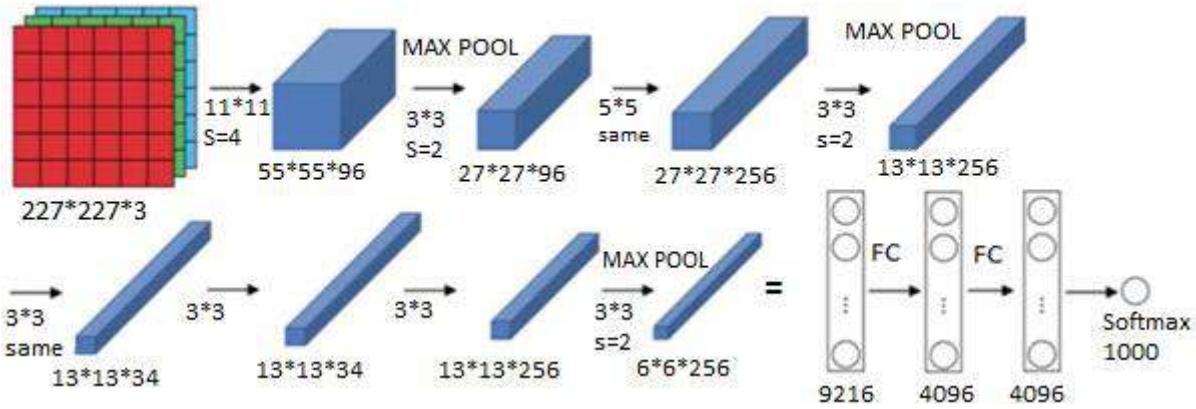


Figure II. 8 : The architecture of AlexNet.

- The AlexNet used ReLU non-linearity activation function after each convolutional and fully connected layer [45].
- It used max-pooling layer after each Local Response Normalization (LRN) layer and the last convolutional layer [45].
- Since it has a larger number of weights (learnable), so to avoid over-fitting it uses several regularization tricks like dropout and data augmentation [45].

II.5.2. LOMO Descriptor

LOMO feature descriptor was developed in order to provide a scale and pose invariant representation, as different poses had typically caused problems for PRe-ID tasks. This is achieved by utilizing an overlapping patch based approach, where for each patch a joint HSV histogram and a Scale Invariant Local Ternary Pattern (SILTP) histogram is calculated. In order to make the feature descriptor scale invariant, a three-scaled pyramid representation is constructed by applying a 2×2 mean filter per step. In order to make the feature descriptor pose invariant, each row of patches

is analyzed and the “local maximal occurrence” is determined by only selecting the largest bin value across all patch histograms in the row.

All row histograms are concatenated, and subsequently concatenated across the three scale steps. Finally, large values in the joint-HSV and SILTP histogram vectors are suppressed using the log operator, the vectors are ℓ2 normalized and concatenated into a single feature vector. In the original implementation the multi-scale Retinex transformation was applied as a pre-processing step in order to match lighting and colors from different cameras. As we utilize a single camera setup this transformation is not applied [48]. Figure II.9 shows LOMO descriptor.

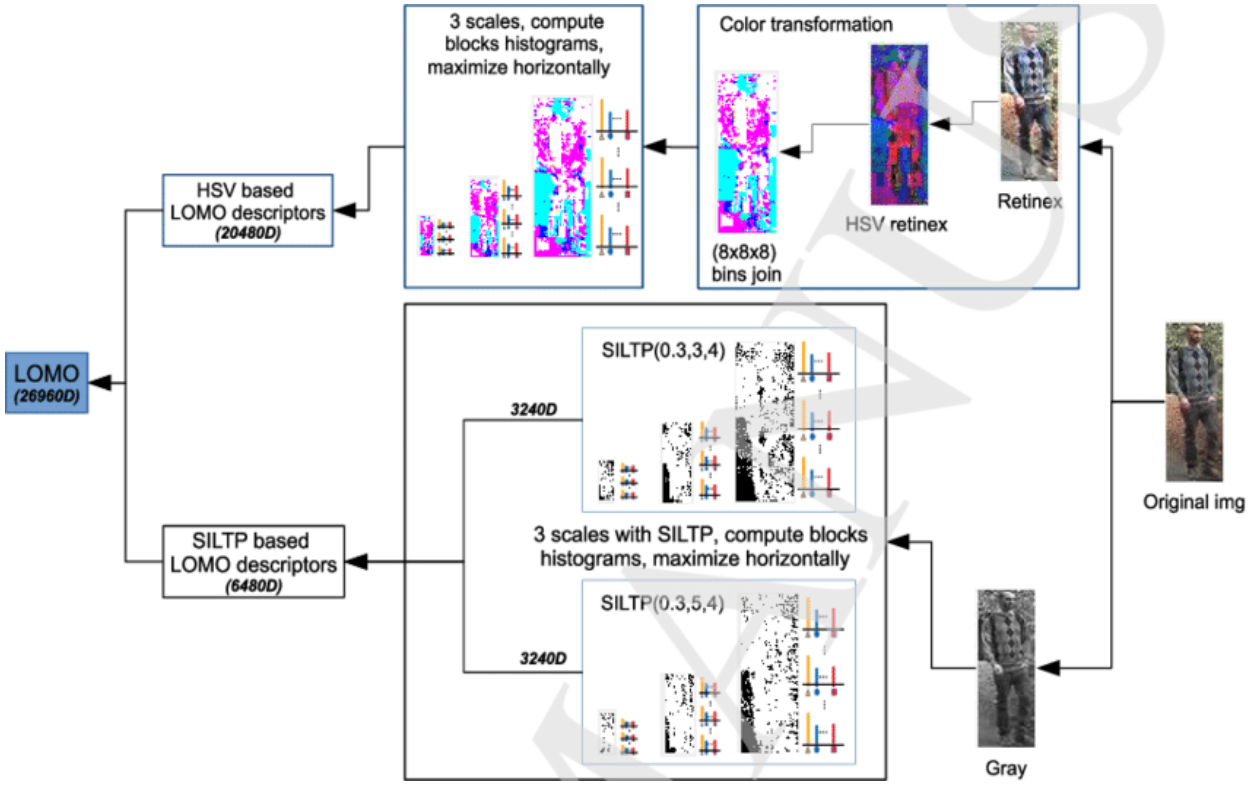


Figure II. 9 : Feature extraction for LOMO descriptor at image-level.

II.6.Conclusion

In this chapter, we reviewed the application and challenges of AI. We also discussed as well one of the most popular deep neural networks which is the CNN. That has an excellent performance in machine learning problems, specially the applications that deal with image data. And we reviewed the LOMO feature descriptor.

Chapter III Implementation and results

III.1. Introduction 28
III.2. Person Re-identification datasets 28
III.3. Overview of the proposed approach 30
III.4. TXQDA multilinear subspace projection 32
III.5. Results analysis 33
III.5.1. Protocol and parameter settings 33
III.5.2. Discussions 36
III.5.3. Comparison with the stat of the art 41
III.6. Conclusion 43

III.1. Introduction

PRe-ID focuses on matching the same person across non-overlapping camera views. It has received a lot of attention from the research community in recent times. Due to its vital role in security-based applications. It's a challenging task in real life scenario. Two fundamental problems in any PRe-ID are feature representation and metric learning. In our work, we select two effective feature types called LOMO and CNN. We propose a new model based on combine CNN and LOMO features in which high order tensor data representation is used. For subspace and metric learning we used a new subspace projection called TXQDA, which, has proven to be very effective with the kinship verification systems [49, 50].

In this chapter we are going to show the effectiveness and accuracy of these three methods on our three challenging datasets VIPeR [51], GRID[1] and PRID450s [40].

III.2. Person Re-identification datasets

There are numerous datasets used to measure the effectiveness of re-identification methods. Some public data sets that we test against our re-identification model and their characteristics are resumed in Table III.1 bellow:

Table III. 1: Datasets and their characteristics.

Dataset	Time	#ID	#image	#camera	#image size (pixels)
VIPeR	2007	632	1264	2	128×48
GRID	2009	250	1275	8	Vary
PRID450S	2014	450	900	2	128 x 48

- **VIPeR**: This dataset captured from two camera views. It contains 1264 images of 632 persons. Each image pair was captured from a random viewpoint under different illumination variations. All images are scaled to 128x48 pixels. 316 identities are randomly selected to build the training set and the rest 316 identities are used for testing. Figure III.1 shows a part of the dataset [51].



Figure III.1: Part of VIPeR dataset.

- **GRID**: It's captured using eight different camera views installed in a busy underground station. It contains 250 pedestrian image pairs. There are 775 additional images that do not belong to the 250 persons which can be used to enlarge the gallery set. The dataset is challenging due to variations of pose, colors, lighting changes as well as poor image quality caused by low spatial resolution [1]. 125 image pairs are used for training, and the remaining 125 image pairs are used for testing. The size of images varies from 29×67 to 181×384. Figure III.2 shows a part of the dataset.



Figure III.2: Random images from GRID.

- **PRID450S**: As shown in Figure III.3. The dataset contains significant poses variations, illumination and viewpoint variations and the occlusions. It contains image pairs of 450 walking pedestrians recorded from two disjoint surveillance cameras. All the images are normalized into 128 x 48 pixels [40].The main dataset challenges are viewpoint, poses and significant differences in background and illuminations. 225 identities are randomly selected to build the training set and the rest 225 identities are used for testing [40].



Figure III. 3: Sample image pairs from the PRID450S dataset.

III.3. Overview of the proposed approach

Like any person Re-identification system [52-54], the general mechanism of our approach relies on a given probe image of a candidate person. Then, it is required to identify images of the same person from different camera in the gallery image datasets. As a result, the system outputs the ranked images of persons similar to the probe image. To achieve this objective, our approach consists of three essential steps: feature extraction and tensor data design, multilinear subspace projection, and matching and score normalization that are applied in training and testing phases:

Phase one (training): the offline phase, all images in gallery and probe sets subjected to the process of features extraction by LOMO, CNN descriptors and our contribution that combined CNN and LOMO features to form our third order tensors with 3 modes: mode 1 presents the extracted features, mode 2 for feature parts and mode3 is set depending on the number of persons in the dataset. Then, we obtained tensors data that projected in new discriminate subspace based TXQDA giving the learning projection matrices.

Phase two (test): the online phase, the pair images of the target person pass through the same steps of the training phase. The matching of the probe with the gallery is through measuring the mahalanobis distance.The proposed multilinear pipeline of our Pre-ID system is illustrated in Figure III. 4.

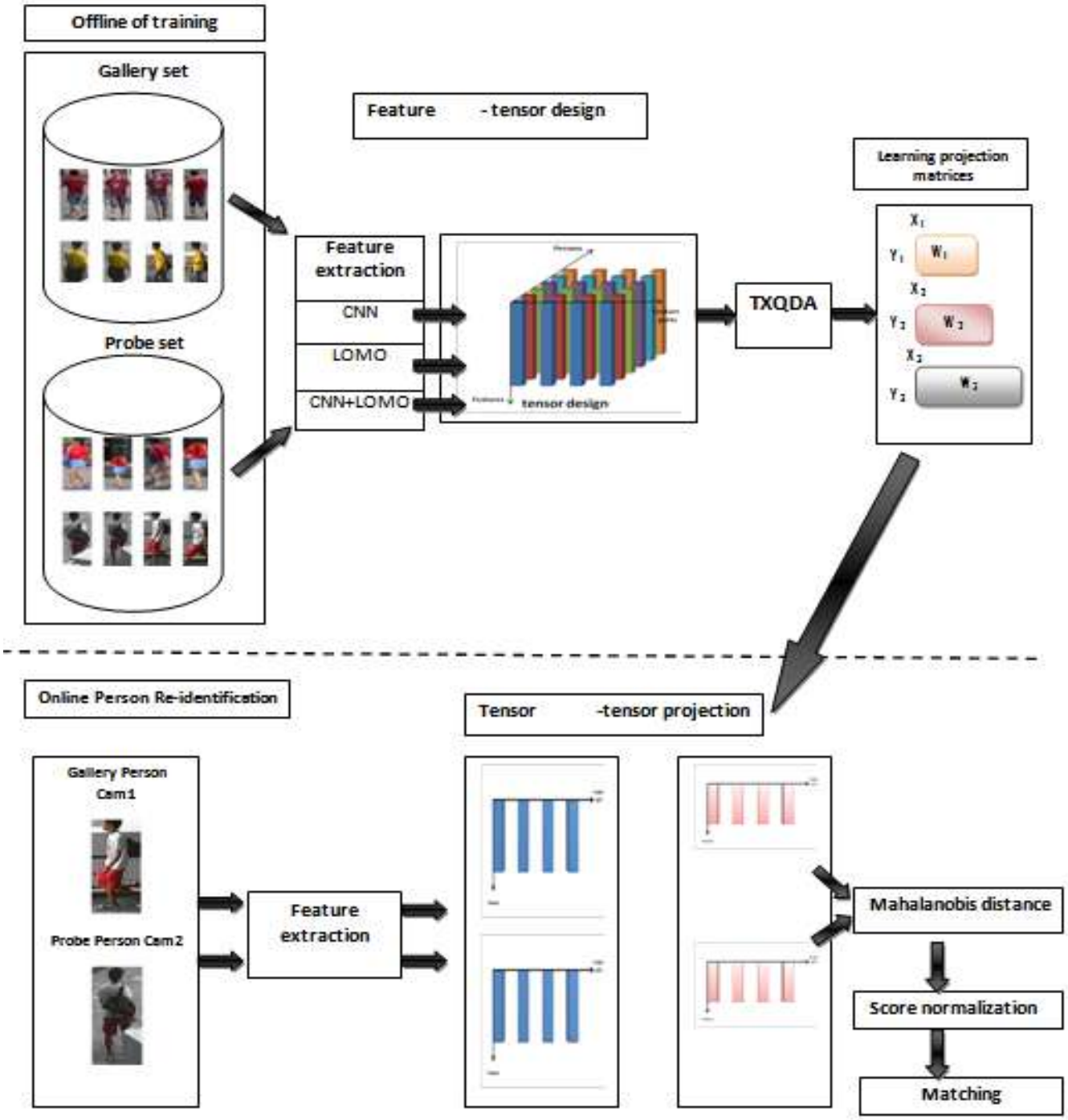


Figure III. 4: Person Re-identification system.

III.4. TXQDA multilinear subspace projection

TXQDA aims to project the multidimensional data into new discriminative subspace, in which the margin of each positive pair (images of the same person captured from different cameras) is reduced as well as each negative pair (images of different persons) is distended. Let X, Y $\in\mathbb{R}^{n_1\times n_2\times\ldots\times n_m}$ be the two cross-view training tensors of l classes (persons). X contains samples of the first view, captured from the first camera, and Y contains the samples of the second view, captured from another camera. The goal of TXQDA is to yield m projection matrices. In our case, we have three matrices, a projection matrix for each tensor mode, P1$\in\mathbb{R}^{n_1\times n_1'}$, P2 $\in\mathbb{R}^{n_2\times n_2'}$and P3$\in\mathbb{R}^{n_3\times n_3'}$.

TXQDA uses an iterative optimization process, which halts when a maximum iterations number is reached or the change in the projection matrix between two iterations is insignificant. The number of iterations is varied to obtain the best results. The training steps of the proposed TXQDA approach are given in the following algorithm.

Algorithm

Input: $\boldsymbol{X}$ and $\boldsymbol{Y}$, the cross-view (Camera1-Camera2) training tensors.
max_itr, the maximum number of iterations.
The projection dimensions n'$_1$, n'$_2$,…, n'$_m$.

Initialization: $P_1^1 = I_{n_1}, P_2^1 = I_{n_2}, \ldots, P_m^1 = I_{n_m}$ where I is the identity matrix.

For iteration= 2 to max_itr do

For k= 1to m do

$$\boldsymbol{X}' = \boldsymbol{X} \times_1 P_1^{itr-1} \cdots \times_{k-1} P_{k-1}^{itr-1} \times_{k+1} P_{k+1}^{itr-1} \cdots \times_m P_m^{itr-1}$$

Deploy the tensor$\boldsymbol{X}'$ to obtain its unfolding matricesX^k (k=1,2,… m)

$$\boldsymbol{Y}' = \boldsymbol{Y} \times_1 P_1^{itr-1} \cdots \times_{k-1} P_{k-1}^{itr-1} \times_{k+1} P_{k+1}^{itr-1} \cdots \times_m P_m^{itr-1}$$

Deploy the tensor $\boldsymbol{Y}'$ to obtain its unfolding matricesY^k (k = 1,2, … m)

Step1: Calculate the covariance matrices$Q_w$and Q_b

Step 2: Calculate the projection matrices P_k in k-mode by eigenvectors decomposition

Step 3: Calculate the new projection matrices $P_k^{TXQDA^{itr}}$

If iteration> 2 and $\left\| P_k^{TXQDA^{itr}} - P_k^{TXQDA^{itr-1}} \right\| < n_k \times n_k \times \varepsilon,\ k = 1, \ldots, m,$

End for

End for

Output

The k-mode projection matrices $P_k^{TXQDA^{max_itr}}$, $k = 1, 2, \ldots, m$.

In the test phase, the cross-view images of the probe are projected into the new multilinear subspace using $P_k^{TXQDA^{max_itr}}$. In which, the eigenvectors of $P_k^{TXQDA^{max_itr}}$ are sorted according to Λ_k in decreasing order.

III.5. Results analysis

To evaluate the effectiveness of the proposed approach, various experimental studies are performed, including comparisons with state-of-the-art methods, on three challenging re-identification datasets: VIPeR [51], GRID [1] and PRID450s [40].

III.5.1. Protocol and parameter settings

Person Re-identification means to find a person of interest in a query image from a large database. There are two main critical problems feature representation and metric learning. In our work, we propose three scenarios:

- **1st scenario**: CNN based AlexNet and TXQDA.

  On a target Pre-ID dataset, we used CNN's features from the vector learned. From CNN's fully connected layer (FC7), we extracted 4096 dimensional features vector for each image, and we subdivided it into four vectors of 1024 feature to get our tensor data with three modes as shown in Figure III.5.

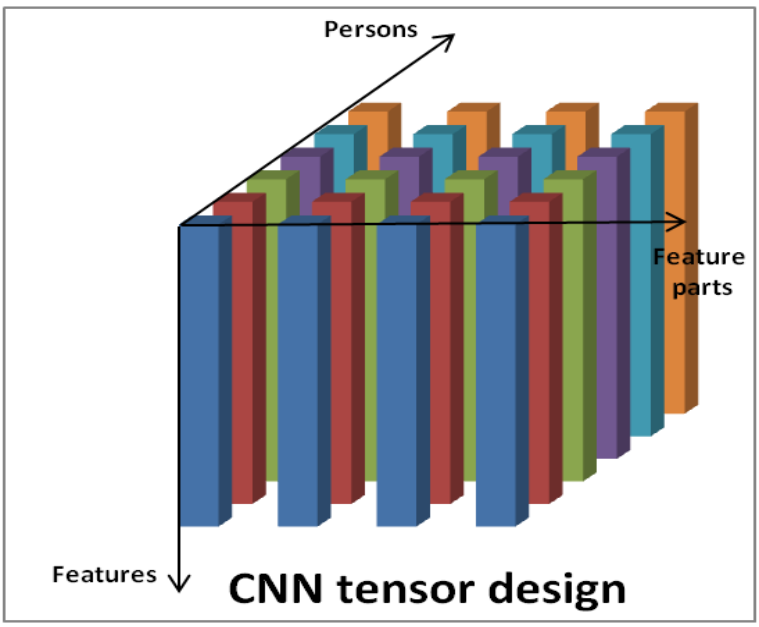


Figure III. 5: CNN 3rd order tensor design.

The size of mode 1 is 1024 feature, mode 2 is 4 and mode 3 is related with the number of images in each dataset in which: GRID 1276, VIPeR632 and PRID450s 900.

- **2nd scenario**: LOMO and TXQDA.

The feature vector with LOMO descriptor is 29690 dimensional features subdivided into 8 vectors of size 3370 to form our tensor data with three modes. FigureIII.6 illustrates LOMO tensor data.

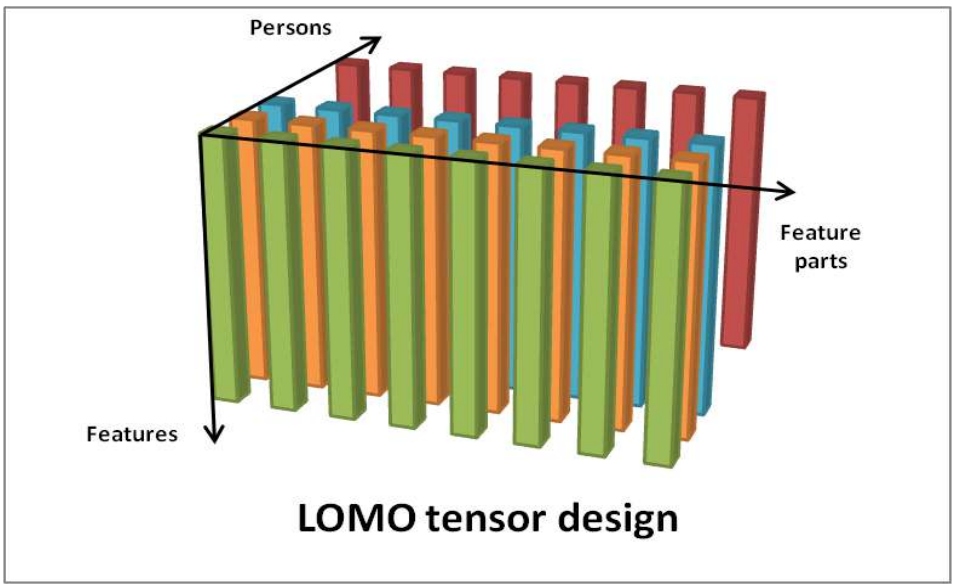


Figure III. 6: LOMO 3rd order tensor design.

The size of mode 1 is 3370 feature, mode 2 is 8 vectors and mode 3 is related with the number of images in each dataset.

- **3rd scenario:** CNN+LOMO

we subdivided the LOMO tensor into two tensors with three modes: mode 1 (dimension of 3370), mode 2 (dimension of 4) and mode 3 related to the number of images of each dataset, then we concatenate the CNN tensor with the two LOMO tensors to get our data tensor based fusion features with three modes: mode 1 (3370*2+1024= 7760), mode 2 (4 vectors) and mode 3 related with the number of images in each dataset. Figure III.7 presents tensor design based features fusion.

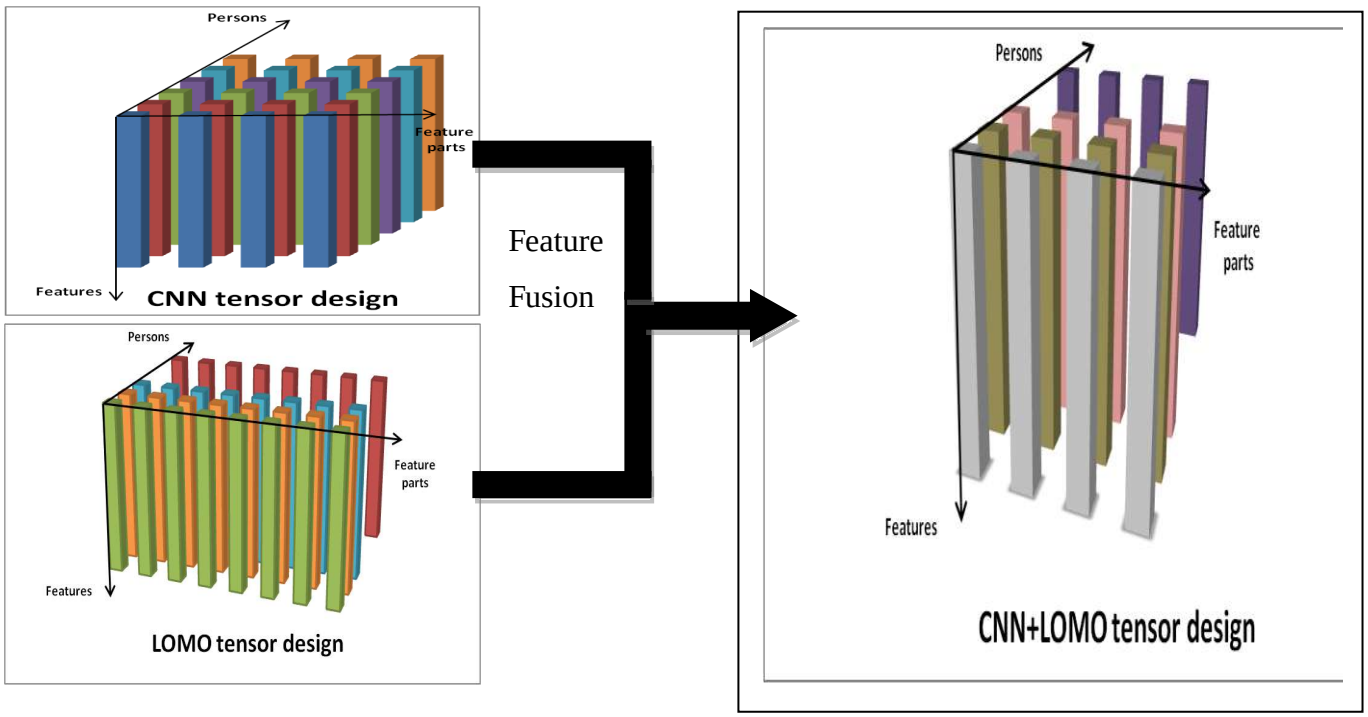


Figure III. 7: The proposed features fusion.

The target dataset is then subjected to metric learning TXQDA to improve discriminative power for person Re-identification. We used the mahalanobis distance, it is a prominent and widely used approach for improving classification results by exploiting the structure of the data [37]. Given n data points$x_i \in R^m$, the goal is to estimate a matrix M such that:

$$d_M(x_i, x_j) = (x_i - x_j)^T M(x_i - x_j) \quad \textbf{(III.1)}$$

Where x_i and $x_j$are two vectors (samples), and M is a positive semi definite matrix.

Because the mahalanobis distance is calculated based on global samples, mutual interference between variables is ruled out. However, when calculating the mahalanobis distance, the required training samples are more than the sample dimensions, otherwise the resulting covariance matrix will be biased [12] .

The cross validation 10-folder strategy is used in the experiments; the three datasets are randomly divided into two sets, training and testing. The test set is equally divided into a gallery / probe sets, in which a gallery set containing the images from camera 1 and the probe set containing images from camera 2. This procedure is repeated for 10 times, the average accuracy of the 10 measures is reported as the result. The goal of PRe-ID is to identify the correct matching image of a probe in the gallery set using the trained mode1. The smaller is the distance between the two

targets, the more they are similar. The ideal case is that the correct matching image Ranks first in the Ranking list, i.e. the correct matching image is the most similar to the probe image.

III.5.2.Discussions

In order to evaluate the performance of method and compare with other the stats-of-the-art methods, we report cumulative matching characteristics (CMC), such as the evaluation metric, at rank-1, rank-5, rank-10, rank-15 and rank-20. The CMC curves of the proposed PRe-ID results along with the CMC curves of existing approaches are shown in Figure III (8-10) for the VIPeR, PRID450s and GRID datasets respectively. In the following, we present the results discussions according to each database.

- **VIPeR**

For the VIPeR datasets the experimental results are given in table III.2 and table III.3 in our experimental settings we select the best feature dimensions , LOMO reported a low Rank-1 accuracy as 21.80% (with 300 dimension) and 69.62% Rank-20 accuracy ,while CNN achieve 38.77% (with 100 dimension) Ranl-1 accuracy and 89.65% for Rank-20. The combination of CNN and LOMO features improved the overall accuracy by 22.87% by achieving 52.41 % Rank-1 accuracy and 16.18% for Rank-20.

Table III. 2 :CMC scores (%) of the used databases VIPeR in CNN and LOMO.

Dimension	CNN					LOMO				
	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20
50	35.22	67.25	79.56	85.28	88.80	16.52	40.85	55.09	62.63	68.23
100	**38.77**	**68.16**	**80.73**	**86.46**	**89.65**	19.40	44.87	57.50	64.94	70.03
150	37.31	68.45	79.43	85.89	89.78	20.73	45.41	57.75	65.13	70.38
200	37.22	67.18	78.67	84.87	88.80	20.98	45.89	57.66	65.00	69.87
250	36.71	65.98	77.18	82.94	86.87	21.20	45.85	58.20	64.56	69.84
300	35.41	63.67	75.66	81.58	86.08	**21.80**	**45.76**	**58.07**	**64.49**	**69.62**
350	34.81	62.41	73.99	80.25	84.30	21.65	45.32	57.97	64.34	69.72

Table III. 3 :CMC scores (%) of the used databases VIPeR in CNN+LOMO.

Dimension	CNN+LOMO				
	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20
50	43.73	76.27	87.37	92.31	94.84
100	49.81	80.95	90.44	94.30	95.92
150	52.12	82.69	90.82	94.53	96.20
200	52.31	83.64	90.70	94.21	96.11
250	**53.16**	**83.32**	**90.19**	**94.02**	**95.82**
300	52.66	83.48	90.19	93.61	95.66
350	52.41	83.35	89.87	93.51	95.63
600	51.71	81.80	89.08	92.75	95.03

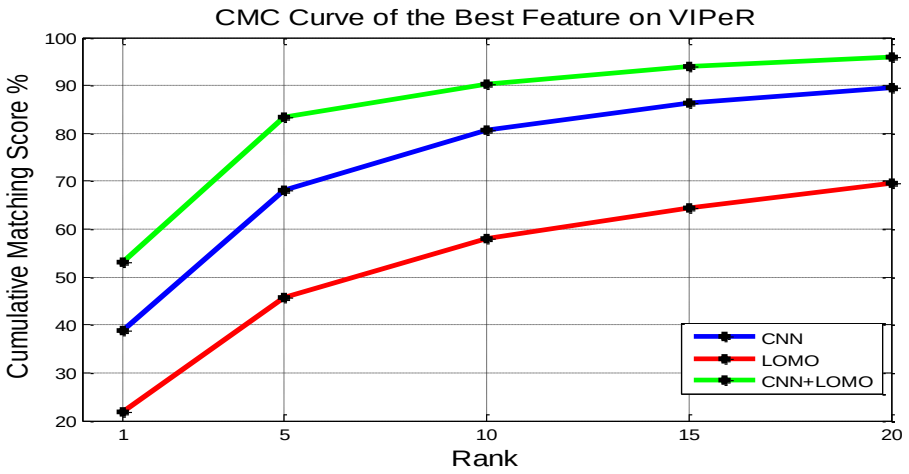


Figure III. 8: CMC curve of the best features on VIPeR.

- **GRID**

The experimental results of this database are shown in table III.4 and table III.5 for the best features. CNN and LOMO features performed almost equivalent with the Rank-1 accuracy rates are 33.92 %(for 100 dimension) and31.44 %(for 300 dimension) respectively. LOMO reported the Rank-20 accuracy as 57.68%, CNN resulted better with 5%. However, the best rates PRe-ID performance is seen when CNN features and LOMO features are combined and resulted in Rank-1 of 86.48% (for 50 dimensions).

Table III. 4 : CMC scores (%) of the used databases GRID in CNN and LOMO.

Dimension	CNN					LOMO				
	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20
50	31.7	45.5	50.5	56.00	59.28	25.52	38.1	44.5	49.36	53.5
100	**33.92**	**47.84**	**53.44**	**57.76**	**62.72**	29.28	41.68	46.64	50.72	55.20
150	33.44	47.68	54.00	58.56	62.96	30.16	43.12	48.72	53.52	57.28
200	33.68	48.16	54.16	58.80	64.16	31.28	43.60	48.72	52.96	56.88
250	33.12	47.84	53.60	58.72	63.52	31.28	44.24	49.76	54.32	57.44
300	32.4	47.6	53.6	58.7	63.6	**31.4**	**45.4**	**50.0**	**54.3**	**57.6**
350	31.76	46.96	53.84	58.56	63.04	32.00	45.20	50.08	54.80	58.16

Table III. 5 : CMC scores (%) of the used databases GRID in CNN+LOMO.

Dimension	CNN+LOMO				
	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20
50	86.48	87.04	87.68	88.16	88.72
100	**86.48**	**87.20**	**88.00**	**88.32**	**88.48**
150	86.32	87.12	87.84	88.32	88.72
200	86.32	87.20	88.00	88.56	88.80
250	86.24	87.12	87.92	88.48	88.88
300	86.32	87.20	87.84	88.56	88.96
350	86.32	87.12	87.76	88.48	88.88

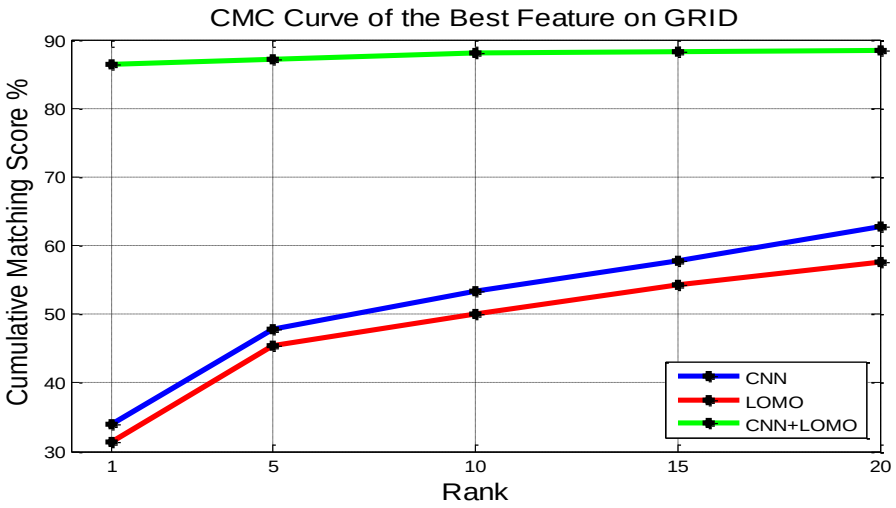


Figure III. 9: CMC curve of the best features on GRID.

- **PRID450s**

The experimental results are presented in tableIII.5 and table III.6. It can be seen that CNN (with 350 dimension) and LOMO(with 150 dimension) descriptors preformed almost the same accuracies 46% with the deference of 0.26% in Rank-1, a slight improvement is seen in LOMO with Rank-20 accuracy of 84.40% in comparison with CNN. Moreover, when the features are combined the overall is improved in Rank-1 accuracy of 70.40% and very high Rank-20 accuracy of 98.76%.

Table III. 6 :CMC scores (%) of the used databases PRID450s in CNN and LOMO.

Dimension	CNN					LOMO				
	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20
50	42.84	67.02	77.24	82.67	86.53	42.44	67.96	76.67	82.27	86.44
100	43.96	65.29	74.67	79.78	83.91	**46.53**	**68.98**	**77.78**	**83.07**	**85.78**
150	44.36	64.31	73.29	77.69	81.42	46.62	68.40	76.62	81.11	84.40
200	45.16	63.96	72.84	77.20	80.44	45.60	66.67	74.84	79.51	83.24
250	44.44	64.71	73.33	77.64	80.71	44.27	65.38	73.42	77.82	81.56
300	46.09	65.82	74.22	78.22	81.91	42.49	63.60	71.24	76.04	79.73
350	**46.36**	**66.31**	**75.16**	**79.69**	**82.84**	40.84	62.04	69.51	73.73	77.69

Table III. 7 : CMC scores (%) of the used databases PRID450s in CNN+LOMO.

Dimension	CNN+LOMO				
	Rank-1	Rank-5	Rank-10	Rank-15	Rank-20
50	62.53	90.09	95.38	97.24	98.27
100	67.73	92.76	96.53	98.09	98.80
150	69.82	93.38	96.40	98.04	98.80
200	**70.40**	**93.64**	**96.44**	**98.04**	**98.76**
250	69.78	93.02	96.22	97.78	98.67
300	69.16	92.80	96.27	97.73	98.76
350	69.47	92.71	96.22	97.82	98.62

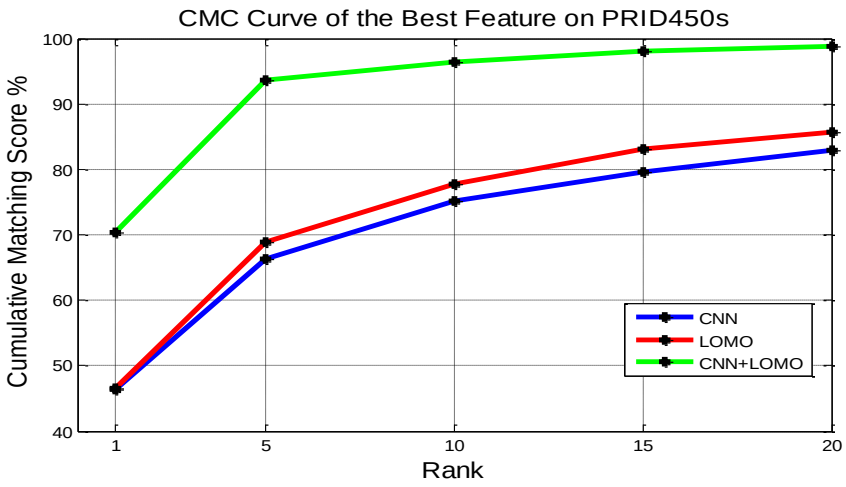


Figure III. 10: CMC curve of the best features on PRID450s.

As we can see CMC curves prove that our approach has a strong discriminative power so that it outperformed in all databases by high accuracies in all ranks. Also, Figure III.11 show the rank-1 rates of VIPeR, GRID, and PRID450s databases. Our approach achieves great results compared to the two other methods (CNN+TXQDA and LOMO+TXQDA). GRID has the highest rank-1 accuracy of 86.48% which is an impressive result according to its complexity compared to the other two databases. We speculate that the proposed approach largely contributes to improve the results

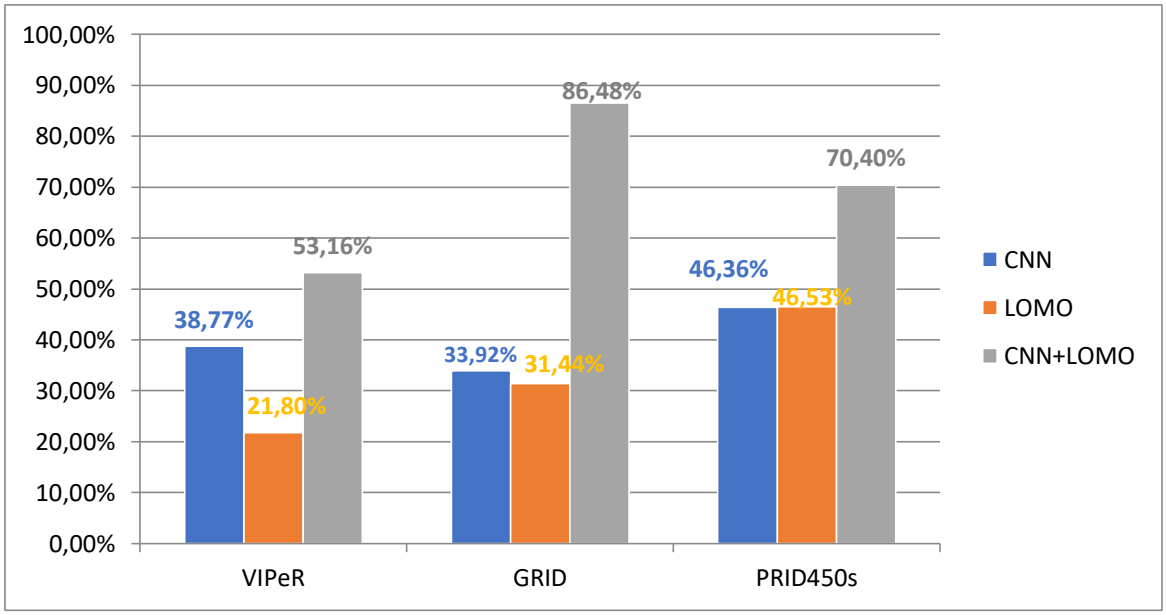


Figure III. 11: The Rank-1 performance on the three datasets VIPeR, GRID and PRID450s.

III.5.3. Comparison with the stat of the art

We compared the performance of the proposed approach against the state-of-the-art via Rank-1 and Rank-20 rates on the three datasets VIPeR, GRID, and PRID450s which are summarized in table III.7 bellow. In this table, the red colored results denote the 1st ranking rates and the 2nd most accurate rates are in bleu.

- **VIPeR**

We compare our method with some state of art methods. Specifically XQDA [55], KEPLER [56], FT-CNN+XQDA [57], SSDAL+XQDA [58], MKSSL+LOMO [59], MKSSL+GOG [59] , EP+LOMO [51] , HIPHOP+LOMO [60], Hybrid+XQDA [40], DIMN [61], Kernel X-CRC[62], VS-SSL[63], DLA[61], EML[64], CNN+XQDA with score normalization [65], Visual-DAC [66], STAR-DAC [66], BNTA [67] and GDNorm [68]. We have the second best results with 53.16% Rank-1 and by arround 95.82% Rank-20 acuuracies. The first best result achieved by GDNorm [68] with 66.1% Rank-1 and Hybrid+XQDA [40] for Rank-20.

- **GRID**

The performance of the proposed approach is compared with some of the existing re-identification approaches introduced in FT-CNN+XQDA [57], SSDAL+XQDA [58], MKSSL+GOG [59], HIPHOP+LOMO [60], Hybrid+XQDA [40], DIMN[61], Kernel X-CRC [62], EML [64], CNN+XQDA with score normalization [65], Visual-DAC [66], STAR-DAC [66], BNTA [67] and GDNorm [68]. The best results obtained using our proposed re-identification approach with Rank-1and Rank 20 accuracies 86.48% and 88.48% respectively. The second best results are for GDNorm [68] with 55.4% Rank-1 accuracy, and 73.30% Rank-20 accuracy for HIPHOP+LOMO [60] approach.

- **PRID450s**

The proposed descriptor is compared with DVR [69], FT-CNN+XQDA [57], SSDAL+XQDA [58], MKSSL+GOG [59], Hybrid+XQDA [40], DIMN[61], Kernel X-CRC [62], VS-SSL [63], EML [64] and CNN+XQDA with score normalization [65].Our method resulted the second best result with 70.40% and 98.76% accuracies for Rank-1 and Rank-20 respectively. Our approach CNN+LOMO+TXQDA achieved the best result of 70.40% and 98.76% Rank-1 and Rank-20 accuracies. CNN+XQDA with score normalization [65] resulted the same Rank-20 accuracy with ours.

Table III. 8: Comparison with the state-of-the-art of Rank-1 and Rank-20 identification rates (%) on the PRID450s, GRID and VIPeR databases.

Approach	Year	VIPeR		GRID		PRID450s	
		Rank-1	Rank-20	Rank-1	Rank-20	Rank-1	Rank-20
DVR [69]	2014	-	-	-	-	28.90	82.80
XQDA[55]	2015	40.00	91.00	-	-	-	-
KEPLER [56]	2015	42.40	90.70	-	-	-	-
FT-CNN+XQDA [57]	2016	42.50	92.00	25.20	64.60	58.20	94.30
SSDAL+XQDA [58]	2016	43.50	89.00	22.40	58.40	22.60	69.2
MKSSL+LOMO [59]	2017	31.20	72.80	-	-	-	-
MKSSL+GOG [59]	2017	40.60	85.90	24.60	64.20	61.60	96.70
EP+LOMO [51]	2017	52.84	86.71	-	-	-	-
HIPHOP+LOMO [60]	2017	54.20	96.90	26.00	**73.30**	-	-
Hybrid+XQDA [40]	2018	50.50	**97.20**	25.50	68.20	66.50	97.00
DIMN[61]	2019	51.20	-	29.30	-	39.20	-
Kernel X-CRC [62]	2019	51.60	95.3	26.60	69.7	**68.80**	**98.4**
VS-SSL[63]	2020	43.90	87.80	-	-	63.30	97.00
DLA [61]	2020	50.89	94.87	-	-	-	-
EML[64]	2020	44.37	-	19.47	-	63.58	-
CNN+XQDA with score normalization[65]	2021	43.16	92.78	35.68	61.92	64.22	**98.76**
Visual-DAC[66]	2022	39.70	-	34.60	69.10	-	-
STAR-DAC [66]	2022	40.80	-	41.20	70.60	-	-
BNTA[67]	2022	57.40	-	51.10	-	-	-
GDNorm[68]	2022	**66.10**	-	**55.4**	-	-	-
CNN+TXQDA(ours)	2022	38.77	89.65	33.92	62.72	46.36	82.84
LOMO+TXQDA(ours)	2022	21.80	69.62	31.44	57.68	46.53	85.78
CNN+LOMO+TXQDA (ours)	2022	**53.16**	**95.82**	**86.48**	**88.48**	**70.40**	**98.76**

Our approach is better than some other, This proves the effectiveness and competitiveness of our method. From the above experiments on various datasets, we can see that our method is the best on certain datasets and surpasses existing methods. Although not the best on some datasets, it is classed the second compaired to the existing deep learning methods. In conclusion, our proposed approach is generally successfull.

III.6. Conclusion

In this work, we tackled the PRe-ID problem using the proposed model feature fusion of the two efficient descriptors LOMO and CNN and metrics learning approach called TXQDA. The CNN and LOMO was used as features representations and they show optimistic results due to its ability and capacity on feature learning. Show the great impact of the re-identification process on the three challenging datasets VIPeR, GRID and PRID450s. Finally, to evaluate our results we've reported the Cumulative Matching Characteristic curves and we compared with the stat of the art.

General conclusion

In the context of smart video surveillance interested in PRe-ID in non-overlapping camera networks. The person's appearance may change significantly between camera views due to different factors related essentially to the conditions of the recording medium and the person's passage through the camera's field of view what make the Re-identification of person of interest a very challenging task. Researchers are trying to improve the performance of the PRe-ID system by developing a robust feature representation and a discriminant metric learning.

The goal of our work was to get a better approach in terms of accuracy for the PRe-ID. The multidimensional representation of data on three challenging datasets VIPeR, PRID and GRID, after using the new descriptor that combine CNN and LOMO based tensor representation to get a best discriminant features with the subspace projection based TXQDA.

Our project can be summarized as follows:

- A general description of biometrics, biometrics' modalities and its applications, also the literature of person Re-identification along with its pipeline, challenges and its application.
- An overview of Artificial Intelligence and machine learning, the classification of its algorithms and its applications.
- The use of a discriminant metric learning called Tensor Cross-view Quadratic Discriminant Analysis.
- Show the great impact of the re-identification process on the three challenging datasets VIPeR, GRID and PRID450s.

It would be interesting as a future work to apply the fusion of CNN and LOMO with other feature representations like the GOG (Gaussian Of Gaussian) descriptor.

References

[1] Singh, N.K., M. Khare, and H.B. Jethva, *A comprehensive survey on person re-identification approaches: various aspects.* Multimedia Tools and Applications, 2022: p. 1-45.

[2] Guo, R., et al., *Density-adaptive kernel based efficient reranking approaches for person reidentification.* Neurocomputing, 2020. **411**: p. 91-111.

[3] Vishwakarma, D.K. and S. Upadhyay, *A deep structure of person re-identification using multi-level Gaussian models.* IEEE Transactions on Multi-Scale Computing Systems, 2018. **4**(4): p. 513-521.

[4] Bolle, R.M., et al., *Guide to biometrics*. 2013: Springer Science & Business Media.

[5] Sharma, A.K., A. Raghuwanshi, and V.K. Sharma, *Biometric System-A Review.*

[6] Bhatia, R., *Biometrics and face recognition techniques.* International Journal of Advanced Research in Computer Science and Software Engineering, 2013. **3**(5).

[7] Wang, Y., T. Tan, and A.K. Jain. *Combining face and iris biometrics for identity verification*. in *International conference on Audio-and video-based biometric person authentication*. 2003. Springer.

[8] Chouchane, A., *Analyse d'images d'expressions faciales et orientation de la tête basée sur la profondeur*, 2016, Université Mohamed Khider-Biskra.

[9] Mohammed, S. and M. Hegarty, *Evaluation of Voice & Ear Biometrics Authentication System.* International Journal of Education and Management Engineering, 2017. **7**(4):

[10] Belhadj, F., *Biometric system for identification and authentication*, 2017, Ecole nationale Supérieure en Informatique Alger.

[11] Dargan, S. and M. Kumar, *A comprehensive survey on the biometric recognition systems based on physiological and behavioral modalities.* Expert Systems with Applications, 2020. **143**: p. 113114.

[12] Wu, W., et al., *Deep features for person re-identification on metric learning.* Pattern Recognition, 2021. **110**: p. 107424.

[13] Su, C., et al. *Pose-driven deep convolutional model for person re-identification*. in *Proceedings of the IEEE international conference on computer vision*. 2017.

[14] Chen, Y., *Person re-identification in images with deep learning*, 2018, Lyon.

[15] Nie, W., et al., *Single/cross-camera multiple-person tracking by graph matching.* Neurocomputing, 2014. **139**: p. 220-232.

[16] Henriques, J.F., et al. *Exploiting the circulant structure of tracking-by-detection with kernels*. in *European conference on computer vision*. 2012. Springer.

[17] Suma, V., *Computer vision for human-machine interaction-review*. Journal of trends in Computer Science and Smart technology (TCSST), 2019. **1**(02): p. 131-139.

[18] Souded, M., *People detection, tracking and re-identification through a video camera network*, 2013, Université Nice Sophia Antipolis.

[19] Zahra, A., et al., *Person Re-identification: A Retrospective on Domain Specific Open Challenges and Future Trends.* arXiv preprint arXiv:2202.13121, 2022.

[20] Khedher, M.I., *Ré-identification de personnes à partir des séquences vidéo*, 2014, Institut National des Télécommunications.

[21] https://towardsdatascience.com/cross-validation-c4fae714f1c5.

[22] Berrar, D., *Cross-Validation*, 2019.

[23] Charniak, E., *Introduction to artificial intelligence*. 1985: Pearson Education India.

[24] SRIDHAR, G.S., *ARTIFICIAL INTELLIGENCE(WAY TO FUTURE)*. G.S.SRIDHAR.

[25] Kakani, V., et al., *A critical review on computer vision and artificial intelligence in food industry*. Journal of Agriculture and Food Research, 2020. **2**: p. 100033.

[26] Osuwa, A.A., E.B. Ekhoragbon, and L.T. Fat. *Application of artificial intelligence in Internet of Things*. in *2017 9th international conference on computational intelligence and communication networks (CICN)*. 2017. IEEE.

[27] Reshamwala, A., D. Mishra, and P. Pawar, *Review on natural language processing.* IRACST Engineering Science and Technology: An International Journal (ESTIJ), 2013. **3**(1): p. 113-116.

[28] Harkut, D.G., & Kasat, K, *Artificial Intelligence - Scope and Limitations*. 2019.

[29] https://www.nature.com/articles/nbt1206-1565.

[30] Harkut, D.G. and K. Kasat, *Introductory chapter: artificial intelligence-challenges and applications*, in *Artificial Intelligence-Scope and Limitations*. 2019, IntechOpen.

[31] https://www.upgrad.com/blog/top-challenges-in-artificial-intelligence/.

[32] El Naqa, I. and M.J. Murphy, *What is machine learning?*, in *machine learning in radiation oncology*. 2015, Springer. p. 3-11.

[33] Rebala, G., A. Ravi, and S. Churiwala, *An introduction to machine learning*. 2019: Springer.

[34] Bell, J., *Machine learning: hands-on for developers and technical professionals*. 2020: John Wiley & Sons.

[35] Alaoui, S.S., Y. Labsiv, and B. Aksasse, *Classification algorithms in data mining.* Int. J. Tomogr. Simul, 2018. **31**: p. 34-44.

[36] Mahesh, B., *Machine learning algorithms-a review*. International Journal of Science and Research (IJSR).[Internet], 2020. **9**: p. 381-386.

[37] Roth, P.M., et al., *Mahalanobis distance learning for person re-identification*, in *Person re-identification*. 2014, Springer. p. 247-267.

[38] Noble, W.S., *What is a support vector machine?* Nature biotechnology, 2006. **24**(12): p. 1565-1567.

[39] https://jain-khushboo20.medium.com/k-means-clustering-unsupervised-machine-learning-algorithm-5a09bd1dbd2.

[40] Perwaiz, N., M.M. Fraz, and M. Shahzad, *Person re-identification using hybrid representation reinforced by metric learning.* IEEE Access, 2018. **6**: p. 77334-77349.

[41] Hurwitz, J. and D. Kirsch, *Machine learning for dummies.* IBM Limited Edition, 2018. **75**.

[42] Kelleher, J.D., *Deep learning*. 2019: MIT press.

[43] Goodfellow, I., Y. Bengio, and A. Courville, *Deep learning*. 2016: MIT press.

[44] Tan, C., et al. *A survey on deep transfer learning*. in *International conference on artificial neural networks*. 2018. Springer.

[45] Ghosh, A., et al., *Fundamental concepts of convolutional neural network*, in *Recent Trends and Advances in Artificial Intelligence and Internet of Things*. 2020, Springer. p. 519-567.

[46] Aburasain, R., *Application of convolutional neural networks in object detection, re-identification and recognition*, 2020, Loughborough University.

[47] Krizhevsky, A., I. Sutskever, and G.E. Hinton, *Imagenet classification with deep convolutional neural networks.* Advances in neural information processing systems, 2012. **25**.

[48] Bruslund Haurum, J., et al. *Re-identification of zebrafish using metric learning*. in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision Workshops*. 2020.

[49] Bessaoudi, M., et al., *Multilinear side-information based discriminant analysis for face and kinship verification in the wild.* Neurocomputing, 2019. **329**: p. 267-278.

[50] Laiadi, O., et al., *Tensor cross-view quadratic discriminant analysis for kinship verification in the wild.* Neurocomputing, 2020. **377**: p. 286-300.

[51] Li, X., L. Liu, and X. Lu, *Person reidentification based on elastic projections.* IEEE transactions on neural networks and learning systems, 2017. **29**(4): p. 1314-1327.

[52] Yang, Y., et al., *Pixel and Feature Transfer Fusion for Unsupervised Cross-Dataset Person Reidentification.* IEEE transactions on neural networks and learning systems, 2021.

[53] Deng, X., et al., *A deep multi-feature distance metric learning method for pedestrian re-identification.* Multimedia Tools and Applications, 2021. **80**(15): p. 23113-23131.

[54] Gou, M., et al., *A systematic evaluation and benchmark for person re-identification: Features, metrics, and datasets.* IEEE transactions on pattern analysis and machine intelligence, 2018. **41**(3): p. 523-536.

[55] Liao, S., et al. *Person re-identification by local maximal occurrence representation and metric learning*. in *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[56] Martinel, N., C. Micheloni, and G.L. Foresti, *Kernelized saliency-based person re-identification through multiple metric learning.* IEEE Transactions on Image Processing, 2015. **24**(12): p. 5645-5658.

[57] Matsukawa, T. and E. Suzuki. *Person re-identification using CNN features learned from combination of attributes*. in *2016 23rd international conference on pattern recognition (ICPR)*. 2016. IEEE.

[58] Su, C., et al. *Deep attributes driven multi-camera person re-identification*. in *European conference on computer vision*. 2016. Springer.

[59] Sun, C., D. Wang, and H. Lu, *Person re-identification via distance metric learning with latent variables.* IEEE Transactions on Image Processing, 2016. **26**(1): p. 23-34.

[60] Chen, Y.-C., et al., *Person re-identification by camera correlation aware feature augmentation.* IEEE transactions on pattern analysis and machine intelligence, 2017. **40**(2): p. 392-408.

[61] Song, J., et al. *Generalizable person re-identification by domain-invariant mapping network*. in *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*. 2019.

[62] Prates, R. and W.R. Schwartz, *Kernel cross-view collaborative representation based classification for person re-identification.* Journal of Visual Communication and Image Representation, 2019. **58**: p. 304-315.

[63] Jia, J., et al., *View-specific subspace learning and re-ranking for semi-supervised person re-identification.* Pattern Recognition, 2020. **108**: p. 107568.

[64] Ma, W., et al., *A new date-balanced method based on adaptive asymmetric and diversity regularization in person re-identification.* International Journal of Pattern Recognition and Artificial Intelligence, 2020. **34**(09): p. 2056004.

[65] Afaf., B. and B. Chahrazed., *Person Re-identification Based CNN and Score Normalization*, 2021, Yahia Fares universty of Medea

[66] Prasad, M.V. and R. Balakrishnan, *Spatio-Temporal association rule based deep annotation-free clustering (STAR-DAC) for unsupervised person re-identification.* Pattern Recognition, 2022. **122**: p. 108287.

[67] Han, K., et al., *Generalizable Person Re-Identification via Self-Supervised Batch Norm Test-Time Adaption.* arXiv preprint arXiv:2203.00672, 2022.

[68] Liu, J., et al., *Debiased batch normalization via gaussian process for generalizable person re-identification.* arXiv preprint arXiv:2203.01723, 2022.

[69] Wang, T., et al. *Person re-identification by video ranking*. in *European conference on computer vision*. 2014. Springer.