N° **D'ORDRE: 41490**

ÉCOLE DES MINES DE DOUAI
Unité de Recherche en Informatique et Automatique (URIA)
UNIVERSITÉ LILLE 1 SCIENCES ET TECHNOLOGIES

THÈSE

pour l'obtention du

Doctorat délivré conjointement par
l'École des Mines de Douai et l'Université Lille 1 Sciences et Technologies
Spécialité: Automatique, Génie Informatique, Traitement du Signal et des Images

présentée par

Yanyun LU

Online classification and clustering of persons using appearance-based features from video images – Application to person discovery and re-identification in multicamera environments

Thèse soutenue le 26 septembre 2014 devant le jury:

Président:	Ludovic MACAIRE	Professeur, Université Lille 1 Sciences et Technologies
Rapporteurs:	Catherine ACHARD	Maître de conférences HDR, UPMC
	Belkacem FERGANI	Directeur de recherches, Université d'Alger
Examinateurs:	Frédéric LERASLE	Professeur, UPS et groupe RAP (LAAS-CNRS)
	Ludovic MACAIRE	Professeur, Université Lille 1 Sciences et Technologies
Directeurs:	Stéphane LECOEUCHE	Professeur, Mines Douai
Co-encadrant:	Anthony FLEURY	Maître de conférences, Mines Douai
	Jacques BOONAERT	Maître de conférences, Mines Douai

Abstract

Video surveillance is nowadays an important topic to address, as it is broadly used for security and it brings problems related to big data processing. A part of it is identification and re-identification of persons in multicamera environments. The objective of this thesis work is to design a complete automatic appearance-based human recognition system working in real-life environment, with the goal to achieve two main tasks: person re-identification and new person discovery. The proposed system consists of four modules: video data acquisition; background extraction and silhouette extraction; feature extraction and selection; and person recognition. For evaluation purposes, in addition to the public available CASIA Database, a more challenging new database has been created under low constraints. Grey-world normalized color features and Haralick texture features are extracted as initial feature subset, then features selection approaches are tested and compared. These optimized subsets of features are then used firstly for person re-identification using Multi-category Incremental and Decremental SVM (MID-SVM) algorithm with the advantage of training only with few initial images and secondly for person discovery and classification using Self-Adaptive Kernel Machine (SAKM) algorithm able to differentiate existing persons who can be classified from new persons who have to be learned and added. The proposed system succeed in person re-identification with classification rate of over 95% and achieved satisfying performances on person discovery with accuracy rate of over 90%.

Contents

List of Figures

List of Tables

List of Algorithms

Introduction

Context

Nowadays video surveillance systems are widespread in various domains that range from public transport to security in towns, from home surveillance system to production line of factory. Numerous automatic methods have been proposed to fast and efficiently extract effective information from huge recorded video data. Object recognition, especially human recognition, is the key assignment in video surveillance system. Human beings can quickly and accurately identify a person or differentiate persons. However, this task is not so easy for computer and giving it capacities to identify people as precisely as human being is still a challenge that has huge prospects in applications. It is still a challenge but has huge prospects in applications. That is the reason why so many researchers still study this topic.

There are many problems that have not yet been addressed in the area of video-based human recognition, such as to deal with the changes in clothes, location or pose, etc. Many human recognition systems can get very accurate results but are designed to operate in highly constrained environments, such as no shelter, single or limited number of person, no change in illumination, location or posture. In realistic environment, not only the information of person is changed due to the different possible positions, poses, expressions or clothes, but also the environmental conditions that could vary with time (illumination, occlusions, etc.).

Objectives of this work to address and problems in online human recognition

The objective of this thesis is to design an online human recognition system to automatically achieve person recognition and new person discovery in real-life environments constituted of one or several video-surveillance cameras. In this research, we investigate the human recognition system not only working in closed indoor environment (e.g., company, office or laboratory) but also involving in open pubic area (e.g., garden or train station). It is a challenge task to perform online person recognition due to the numerous changes in both intrinsic attributes of person and extrinsic environmental conditions.

First of all, how to effectively represent a person is essential and significant. In certain video sequences in real-life environments, the face of a person can not be seen due to the fact that he/she is back to the camera. Since this case possibly happens, face-based features human recognition approaches are not suitable for our system. Besides, gait-based approaches are sensitive to multiple persons, walking speed changes, etc., as well as they require an optimal silhouette of the person [WTNH03]. Appearance, that contains the clothes and the visible parts of persons, could be a good choice for the representation, specially for the case of novelty detection without required supervisor.

Various features based on simple color or texture, or based on more complicated feature descriptors can be extracted for human representation. What kind of features can be easily extracted and can effectively represent a person is an addressed problem in this work. Besides, the extracted features of persons obtained in video sequences can be various and numerous but redundant. To select efficient features is important for improving classification performance. Several features selection methods have to discussed before using them in the proposed human recognition system.

Since the recognition system is applied in real-life environments with a tremendous amount of possible variations, it is not possible to have a complete knowledge of the person in order to online recognize him/her in video sequences. To improve the classification performance, the traditional approaches of person recognition require a large percentage images to train, so that the variations of the person's presentation or of the environment are taken into account in the model [FRP07]. However, these approaches are slow and time consuming, and in most of the cases cannot be effectively applied in real-life applications. Solutions including an updated the classifier and learn incrementally should be an improvement to overcome these problems. As a consequence, in this thesis, we will present an online supervised learning algorithm with incremental and decremental capabilities to recognize persons in online setting with challenging conditions (illumination and person pose changing, bag, paper or a cup of coffee carrying), which only needs a few training images for initial learning.

Since the applications of this work can be either indoor (closed environment) or outdoor (public environment), two specific applications are interesting to mention: person re-identification and new person discovery. First of all, the recognition system should have the ability to detect persons in video sequences, to identify or re-identify them from image captured by one or multiple cameras, and also to work in an online setting from the video feeds with variable and challenging conditions. Furthermore, in outdoor public environments or in special security surveillance systems, the recognition system should have self-learning ability to differentiate new persons, to create new models for unknown persons, and to adaptively classify them. This task cannot be achieved by the classical supervised learning algorithm and will be presented using an original method. Indeed, in this new person discovery application, we have no knowledge of the entire set of person that could come or leave the area (and supervised learning required the knowledge of the number of classes and of an initial set of samples for each class to initialize the classifier). Another unsupervised learning algorithm will be presented and used to achieve new person discovery in online setting.

Organization of the thesis

This thesis will be organized as follows: Chapter 1 will present a brief review of video-based human recognition based on the features of face, gait and appearance. In Chapter 2, an offline human recognition system is described, in which an initial set of features related to appearance are extracted and the more relevant features are selected by three different methods (PCA, Wrapper and Correlation-based feature selection). Four feature datasets (the whole feature dataset and the other three achieved by feature selection) are prepared and will be used after for the different human recognition tasks. Chapter 3 reminds the properties and the functioning of an online supervised learning algorithm (MID-SVM), which is applied to automatically recognize persons in online setting. Then another unsupervised learning algorithm (SAKM) is introduced and employed to address the problem of new person discovery. The experimental results on a publicly available database (CASIA) show the performances of these two algorithms. In Chapter 4, two specific applications of the proposed human recognition system are taken into account: person recognition and new person discovery. Recognition experiments on our own database (IA Database) will test the performances of two proposed algorithms in these two specific applications. Finally, a conclusion will give an overview of the contributions of the thesis and present the future works that could improve the performance of the previously presented human recognition system.

Chapter 1

Human recognition system in video surveillance

1.1 Introduction

Human recognition approaches are used in various domains that range from public transport to security surveillance, from entertainment (e.g., human computer interaction) to smart cards (e.g., passports). In this work, we focus on the application of human recognition in video surveillance and do not consider more details in perception tasks, such as activity recognition or pose recovery.

Human recognition is an important but still challenging problem in video surveillance. In real-life systems, person detection becomes a painful task owing to the fact that there could be difficult to follow a person in an open environment, with possible interacts and occlusion. Many video surveillance systems need a group of cameras cooperating to recognize persons, which bring new difficulties for person re-identification through the network of cameras. It is difficult to manage the large volume of information of multiple cameras, and due to changes of illumination or of sensor characteristics, it is also difficult to find a common and optimised set of features to describe the person. Generally, features based on face, gait and appearance are extracted for person recognition. The choice of the kind of features is related to the aim of the specific application and to the conditions of the environment (resolution of the cameras, etc.). No matter what kind of features are extracted to represent a person, more precise features give more information of the person, which should result in more discrimination power. However, high dimensional features easily have overfitting problem. Besides, they inevitably contain irrelevant and redundant information. Feature selection is necessary and useful to identify and remove these useless information and keep the main useful information.

The organization of this chapter is as follows: after this introduction, an overview of general human recognition system is given. Then, based on various human representations, face-based, gait-based and appearance-based human recognition are reviewed respectively, including their main methods, applications and difficulties. After a short comparison of these three categories, four representative works on appearance-based human recognition are presented in more details. Considering all these development, a conclusion is finally drawn.

1.2 Overview of human recognition systems

Generally, human recognition system in video surveillance consists of two main tasks: (1) people detection and tracking; (2) people recognition. Figure 1.1 gives a general layout of such a system. The input data are static images or video sequences captured by one or several cameras. The pre-processing consists of background subtraction, silhouette segmentation, etc. Human features are extracted for person representation and a final feature set is considered as input data of a classifier. In most cases, in order to obtain efficient features and to reduce the dimension, feature selection process is done before classification.

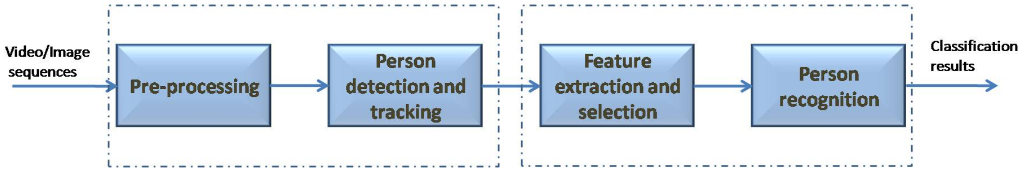


Figure 1.1: General framework of human recognition system

An efficient human recognition system should have the following abilities: (1) detect and track persons in video sequence; (2) classify multiple persons in online applications, from possibly different cameras, with variable conditions (illumination changing, person overlapping, etc.); (3) recognize any person including one who is totally new to the system and have self-learning abilities to update the classifier.

As a consequence, there are many problems for human recognition occurring in real-life applications. The first is that the system should be able to on-line adapt to the changes that can occur (e.g., clothes, expression or orientation) or to the variation of the environmental conditions (e.g., illumination, location and camera angle). Current human recognition systems, which get accurate results, are designed to operate in highly constrained environments, such as single person or constant artificial lighting [DSTR11].

Secondly, it is difficult to have a complete knowledge for person recognition, especially in online applications and in open spaces in which people can come and leave freely. Features extracted for human representation are generally based on simple color, texture, or complicated descriptors, such as bag-of-features model [SC00]. What kind of feature should be extracted to correctly represent different persons is still an open problem.

Furthermore, in real-world applications, we need not only to achieve person identification, but also to solve the problem of new person discovery. It is a challenge to identify new persons from an existing database and save them as new classes in the recognition system.

The existing researches started to address these problems, such as in [BRL$^+$11, KDM$^+$09], human recognition system are applied in online settings. Person re-identification task has also been studied on person representation and classification [BMP04, BDSP07, TCAKD09]. Generally, in most of the video surveillance systems, human recognition is based either on biometric attributes (face or gait) or on non-biometric attribute (appearance). In the next section, face-based, gait-based and appearance-based human recognition are described with their respective methods, applications and difficulties.

1.3 Representation methods

As presented in Section 1.2, human representation is generally based on features extracted from face, gait and appearance or some specific models depending on the application or on the environmental conditions. In this section, we will detail a literature review of human recognition based on face, gait and appearance, respectively.

1.3.1 Face-based human recognition

1.3.1.1 Literature review

Face perception is an important part for a human recognition system. Compared with the other biometric attributes (voice, iris, fingerprints, hand veins, etc.), face attributes are easy to obtain with high resolution cameras and are practical to apply in video sequences. Face-based recognition technologies have numerous applications: security surveillance, criminal justice systems, age and gender classification, human computer interfaces, social networks, indexation, etc. [FGH10, SSM07, MY02b, SWZ$^+$09, HMD13].

Depending on the method of face data computation, face-based recognition can be divided into three categories [JA09]: face data from static images, from video sequences and from the other sensory inputs (e.g., infra-red imagery). The methods for acquiring face data depend on the underlying application. Here, we pay more attention to face recognition from video sequences. A general statement of face recognition from video can be formulated as follows: given sequences of a scene, (1) detect (and track) human face, (2) represent face image by the extracted features, (3) identity (and classify) a person. As most video-based face recognition systems choose good image frames from video sequences and apply similar techniques that for static images to identify persons, a brief review of face-based recognition methods for static images is presented hereafter.

Generally, face recognition methods from video or images could be divided into three categories according to the representation: Feature-based matching methods, Holistic matching methods and Hybrid methods.

- **Feature-based matching methods [BP93, CGY96]:** In these methods, facial component features such as eyes, mouth and nose are extracted and their geometric parameters are measured. Then, these parameters are used for classification by statistical pattern recognition techniques. Pure geometry features are used for face recognition in early periods, such as distance between eyes and width of the head [Kan74]. Then deformable templates [Yui91, FGMR13], Hough transform [FYN$^+$12] and other templates [WL11, BM10] are used for face recognition. In 1997 [WFKvdM97], Wiskott et al. proposed a face recognition method, named Elastic Bunch Graph Matching (EBGM), in which Gabor wavelet coefficients are usually used for local feature representation. Afterward, this method have been widely used in face recognition [SHH09, HGV13]. Since feature-based methods measure structural characteristics of each special face, it is easy to get high speed matching of the image to that of the person. However, it is difficult to decide which features are better to represent the other different faces [JA09].

- **Holistic matching methods [KP07, KLJ11]:** These methods are based on the entire face image instead of facial component features. Eigenfaces [TP91], Fisherfaces [BHK97] and Gabor wavelets [LW02] are widely used to represent face appearance. Statistical methods or Artificial Intelligence (AI) methods are then used for recognition. In statistical methods, Principal Component Analysis (PCA) [Bou09, ZB08], Independent Component Analysis (ICA) [BMS02], and Fisher Discriminant Analysis (FDA) [LHLM02] are widely used. Machine learning techniques, as SVM, HMM and NN can also be used to classify face images. The main advantage of holistic methods is to keep the completeness of face information, instead of concentrating only on some regions of the image. However, this property is also the disadvantage of such methods, which causes high computational cost. Besides, the features extracted by such methods are not invariant with the changes of illumination, pose, etc., due to the fact that holistic methods consider all the pixels of an image with equal importance. As a consequence, the models are difficult to construct.

- **Hybrid methods**: These methods consider both the whole face region and facial component features. As discussed above, feature-based methods and holistic methods have their respective advantages and disadvantages. In some applications, hybrid methods could obtain better performances. In [HAA07], holistic features and facial component features (nose, mouth, left and right eyes) are combined, the experimental results show that the proposed method performs better than Eigenface method. In [Bou09], a hierarchical method is proposed for face recognition, and performances of classification are compared, based on Eigenfaces, SVM and PCA combined with facial components (right face, top part of face, left and right mirrored face, etc.). The results show that the combined features work well on two benchmark data sets (Yale faces and ATT faces) and always achieve better classification rate than any the individual features.

Since one important application of face recognition is security surveillance, the researches on video-based face recognition approaches are going on in these years. Most of these approaches [CGS12, MR03] are originated from still-image-based methods, which are discussed above. Video-based recognition system should have the ability to automatically detect, segment the face from an image sequence, and then identify the face in real-time using still-image-based face recognition methods. Comparing with recognition from still images, face recognition from video images appear to be disadvantaged: low quality images, cluttered backgrounds and a large amount of data to process [PRJ13].

1.3.1.2 Difficulties of face-based recognition

Even though major advances have been achieved in face recognition, real-world application in open environment remains a challenge. It comes from the fact that face acquisition process can be affected by lots of conditions. Following, general difficulties and restrictions for face recognition are given [ANRS07]:

- **Strong limit in face recognition:** Face should be detected first, that is to say, the image should contain a face in frontal or side view at least. However, in real-life video surveillance applications, there are all kinds of possibilities (back view, top view, overlap, etc.).

- **Influences of intrinsic factors:** Intrinsic factors are caused by the physical nature of the face, such as the changing of facial expression, hair style, glasses, etc. For example, face features are different when people are crying or laughing.

- **Influences of extrinsic factors:** Extrinsic factors (illumination, pose, camera resolution, focus, etc.) affect the face image by the interaction of light. For example, changes in the pose can introduce projective deformations and self-occlusions. Face recognition systems have a real need of high resolution images.

Furthermore, as mentioned in [ZCPR03], video-based recognition has more challenges:

- **Small face image:** Since surveillance cameras are generally installed upright (on a ceiling or a wall), face images from video sequences are often small, which affect both face detection and classification.

- **Low video quality:** Video sequences are taken indoors or outdoors with large and random variations of environmental conditions, which may affect the quality of the images captured from video. Moreover, video images have to be compressed for easy storage and transmission, which could yield to important information loss.

- **Characteristics of faces:** In video sequence, the changes of face are smaller and harder to detect comparing to the whole body, which cause that some face detection and recognition approaches can not be applied.

As a consequence, it is quite difficult to compute sophisticated face features in on-line video human recognition system.

1.3.1.3 Conclusion

Face recognition has a wide range of applications (e.g., human computer interaction, image and social networks and security surveillance) and is a very important topic in human recognition. However, it has some difficulties due to the fact that face features could be affected by numerous factors, such as head pose, glasses, hairstyle, facial expression and occlusions [ZCPR03]. In particular, video-based face recognition is possible only when the person has an adapted orientation comparing to the camera and is close enough for sufficient details on the image.

1.3.2 Gait-based human recognition

1.3.2.1 Literature review

Gait attributes contain the body shape (such as height, leg length, etc.) and the dynamic of the person while walking or moving. Generally, gait recognition is used to detect, track, and recognize persons or to analyse human behaviours. Gait-based recognition technologies offer potential for human recognition when persons are at a distance from the camera or with low image resolution. Besides, they have possibilities to get good results when a person is back to the camera or changes his clothes. There are lots of work on gait recognition methods and also several widely used databases, such as CASIA Gait Database [CAS01], CMU Motion of Body Database [GS01].

Gait recognition are divided into three categories according on the sensor used for gait information acquisition [PHS11]: floor sensor (FS) based, wearable sensor (WS) based and machine vision (MV) based. Several applications based on these sensors are used for medical purposes. In MV-based recognition, gait is obtained by a video-camera at a distance without other special equipment. In general, two main techniques exist for MV-based gait recognition: model-based approaches and motion-based approaches.

- **Model-based gait recognition [YNC02, TB01]:** These methods use models that define human bodies or their motions. For each frame, model matching is performed and parameters are calculated for gait representation. Ellipse [XCL$^+$10] or Skeleton [MZBH09] model are used to define the body. Besides, Hough transform [RWZD07], Fourier analysis [NCC$^+$02] or Hidden Markov Model [Mey97] are adopted to measure gait features from leg angles or trajectories of head and feet, etc. Then these features are used for setting up the gait model for recognition. The main disadvantage is that it is hard to determine the best efficient model for each specific scenario. For the moment, the choice of the model is empirical. Furthermore, model-based methods are time consuming.

- **Motion-based gait recognition:** Most of the researches on gait recognition are based on motion-based methods, which often operate on human silhouettes. The processing includes first a background subtraction, then the segmentation of the silhouette, and finally gait recognition. These approaches generally consist of three major parts: human detection and tracking, gait feature extraction and gait classification. Existing methods can be divided into two main categories: state-space methods and spatio-temporal methods. The former methods represent the person motion as a sequence of static poses by considering temporal gait variations. In [HHN99], human shape features and their temporal modifications are used to discriminate different gaits. HMMs can also be successfully used for gait recognition [SR03]. Spatio-temporal methods describe gait features by the distribution of motion through space and time. In [HAN02], an automatic gait recognition method based on spatio-temporal symmetry is described. Philips et al. proposed a baseline algorithm for gait recognition and presented its performance on a large and challenging database [PSR+02]. Wang and Tan proposed a simple and efficient algorithm for gait recognition using spatial-temporal silhouette analysis, in which 2D silhouette image is converted into 1D distance signal by eigenspace transformation [WTNH03].

1.3.2.2 Difficulties of gait-based recognition

Although the performances of gait-based recognition are encouraging, there are some factors which could negatively affect the performance. In [SPL+05], Sarkar et al. create a challenging database with 122 persons in 32 different sequences, which includes the changes in five factors: camera angle (right view or left view), shoe type, walking surface (grass or concrete surface), carrying or not carrying a briefcase and time. The experimental results show that the classification rates range from 78% on the easiest experiment (in which the difference is only on shoes) to 3% on the hardest one (differences on surface, time, shoes and clothes). In general, there are some challenges in gait recognition described as follows:

- Multiple persons: to the best of our knowledge, there is no good solution for gait recognition when there are multiple persons in one image, especially when they are overlapping.

- Speed of walk: gait attributes are quite different when people are running or walking.

- Physiological changes: some internal factors cause gait changes, such as foot injury, sickness, lower limb disorder, losing or gaining weight, etc.

- Object carrying: in real-life condition, the fact that persons often carry handbag, knapsack or laptop causes variable gait features.

- The angles of observation: variant views (frontal view, right-side or left-side view) could give variant gait shapes. As proven in [SPL+05], right and left views have different gait silhouette shapes and strictly frontal view is impossible to handle when using gait period detection method [LS04].

- Walking surface conditions: person walks in different style depending on ground surface (e.g. wet or dry, grass or concrete surface).

- There are other factors, such as lighting condition (day or night, indoor or outdoor) and shoe types (sport shoes or high-heeled shoes). Change of lighting condition affects image resolution and acquisition rate, which could also bring the problem of shadow during silhouette segmentation.

The above challenges are due to both external and internal factors. Although some of the external factors (carrying or not, lighting changes, etc.) have been addressed, the effects of internal factors are not investigated yet.

1.3.2.3 Conclusion

Gait-based methods offer possibilities to get good results in human recognition at a distance and are more robust to illumination and clothe changes. Similarly to face-based methods, gait recognition systems have numerous disadvantages, especially to be sensitive to multiple persons crossing their paths in one sequence, walk speed, view angle, etc. As a consequence, most of the works on gait recognition focus on human motion analysis or activity recognition.

1.3.3 Appearance-based human recognition

1.3.3.1 Literature review

Biometric features (such as face, fingerprint and iris) allow a very high recognition rate as they are very specific for a person. However, some strong restrictions limit their applications. Fingerprint and iris features are impossible to use in video sequences. Face recognition is highly affected by image resolution and face views, as well as gait recognition has restrictions in application when we can find multiple persons in one image. Non-biometric features (appearance) could be used for human recognition in video surveillance with weaker restrictions on the acquisition modality. Appearance is defined by the person's visible clothes and body parts, which can be easily obtained after background subtraction. For a short period of time, the appearance of a person is expected to be invariant.

Generally, in an appearance-based human recognition system, image sequences are captured by a camera or a group of cameras, persons are segmented from images after background subtraction, and appearance features are extracted to represent them, which mainly include color, texture and shape. At last, classifiers are applied for classification. Figure 4.16 shows the general outline of appearance-based recognition.

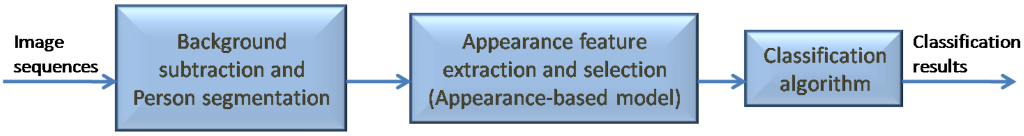


Image sequences
Background subtraction and Person segmentation
Appearance feature extraction and selection (Appearance-based model)
Classification algorithm
Classification results

Figure 1.2: General outline of appearance-based human recognition system

A significant amount of work has been performed on appearance-based human recognition, especially on person re-identification [TCAKD09, TCKA+10, TCA10, FBP$^+$10]. Generally, appearance-based person recognition techniques are organized in two main categories: single-shot [GT08, LD08, SD09] and multiple-shot [CMBT03, GSH06]. The first considers the appearance information of a person from one image only, the second integrates such information from multiple images of the same person. Person's appearance is usually represented by features (color, texture, etc.) or appearance models. As presented in [LD08], appearance representation could be categorized into two groups: holistic methods and part-based methods.

- **Holistic methods:** Holistic representation methods extract appearance information from the whole body. Histograms-based models (e.g. color and structural histograms) are usually used in holistic representation. In [HKK04], appearance is represented by color histogram and texture features of full body. Previous results proved that color features perform better than texture and the combination of both obviously improve the recognition rate. Similarly, Lin and Davis [LD08] extract features based on color-position histogram (color distribution in the silhouette) for multi-category human recognition. Besides, low-level features are exploited for holistic representation. In [GT08], Gray and Tao propose a viewpoint invariant pedestrian recognition system, in which AdaBoost is used to learn appearance model composed by simple localized features (image position and intensity).

- **Part-based methods:** These methods describe appearance from one or several parts of the body. Interest point operators and model fitting are often used. In [PJK$^+$06], the detected person is segmented into three parts by height ratio and features are extracted by the combination of color and height information of each part. Similarly, Hamoudi et al. [HBL10] process human body in three different parts, extract color (average values and histograms) and texture features (cooccurrence matrix-based texture features) for each part, then reform one-class SVM classification for recognition. In [FBP$^+$10], Bazzani et al. propose a robust appearance model for person re-identification. Three parts of body (head, torso and legs) are divided based on body symmetry axis and color histograms are computed for each part. The model works well both in the single-shot and the multi-shot modality and deal with the problem of a large and variable number of persons.

1.3.3.2 Difficulties of appearance-based recognition

Generally, there are some difficulties concerning appearance recognition:

- **Inner disadvantage of appearance:** Appearance features are useful only on a short period of time. Since the main features of appearance are based on person's clothes, it is a big challenge to classify persons based on appearance when they change clothes over a long period. Besides, it is difficult to recognize persons when they wear very similar clothes at the same time.

- **Changes of environmental conditions:** The changes of illumination could lead to the problem of shadow, which could increase difficulties on human segmentation from background. Appearance recognition is sensitive to bad segmentation. Besides, background is complex and dynamic in real-life settings, how to effectively detect and recognize persons is still an open challenge.

- **Choice of features:** The effectiveness of features extracted to represent persons has a direct effect on classification performance. As a result, what kind of features should be used and how to make them robust to the changes of illumination should be investigated.

1.3.3.3 Conclusion

Appearances, as non-biometric features, are easy to obtain in video sequence, since they do not require complex computation or seeking for specific parts of the body. Appearance recognition has less requirements concerning camera resolution, position of the person, walking speed, etc. As a consequence, appearance-based recognition methods are widely used in video surveillance, especially in human re-identification. However, automatic human recognition based on appearance in real-world application is still a challenge due to the inner disadvantages of appearance and the complicated environmental influences.

1.3.4 Discussion on representation methods

Biometric features (face, gait) and non-biometric features (appearance) are considered for human recognition. The inner disadvantages of each kind of these three human representations restrict their applications. The objective of this work is to automatically recognize persons or groups of person (with some specific characteristics) from video sequences created in real-world settings. In this online application, human recognition could be quite complicated due to the fact that person pose, background and illumination are variable. There could be one or several persons in one video sequence, furthermore, poses of persons are random (face/back related to the camera or with an angle to the camera). As we know, face-based methods rely on frontal view of face and high camera resolution, and gait-based recognition approaches are sensitive to the changes in walking speed, walking surface condition, etc. As a consequence, appearance-based methods are chosen in this work.

Next section will mention some approaches of appearance-based recognition in different applications, such as person re-identification in a network containing multiple cameras, person recognition over different days, etc. These articles have solved some problems of appearance-based human recognition with their proposed approach.

1.4 State of the art of appearance-based recognition

Appearance-based human recognition have been widely applied. In some applications, appearance-based human recognition systems are easier and more efficient than the ones based on gait and face features. However, in real-life applications, as the recognition environment could be variable and complicated, there are also some limitations of appearance-based system, such as clothe changes, overlap of viewpoint, carrying of goods, etc. Lots of researches focus on solving these problems, not only for improving the performance of classification, but also for finding solutions to reduce restrictions of the recognition methods.

In the next subsections, several appearance-based human recognition systems will be briefly presented in terms of features extraction, classification and analyses of experimental results. Firstly, human re-identification through a video camera network is briefly reviewed. Secondly, four articles that describe human recognition systems in different conditions (multiple cameras, changes of clothes or illumination conditions, etc.) are presented extensively. They have proposed efficient feature descriptors and classifiers in their own applications. All of these results are instructive and meaningful in the objective to design an online multi-category human recognition system.

1.4.1 Human re-identification system: Person detection, tracking and re-identification

The basic task of human re-identification system is to determine whether a given person has already been observed over a network of cameras [BCBT10]. The system generally consists of person detection, tracking and re-identification. Human detection and tracking are fundamental and significant for human identification. In [MLS12], Meden et al. present a pedestrian tracking system, in which a markovian tracking-by-detection framework is set up to monitor Multi-Object Tracking (MOT) from Non Overlapping Fields Of View (NOFOV) networks. Here, person re-identification is embedded in this tracking framework. There are two levels of tracking: local, at the camera level and global, at the network level. The measure between tracker and detection contains not only Euclidean distance evaluated by a Gaussian kernel, but also includes appearance model and identities distribution.

The distributed mixed-state filter combined with topology knowledge is used for the local identification. The experimental results on both synthetic and real data show that the proposed tracking-by-reidentification approach can achieve person re-identification in NOFOV camera networks.

In the thesis of Bak [Bąk12], six appearance-based techniques are proposed and compared for human re-identification. Firstly, the detection of the body parts is based on features learnt using an Histogram of Oriented Gradient (HOG) or an asymmetry characteristic of the body. Single-shot approaches extract person characteristics from only a single image, two methods are proposed: Dominant color descriptor (DCD) and spatial Covariance regions (SCR), which are both highly dependent on the accuracy of the body detection. Then four multiple-shot approaches (which use multiple images of the same person as training data to obtain a more reliable representation) are employed: Haar-like features, Mean Riemannian covariance (MRC), Mean Riemannian Covariance Grid (MRCG), COrrelation-based Selection of Covariance MATrIces (COSMATI). In the former two methods, one-against-one learning scheme is used to extract a discriminative representation of the person appearance and boosting approach is used for learning. The latter two methods are more efficient with a lower computational coat. Especially, COSMATI approach is faster than MRCG. However, it requires a step of offline learning to obtain the distinctive representation. The experimental results show that the covariance feature is the best comparing with other representation approaches, which is more invariant to the changes of rotation and illumination. However, they do not perform so well for low resolution images.

Based on [Bąk12], Souded [Sou13] also presents a person re-identification system, in which many methods are improved. The proposed improvements are instructive for this thesis work. For instance, normalized color attributes before feature extraction makes the algorithm more robust to the changes of illumination. The covariance descriptors based on texture and color features instead of single color information could improve the discrimination power. Visible side signature can be used for person classification to offer more reliable information with easy and fast computation. An adaptive weight is employed for each person, which gives better re-identification performance. However, person detection and recognition are still difficult when the size of the person is small comparing to the size of the image. Besides, in order to get better performance, more images for each person are needed for training, which will inevitably extend computing time. It is not efficient for online application when either the number of person or the number of camera is large.

1.4.2 Automatic human recognition based on gradient and color features

Horster et al. [HLL07] propose an automatic human recognition system based on appearance. After capturing the video data, foreground-background segmentation is applied to separate the moving person from the background. Gradient and color features are extracted to represent the segmented persons. Then, a statistical model (a bag-of-visual words) for each person is built and persons are recognized using a K- Nearest Neighbour (K-NN) classifier.

1.4.2.1 Appearance model

Two sets of local features are proposed to describe the person's appearance.

- **Gradient features:** Scale-invariant feature transform (SIFT) features [Low04] are computed from histograms of gradients and compose a 128 dimensional matrix. SIFT features are invariant to variation of orientation and also partly invariant to perspective transformation and illumination.

- **Color features:** Both local RGB and HS (H and S channel of HSV color space) color histograms are computed and normalized. All three channels in RGB color space are combined into one histogram and HS histogram is built and added. The number of RGB and HS features dimensions is 126 and 84, respectively.

A bag-of-visual-words model is used in this article by considering person features as words. k-mean clustering algorithm is used to obtain k clusters of visual words for each type of features separately. Codewords are then defined as the centres of the learned clusters. Based on a bag-of-visual-words model, each person is represented by N model vectors stemmed from the N selected video frames of the person.

1.4.2.2 Classification method

K-Nearest Neighbour method is used for person classification. The distances between the vector representing the test person and model vectors are computed by normalized histogram intersection I. $I(H_t, H_m)$ is the normalized histogram intersection between the model vector H_m and the test vector H_t, which is computed by:

$$I(H_t, H_m) = \frac{\sum_{j=1}^{M} \min(H_t^j, H_m^j)}{\sum_{j=1}^{M} H_m^j}$$

where M is the codebook size.

1.4.2.3 Experimental results

The database contains four persons, as shown in Figure 1.3, captured at different time of the same day and during different days with various clothes, light, etc. However there are similar appearances between the training and test images for each person. The comparative results with different conditions (number of person, number of frames, feature type, etc.) are shown in Table 1.1, in which between two and four persons are used in five experiments. Considering the second line of this table for example, it means that two persons are used in this experiment, 100 clusters are set in this bag-of-visual-words model, there are two video sequences for each person and each video has 800 frames, only RGB color histograms features are used and recognition rate is 75.71%. The recognition rates show that local RGB and HS histograms perform similarly and are better than SIFT features. Furthermore, the recognition rate can be improved. Experimental results show that the recognition rate can increase from 90.00% to 95.71% using $K = 5$ instead of $K = 1$ in another experiment.

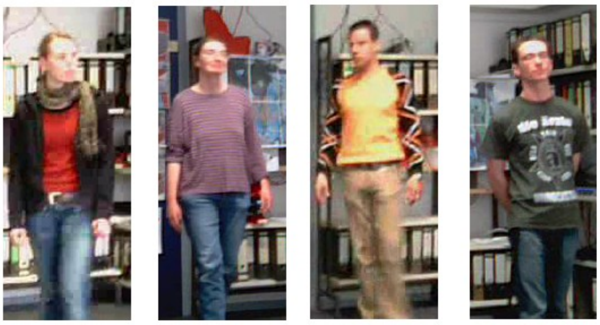


Figure 1.3: The four persons used for the experiments [HLL07]

N. Persons	N. Clusters	N. Videos/person	Frames/video	Feature type	Recognition rate
2	100	2	800	RGB	75.71%
2	150	2	300	HS	70.00%
4	250	4	150	HS	67.14%
3	250	2	300	HS	77.14%
2	100	2	300	SIFT	61.42%

Table 1.1: Recognition results for five experiments with different parameters [HLL07]

1.4.2.4 Discussion

In this system, color and gradient features of appearance are extracted to represent persons and k-NN method is used to classify persons based on normalized intersection of histograms. The experimental results show that the proposed system is usable. However, the recognition rates in Table 1.1 are below 80%, some methods could be considered to improve the performance of classification, for example extract more complete features (adding texture features) or consider less simple classifiers (e.g. SVM). Besides, in this article, only four persons are chosen in the experiments. In this person recognition system, there is almost no changes of illumination and location in the images used for learning and testing.

1.4.3 Human re-identification in complex environments

Truong Cong et al. [TCKA+10] present a system for reidentifying moving persons observed by multiple non-overlapping cameras in different locations. Color-position histograms are proposed as features to characterize person silhouettes in images. Then spectral analysis is used to reduce the number of features and SVM is chosen for classification.

1.4.3.1 Feature extraction and dimensionality reduction

A feature called color-position histogram is proposed, which combine both color and spatial information. To compute color-position histogram, person silhouette is first vertically divided into n equidistant parts, shown in Figure 1.4, then the average color of each part is computed and considered as feature vectors. The color-position histogram is composed of $n \times 3$ values (R, G and B, the three color channels). As presented in this article, color-position histogram is fast, uses less memory than classical color histogram, and can offer more reliable measures to separate different silhouettes.

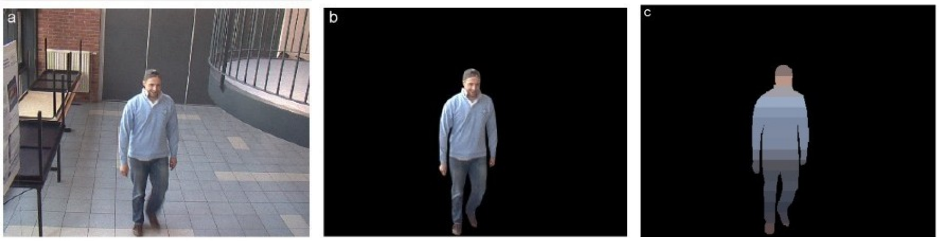


Figure 1.4: Color-position histogram: (a) original image; (b) localization of the silhouette; (c) color distribution in the silhouette [TCKA+10]

In order to get invariant color features, normalization is necessary. A color (R,G,B) is normalized as (r,g,b), where R, G, B is intensity of Red, Green and Blue respectively. Five normalization methods are compared in this article: (1)

chromaticity space: $r = \frac{R}{R+G+B}$, (idem for G and B). (2) Grayworld normalization: $r = \frac{R}{mean(R)}$, (same for G and B). (3) Affine normalization: $r = \frac{R-mean(R)}{std(R)}$, (same for G and B). (4) Comprehensive normalization [FSC98]: it is an iterative algorithm working in two stages to be invariant to intensity and color changes. (5) RGB-rank [FHST05]: the rank measure for each level in three channels are computed. Color normalization is done for each person's silhouette before calculating color-position histogram. The results of comparison of RGB space and five other normalization methods are shown in Table 1.2. The Grayworld normalization, RGB-rank and affine normalization achieve better performances than original RBG color. Grayworld normalization is the best one.

Spectral analysis [VLBB08] is used to reduce dimensionality representation space, which is obtained by considering the n lowest eigenvectors. In this work, only two significant eigenvalues are used to replace the initial $n \times 3$ features (n=8). The image is, as a consequence, projected to a 2D space based on spectral analysis.

1.4.3.2 Experimental results

In order to test the performances of spectral analysis, three tests have been done. The first test dataset is composed of two sequences of two persons differently dressed. The second contains two sequences of two people very similarly dressed. Finally, the last test consists of two sequences of the same person wearing the same clothes captured in different locations (indoors and outdoors). In the experimental results, the sequences of these three tests are respectively well-separated, less easily separated and strongly overlap. Since the recognition is based on person's color appearance, these results are satisfying and show the usefulness of both color-position histogram feature and spectral analysis [TCKA+10].

SVMs are then used for persons re-identification. A real-world database is created, named INRETS database, in which 40 persons are collected in two different locations by two cameras: indoors (in a hall with windows) and outdoors (with natural light). The experimental results show that the proposed method provides reliable results with true reidentification of 95%, as described in Table 1.2, where TRR stands for True Re-identification Rate.

1.4.3.3 Discussion

The proposed system can track moving persons observed by multiple cameras in different locations. The used features are based on color-position histogram with illumination invariant methods. The experiments have proved that normalized features can improve recognition performances. SVM classifier combined with spectral analysis gave satisfying classification results. In order to improve this work, more temporal and spatial features or biometrics features (face and gait)

	TRR at the optimal point(%)
RGB	86
Chromaticity space	65
Greyworld normalization	95
Comprehensive normalization	80
RGB-rank	88
Affine normalization	91

Table 1.2: Results of TRRs of RGB space and five normalization methods [TCKA$^+$10]

could be considered. Furthermore, this work has not given the solution to more challenging problems, such as to identify different persons who wear very similar clothes.

1.4.4 Person and pose classification in the same and in a new day

A real-time person recognition system is described by Nakajima et al. in [CMBT03], which aims to recognize not only persons but also their poses in images collected indoors. In this work, person image sequences are acquired by a fixed camera located in their lab, where illumination and background remain almost unchanged. However, the poses of person are unconstrained (face, back or side to the camera) and the image resolution is not very high. The system includes two main steps: image pre-processing (which consists of person detection and feature extraction) and person/pose classification.

1.4.4.1 Feature extraction

After background subtraction and edge detection, persons are segmented from the background. Then, color and shape features are considered to represent them. Four feature sets are obtained:

- (1) RGB color histogram: Each color histogram is calculated with 32 bins. 96 (32×3) features are extracted from one image;

- (2) Normalized color histograms: Color channel R and G are normalized as $r = R/(R+G+B)$, $g = R/(R+G+B)$. 32 bins are also chosen for each channel and 1024 (32×32) features are extracted from each image;

- (3) RGB color histogram and shape histogram: Shape histogram is calculated by counting pixels along rows (30 bins) and columns (10 bins) of each image. Then 136 features are obtained, 96 (32×3) for color histograms and 40 ($10+30$) for shape histograms;

- (4) Local shape features (75 features): Local shape features are computed by convolving 25 local shape patterns, which are introduced in [KHM97]. These features are extracted from three color channels: $R+G-B$, $R-G$ and $R+G$. 75 (25×3) features are considered.

1.4.4.2 Classification methods

In this work, two methods are used for person/pose classification: SVM and k-NN. Since the authors need to recognize multiple persons, three types of multi-category SVMs classifiers are used and compared. The first is one-against-all classifier and the other two are pairwise classifiers (Bottom-up decision tree and Top-down decision graph). Besides, k-NN classifier is also tested by choosing different values of k ($k=1$, $k=3$ and $k=5$).

1.4.4.3 Experimental results

In their experiments, images are captured by a color camera at a fixed location. Background and lighting are almost invariant. The first experiment aims to recognize four persons and estimate their poses (front, back, left and right). The example images are shown in Figure 1.5, 640 images (160 images for each person, 40 images for each pose) and 418 images have been respectively used for training and testing. Five multi-category classifiers are used, one is trained to identify persons and the other four are trained for each person separately and determine the pose of the person who is recognized by the first classifier. The results of the recognition and pose estimation show that the three types of SVMs have similar recognition rates, which are slightly better than k-NN classifiers. The performances of person recognition are higher than the one of pose recognition. Pose estimation is more difficult because of the similarity between right or left poses and front or back poses. In the second experiment, 1127 images of eight persons collected during 16 days are used. Five different sets of experiments (about 90%, 80%, 50% and 20% images for training) are performed and their recognition rates are shown in Table 1.3. The recognition rates reach more than 88% and are satisfying when the training and testing images are captured in the same day (first four experiments). However recognition rate decreases to about 50% when the test images are collected in a new day (last experiment). That is explained by the fact that person may change clothes everyday.

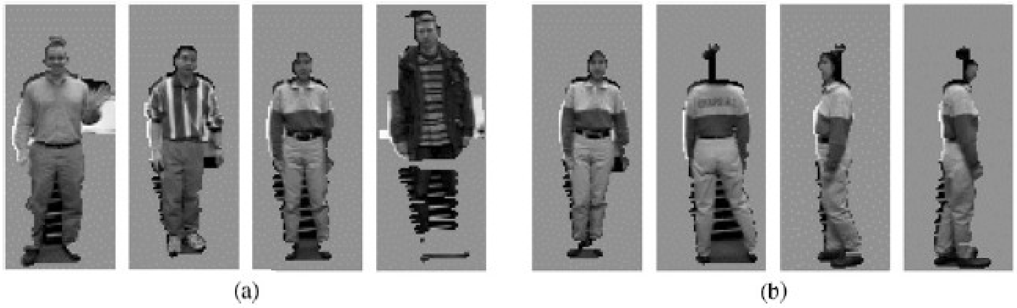


Figure 1.5: (a) Examples of the four people in the frontal pose; (b) Examples of the four poses for one person [CMBT03]

	(test:training)	1:9 (113:1014)	1:5 (188:939)	1:1 (564:563)	5:1 (939:188)	New day (122:1005)
	Top-down	92.3	91.2	90.5	73.3	45.9
	Bottom-up	90.6	91.7	90.6	66.1	45.9
SVM	One-vs-all	87.2	90.6	85.9	84.6	49.2
	One-vs-all (Polynomial)	**98.3**	**96.4**	**94.7**	**88.1**	52.9
	$k = 1$	92.9	92.0	92.7	85.1	**53.3**
KNN	$k = 3$	92.9	92.0	92.2	81.3	50.0
	$k = 5$	94.7	91.0	90.1	76.0	50.8

Table 1.3: People recognition rates of different classifiers for eight persons based on the normalized color features [CMBT03]

1.4.4.4 Discussion

In this system, images are represented by color and shape features. The experiments show that the feature set of normalized color histograms is better than the others. SVMs and k-NN classifiers are compared, and the results show that polynomial one-vs-all SVM achieves the best recognition performance.

This work can deal with multiple person classification, but it does not solve the problem of new person discovery. More invariant features could be considered to extend the capabilities of this system. Besides, this proposed recognition system does not consider adaptive learning and can not update with new input data, which is not suitable for many real-life applications.

1.4.5 Person recognition using Partial Least Squares (PLS) models

Schwartz and Davis [SD09] describe a framework to build discriminative appearance-based models using a statistical method called Partial Least Squares (PLS) [GK86].

A rich set of features based on color, texture and edge are extracted. As a result, a high dimensional feature set is obtained. However, high dimensional data may cause high computational time and low performance for classification. PLS method can model relations between observed features sets by means of latent features. In this work, PLS analysis is used to predict the class labels while reducing feature dimensions.

1.4.5.1 Feature extraction

For each person appearance represented by an image as shown in the first row of Figure 1.6, a high dimensional feature vector is computed. This vector consists of color features, texture features and edge features.

Color feature: Normalized color histograms are used to incorporate color information.

Texture feature: Texture features are extracted from co-occurrence matrices. Twelve descriptors of Haralick features [HSD73] are chosen.

Edge feature: Edge features are calculated by histograms of oriented gradients (HOG) [DT05].

1.4.5.2 Dimension reduction and classification

PLS model computes relations between observed variables depending on latent variables [Ros03]. In this work, PLS analysis is based on the classical Non-linear Iterative Partial Least Squares algorithm (NIPALS) [GK86]. When a test sample is presented, PLS method is used to reduce dimension by projecting the sample's feature vector onto a set of orthogonal weight vectors (projection vectors) in the latent subspace. The desired weight vectors are extracted by maximizing the covariance between the feature vector of the sample and its class label.

In this work, the features in the locations containing discriminatory attributes are associated to higher weights by PLS method. Figure 1.6 shows spatial distribution of weights of eight persons taken in the ETHZ dataset based on PLS models. Red shows high weight and blue is low one.

As described in Figure 1.7, the PLS discrimination model is learnt by one-against-all method (the appearance of a person is considered as positive sample and the appearances of the others are considered as negative samples). The feature vectors of learning appearances are projected onto their weight vectors, which have lower dimensions than the original ones. When a testing sample is coming, the feature vector of this sample will be projected into each latent subspace estimated by each training appearance. In addition, Euclidean distance between this sample and each one of the training set is computed after projection. The testing sample belongs to the person with the smallest Euclidean distance.

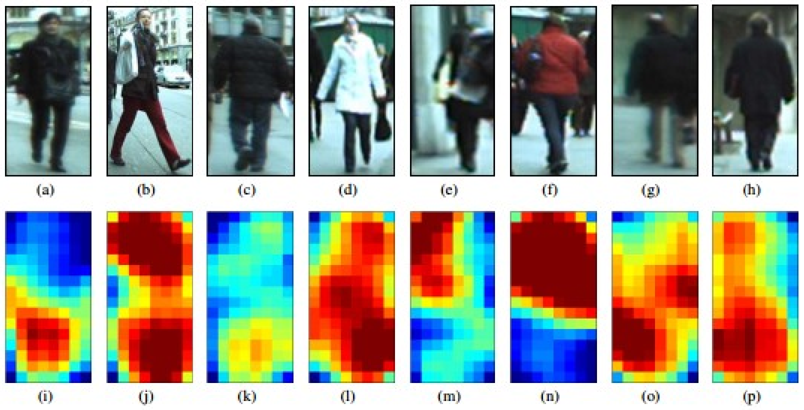


Figure 1.6: The images in the first row are appearances of eight persons in ETHZ Dataset and the second row shows the corresponding weights given by PLS models [SD09]

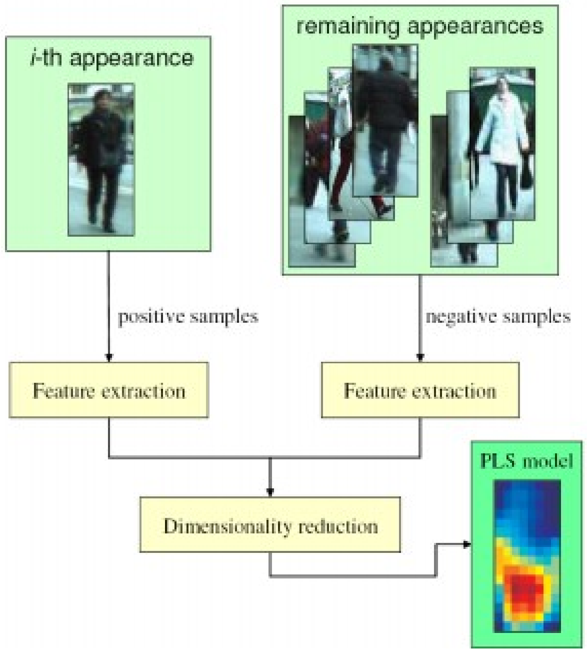
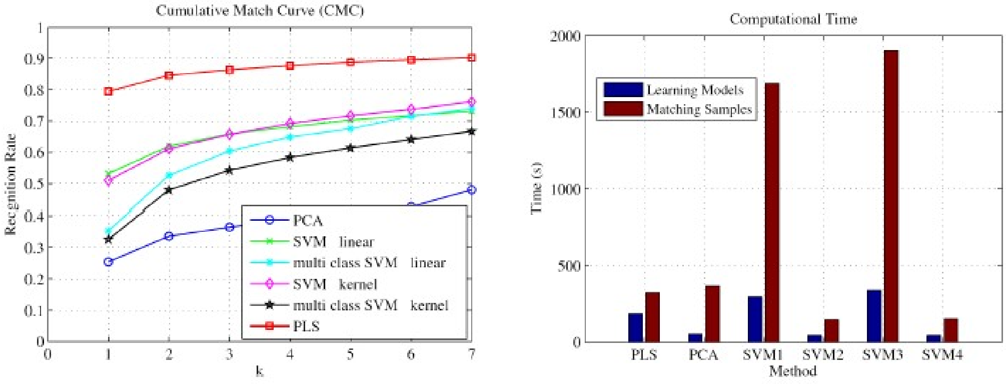


Figure 1.7: One-against-all strategy is used to learn a PLS discrimination model [SD09]

1.4.5.3 Experimental results

In ETHZ Database, three video sequences, captured by moving cameras, are used separately in three experiments. Sequence 1 has 1000 frames with 83 persons; Sequence 2 is composed of 451 frames with 35 persons; and Sequence 3 contains 354 frames with 28 persons. The comparable experiments are done with PCA and SVM. In the first comparison, dimensionality has been reduced by PCA and the classification method is the same as the one for PLS model. For the second one, data sets are classified directly in the original feature space. Four kinds of SVM are compared: linear one-against-all SVM (named as SVM1), linear multi-class SVM (SVM2), polynomial kernel one-against-all SVM (SVM3) and polynomial kernel multi-class SVM (SVM4).

The results on the three experiments show that the proposed PLS discriminative model generally gives higher recognition rate and costs less computational time. Taking the experiment for Sequence 1 as example, the results are shown in Figure 1.8, where CMC means the Cumulative Match Characteristic curve. Compared with PCA-based method, PLS model obtains a higher recognition rate and lower dimensional latent space. Furthermore, the results also show that PLS approach achieves better performance than SVM classifier when there are a few training samples and high dimensional features.



(a) Recognition rates for sequence #1

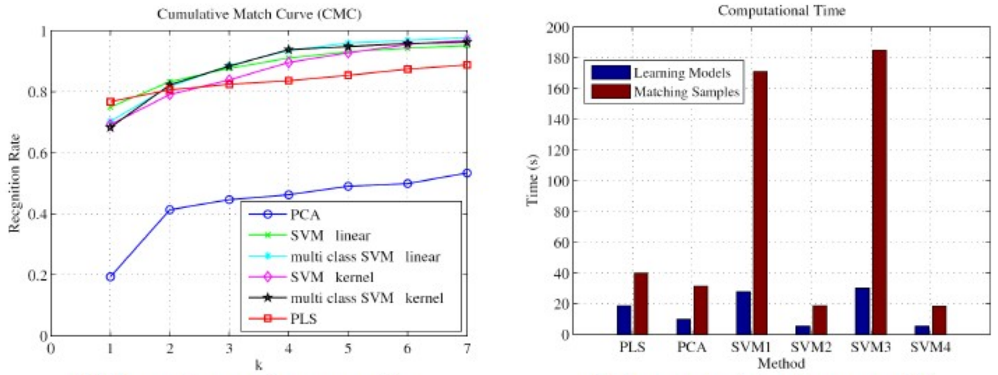
(b) Computational time for sequence #1

Figure 1.8: Recognition rate and computational time comparisons considering proposed PLS method for Sequence 1 [SD09]

1.4.5.4 Discussion

This work shows that a richer feature set can improve the performance of classification and the proposed PLS discriminative models give satisfying results in dimension reduction and classification (combined with one-against-all approach).

However, there are still some problems. As shown in Figure 1.9, PLS methods do not always obtain the highest recognition rate, SVM4 (polynomial kernel multi-class SVM) gives the best results on classification and computational time for Sequence 3. The authors have determined that the misclassifications are caused by the changes of clothes, illumination and occlusion. Besides, the experiment for Sequence 1 also show that the increase of the number of persons and the similarity in their appearance raise the difficulty of such system.



(e) Recognition rates for sequence #3

(f) Computational time for sequence #3

Figure 1.9: Recognition rate and computational time comparisons considering proposed PLS method for Sequence 3 [SD09]

1.5 Chapter summary

After an overview of general human recognition system, we have reviewed face-based, gait-based and appearance-based human recognition describing their main methods, difficulties and applications. Even if face features have the advantage of long period of validity, they have high restrictions on camera resolution and views of the face. Similarly, gait-based recognition systems are affected by numerous factors and can not handle the problem of multiple persons occurring in one video sequence. As a result, appearance-based recognition approaches seem more suitable in our study.

For appearance-based recognition, numerous researches have been developed to improve performances of classification and address the problems caused by changes on appearance or outer environment conditions. Features based on color, texture or shape are usually extracted to represent appearance, such as color histogram [HLL07], shape histogram [CMBT03], color/path-length features [YHD06], color-position histogram [TCKA+10], cooccurrence matrix-based texture features [HBL10], etc. The combination of two or several kinds of features could offer more information for appearance. [HKK04] prove that the combination of color and texture features improves classification performance. According to part-based methods, interest regions as head, shirt or legs are considered instead of the whole body. In [LD08, FBP$^+$10], human features is extracted from three parts (head, top and bottom). It could have better representation for a person with different color in shirt and trousers. These previous works are instructive on appearance recognition and will seem as a guideline for our work.

However, there are still some unsolved challenges for appearance-based recognition system in real-world application. The main limitation is caused by the disadvantages of appearance. The features of appearance are very different when persons change clothes and it is hard to distinguish persons when they wear very similarly colored clothes. Similarly to face and gait recognition, external environment influences (e.g., illumination, location) cause difficulties in appearance detection and classification. Besides, it is difficult to have a complete information representing a person. That is why extracting effective appearance features is very important in human recognition system. Furthermore, the challenge that consists in recognizing new persons has not been well addressed in existing researches. This work will continue research on appearance-based human recognition in order to achieve better performances both in offline and online applications.

In the following chapter, a new appearance-based human recognition system will be proposed and tested in an offline setting. The principle of appearance features extraction and selection will be described, and finally multi-category classification will be discussed.

Chapter 2

Offline human recognition system

2.1 Introduction

Chapter 1 presents a review of human recognition capacity in video surveillance system, in which a generic framework and main human representation methods are described. Besides, this has emphasized the need of specific features and efficient classifiers. In this chapter, an appearance-based recognition system will be described to classify multiple persons in offline application. Support vector machine (SVM) classifier has been chosen for this first part of multiple persons recognition in this thesis, because SVM is considered as one of the most powerful method for classification and performs well in lots of various applications [CV95, BL02].

Generally, when a video sequence is acquired in a recognition system, initial features are extracted to represent the person, then feature selection methods are implemented to choose a good feature subset by removing irrelevant and/or redundant features. A good subset is the one that includes features uncorrelated with each other (to obtain the least dimension possible) but highly correlated with the class [Hal99]. Many factors can affect the performances of human recognition system (and more generally of a classifier), and the effectiveness of selected features is one of them. Theoretically, a higher dimensional feature set which contains more information can lead to higher discrimination performance in classification. However, higher dimensional data easily raise the curse of dimensionality. Besides, learning algorithms in higher dimensional spaces inevitably have higher computational costs. As a consequence, many algorithms (either supervised or unsupervised) have been proposed to perform dimensionality reduction and feature selection. This will be discussed in this chapter.

It starts by reviewing SVM techniques ranging from binary SVM to several approaches of multi-category SVM. Then literature reviews on feature extraction and selection are presented and three methods of feature selection are also described. Finally, offline human recognition system based on multi-category SVM is tested on a publicly available database (CASIA), that will be introduced.

2.2 Support Vector Machines

Support Vector Machines (SVMs), introduced by Vapnik and his co-workers [Vap00, CV95, BGV92], are considered as one of the most important and powerful method in supervised classification. SVMs are efficient to handle large-scale classification problems and achieve great success in many applications, such as handwritten digit recognition, text categorization or face detection [CST00].

Originally, SVM is designed for binary classification by finding an hyperplane with a maximum margin between two classes. Then, binary SVM has been extended to solve multi-category problems. A brief review of binary SVM and several methods of multi-category SVM will be presented in the next sections.

2.2.1 Binary SVM

Binary SVM is initially designed as a function classifying two sets of linearly distinguishable data. Let's consider a set T of n pairs (x_i, y_i), where $x_i \in R^d$, $i = 1, 2, \cdots, n$, $y_i \in \{1, 2, \cdots, K\}$, K is the number of classes and n is the number of samples. For binary SVM ($K = 2$), we consider $y_i \in \{-1, +1\}$. For linearly separable classes, there exists an infinite number of separating hyperplanes $w^T x_i + b = 0$, where $w \in R^d$ and $b \in R$. SVM specially search for the optimal hyperplane (with the biggest margin, that is to say the distance with the "closest" points of each class) from all hyperplanes satisfying $y_i(w^T x_i + b) \geq 1$ for $i = 1, 2, \ldots n$. It is equivalent to solve the following problem [Bur98]:

$$\min_{w,b} \frac{1}{2} w^T w \quad s.t. \quad y_i(w^T x_i + b) - 1 \geq 0, i = 1, 2, \ldots n \tag{2.1}$$

Equation 2.1 is a quadratic constrained optimization problem, which could be solved by its dual problem:

$$\begin{aligned} &\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j x_i^T x_j \\ &s.t., \sum_{i=0}^{n} \alpha_i y_i = 0, \alpha_i \geq 0, i = 1, 2, \ldots n \end{aligned} \tag{2.2}$$

where α_i are the Lagrange multiplies.

Afterwards, linear SVM has been extended to solve non-linear problem by mapping initial data x into a higher dimensional space based on a function Φ and finding a linear optimal hyperplane in this new space. We get $x \to \Phi(x)$ and $x_i^T x_j \to \Phi^T(x_i^T)\Phi(x_j)$. We consider a set of functions (kernels) that act as dot products (Mercer's conditions) and as a consequence $\Phi^T(x_i^T)\Phi(x_j)$ is replaced by $k(x_i,x_j)$. Therefore, conditions 2.2 can be generalized as:

$$\max_{\alpha_i} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \alpha_i\alpha_j y_i y_j k(x_i,x_j)$$
$$s.t., \sum_{i=0}^{n} \alpha_i y_i = 0, \ldots \alpha_i \geq 0, i = 1,2,\ldots n \tag{2.3}$$

There are three typical kernel functions, as shown in Table 2.1.

Kernel function	Inner product kernel $k(x,x_i)$, $i = 1,2,\cdots,n$
Polynomial kernel	$k(x,x_i) = (x^T x_i + 1)^d$
Gaussian (Radial-basis) kernel	$k(x,x_i) = exp(-\vert\vert x - x_i\vert\vert^2/2\sigma^2)$
Multi-layer perceptron (sigmoid)	$k(x,x_i) = tanh(\beta_0 x^T x_i + \beta_1)$, β_0 and β_1 are decided by the user

Table 2.1: Three typical kernel functions

However, due to the effects of noise, it is possible that no hyperplane can strictly divide the data. Soft margin SVM has been proposed to solve this problem by finding an error-tolerant margin. Positive slack variables ξ_i are introduced, which measure the degree of misclassification of the data x_i. Equation 2.1 is then generalized as:

$$\min_{w,b,\xi_i} \frac{1}{2} w^T w + C\sum_{i=1}^{n} \xi_i \quad s.t., w^T \Phi(x_i) + b) \geq 1 - \xi_i, i = 1,2,\ldots n \tag{2.4}$$

where $C \geq 0$. C is the trade-off between margin maximization and a low number of errors in the training data. When $C = 0$, no training error is allowed. On the contrary, when $C = \infty$, the classifier has very soft margin.

2.2.2 Multi-class SVM

There are two main categories of multi-category SVM. One is constructed by combining several binary classifiers, e.g., one-against-one SVM and one-against-rest SVM. The other has only one classifier, in which all data are treated by one optimization formulation, as all-against-all method [HL02].

2.2.2.1 One-against-rest SVM

One-against-rest method is probably the earliest approach implemented for multi-class classification. Let's consider K-class classification ($K > 2$) with a set $C = \{C_1, C_2, \cdots, C_K\}$ of classes, some training data $(x_1, y_1), \cdots, (x_j, y_j), \cdots, (x_n, y_n)$, where $x_j \in R^d, j = 1, 2, \cdots, n$ and the associated classes $y_j \in \{1, \cdots, K\}$. One-against-rest constructs K binary SVM classifiers using $C_1 \cdot against \cdot C \backslash \{C_1\}$, $\cdots$, $C_i \cdot against \cdot C \backslash \{C_i\}, \cdots, C_K \cdot \cdot against \cdot C \backslash \{C_K\}$, in which the i^{th} SVM ($i = 1, \cdots, K$) needs to solve the following problem:

$$\begin{aligned} \min_{w_i, b_i, \xi_j^i} \quad & \frac{1}{2}||w_i||^2 + C \sum_{j=1}^{n} \xi_j^i \\ s.t. \quad w_i^T \Phi(x_j) + b_i \geq 1 - \xi_j^i, & \qquad \forall x_j \in C_i \\ w_i^T \Phi(x_j) + b_i \leq -1 + \xi_{ij}, & \qquad \forall x_j \in C \backslash \{C_i\} \\ \xi_j^i \geq 0, j = 1, \cdots, n & \end{aligned} \tag{2.5}$$

For a testing sample x, $f_i(x) = w_i \cdot x + b_i$ can be obtained by SVM_i. The testing sample x belongs to j^{th}class where $j = \arg \max_{i=1,\ldots,n} f_i(x)$.

Figure 2.1 shows a simple example of one-against-rest SVM applied to three-class recognition. There are three classifiers: Class1 vs Class(2,3); Class2 vs Class(1,3) and Class3 vs Class(1,2). Such kind of multi-class SVM is easy to understand and to compute. However, it exists large areas of difficult decision (overlapping of all areas), which are colored on Figure 2.1. In these case, the usual is to consider the classifier that has the largest margin as the "safer" decision.

2.2.2.2 One-against-one SVM

The one-against-one method is introduced in [KPD+90], which constructs binary SVM classifiers for all pairs of classes, such as C_1-against-C_2, C_2-against-C_3, etc. For K-class problem, the total number of binary SVMs is $K(K-1)/2$. For the training data from the ith and jth classes, we get the following:

$$\begin{aligned} \min_{w_{ij}, b_{ij}, \xi_t^{ij}} \quad & \frac{1}{2}||w_{ij}||^2 + C \sum_{t=1}^{n^{ij}} \xi_t^{ij} \\ s.t. \quad w_{ij}^T \Phi(x_t) + b_{ij} \geq 1 - \xi_t^{ij}, & \qquad \forall x_t \in C_i \\ w_{ij}^T \Phi(x_t) + b_{ij} \leq -1 + \xi_t^{ij}, & \qquad \forall x_t \in C_j \\ \xi_t^{ij} \geq 0, t = 1, \cdots, n^{ij} & \end{aligned} \tag{2.6}$$

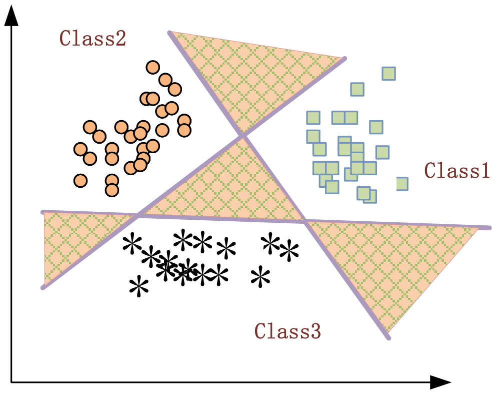


Figure 2.1: Diagram of one-against-rest SVM applied to a three-class classification problem

where n^{ij} is the number of data that belong to $C_i \cup C_j$.

An example of three-class recognition is shown on Figure 2.2. There are three classifiers: Class1 vs Class2; Class1 vs Class3 and Class2 vs Class3. Comparing Figure 2.1 and 2.2, the colored area in Figure 2.2 is largely smaller. It shows that one-against-one can achieve better performance when test data are in the confused area.

There are many function methods to decide the class of a sample from $K(K-1)/2$ classifiers. One common method is Max Wins Strategy [Fri96], which assigns a sample to the class that has the largest number of votes.

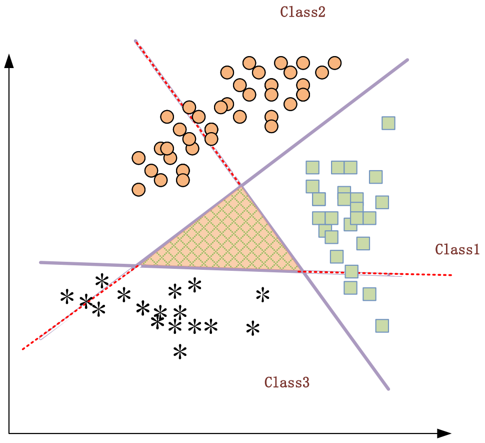


Figure 2.2: Diagram of one-against-one SVM applied to a three-class classification problem

2.2.2.3 All-against-all SVM

Similar to binary SVM, all-against-all method solves a K-class problem by addressing a single quadratic optimization problem of size $(K-1)n$. That is to say, it needs to find an optimal hyperplane for separating all classes [HL02]. In K-category classification, the margin between classes i and j is $2/||w_i - w_j||$. In order to get the largest margin between classes i and j, minimization of the sum of $||w_i - w_j||^2$ ($i = 1, \cdots, K-1$ and $j = i+1, \cdots, K$) is computed. Also, as described in [BB99], the regularization term $\frac{1}{2}\sum_{i=1}^{K} ||w_i||^2$ is added to the objective function. In addition, a loss function $\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_t \in P_{ij}} \xi_t^{ij}$ is used for the non-separable case [BB99], where $C_{ij} = C_i \cup C_j$ and the slack variable ξ_t^{ij} measures the degree of misclassification of the t^{th} training vector, related to the hyperplane ij. Therefore, the results of K decision functions are based on the solution of the following formulation:

$$\begin{aligned} \min_{w_i,b_i,\xi_t^{ij}} \sum_{i=1}^{K}\sum_{j=i+1}^{K} \left\|w_i - w_j\right\|^2 + \frac{1}{2}\sum_{i=1}^{K} \|w_i\|^2 + C\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_t \in P_{ij}} \xi_t^{ij} \\ s.t., (w_i - w_j)^T\Phi(x_t) + (b_i - b_j) \geq 1 - \xi_t^{ij}, \forall x_t \in C_i \\ (w_i - w_j)^T\Phi(x_t) + (b_i - b_j) < -1 + \xi_t^{ij}, \forall x_t \in C_j \\ \xi_t^{ij} \geq 0; i = 1, \ldots K; j = i+1, \ldots K; t = 1, \cdots, n \end{aligned} \tag{2.7}$$

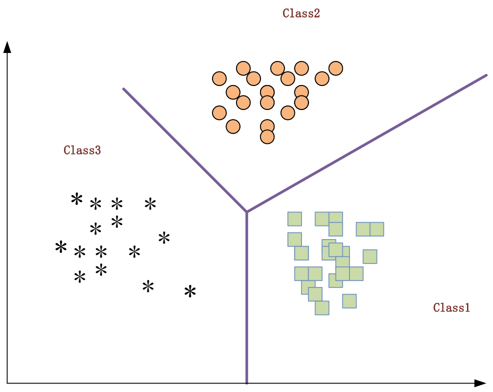


Figure 2.3: Diagram of all-against-all SVM applied to a three-class classification problem

Taken three-class recognition for example, as shown in Figure 2.3, which describes the classification boundaries and the biggest margins for the three classes.

Since only a single classifier is used to separate all classes, there is no confused area.

2.2.2.4 Discussion

Three kinds of multi-class SVM approaches have been presented. As far as the training cost is concerned, one-against-rest SVM is preferable, because it needs only K binary SVMs. However, it has a largest areas of difficult decision, as shown by the comparison of Figures 2.1, 2.2 and 2.3. Compared to one-against-all SVM, one-against-one SVM algorithm is more time-consuming as it needs to solve $K(K-1)/2$ binary SVMs. However, it generally performs better and is more suitable for practical use. Nevertheless, all these methods are usually applied in offline applications, in which test step can begin only when training step has finished and all parameters of SVMs have been chosen.

All-against-all SVM method consists of only one SVM solving a single optimization problem. If it finds the optimal hyperplane, there is no difficult decision area. The most important is that it could be used in online system by updating the decision functions parameters with the changes of data.

2.3 Feature extraction

In order to recognize people based on his/her appearances, we need to characterize appearance by using a set of features. Feature extraction corresponds to convert the information of an image carried by the pixels to a more repeatable and condensate form. Appropriate features can correctly represent one class (person) and easily differentiate it from the others. The most adapted method of feature extraction should be sufficiently discriminative, effective and with reasonable dimension for the feature vector. According to the location of the effective features for the image content characterization, the existing approaches are separated into two categories: global feature and local feature. The methods of the former are generally based on statistical analysis of the whole image (pixel by pixel). However, the methods of the latter compute features from a small neighbourhood with a predefined size and form around a particular area of the image. Since global features are computed from the analysis of the whole image instead of a more expensive local object detector, the computation is cheaper and faster.

According to the nature of the extracted appearance features, they are divided into three categories: color features, texture features and shape features.

- **Color features.** Color features are easy to obtain and compute. Besides, since they are based on the value of the pixels, they are less sensitive to translation, rotation and insensitive to scale after correct normalization. As described in numerous references [HKK04, HBL10, TCKAD10, HLL07], average value, variance value, histogram, color moments, etc., are widely used for person representation.

- **Texture features.** Texture features contain important information about the structural arrangement and describe the distinctive physical composition. The big varieties of texture on clothing could be used for appearance-based human recognition. Typically, the methods of texture feature extraction fall into two main categories: structural methods and statistical methods [KT10]. Structural methods analyse texture based on characteristic of structural primitives, such as gabor filters [GPK02] and wavelet transforms [MM96]. Statistical methods compute grey values at each point in the image and then get the statistics from the distribution of the local features. There are first-order and second-order statistics, such as grey-level co-occurrence matrix (GLCM) [Cla02] and Haralick texture features [HSD73].
- **Shape features.** Shape features are the surface configuration characteristics of the object, such as outline and contour. Shape representations can be divided into different ways: contour-based or region-based, space domain or transform domain, information preserving or non-information preserving [YKR08]. Many shape descriptors have been developed, such as Edge Histogram [PJW00], Histogram of Oriented Gradients [DT05] and Scale Invariant Feature Transform (SIFT) [Low04].

In order to extract a visual information that is as complete as possible, different features can be fused as an initial feature set for representing persons. As proved in [HKK04], the combination of color and texture features achieves better recognition results, compared to the use of only color features. Since histograms are spatial information, Truong et al. [TCA10] propose a feature called ”color-position” histogram used for people reidentification, which includes both color and spatial features. In this work, the initial feature set is composed by both color and texture features extracted from person appearances.

2.4 Feature selection

2.4.1 Introduction

With the emergence of big data and huge dataset, reducing data dimension is necessary and important in most of pattern recognition and machine learning problems. Generally, there are two main ways to reduce dimension, one is to transform to a new reduced feature set (feature transformation), the other is to select a subset of the initial space (feature selection).

Feature transformation methods transform the original feature set to a more compact one, while keeping as much information as possible. Principal Component Analysis (PCA) is one of the well-known unsupervised methods for feature transformation. PCA uses an orthogonal linear transformation to project the original samples into a lower dimensional space. Original features are converted to a set of linearly uncorrelated variables named as principle components. PCA has the ability to identify the most representative combination of features by the order of principle components (which show the largest directions of variance in the data). As a result, PCA reduces dimensionality of original features without losing the main information [Jol05]. In this work, PCA is chosen to reduce the dimension of the initial feature set.

Compared to feature transformation techniques, feature selection methods reduce dimension by selecting a good subset of the original feature set, which only contains relevant and discriminative features. Generally, there are four basic parameterization in a typical feature selection method, as shown in Figure 2.4 and described as follows:

- **Starting point.** In the beginning of the feature selection process, a starting point (a special feature set) is chosen as the origin of the search, which may affect the result. Generally, starting point can be chosen in the following three options: (1) to begin with no features and iteratively add features; (2) to begin with all features and successively remove features; (3) to begin with a random set of features and move outwards from this point.

- **Search strategy.** Search strategy is one of the most predominant aspect of feature selection. It influences not only on the computational cost but also on the result of the selection. Section 2.4.2 discusses some common search approaches.

- **Evaluation strategy.** How to evaluate the goodness of the selected feature subset is another important point. Generally, according to the evaluation criterion and the dependency to the classifier, feature selection methods fall into two main categories: filter methods and wrapper methods, which are described in detail in Section 2.4.3.

- **Stopping criterion.** A feature selector has to decide when to stop the searching procedure. Both search procedure and evaluation strategy can influence the choice of a stopping criterion.

In recent researches, search procedure and evaluation strategy are the two main aspects. In the next section, these two aspects will be described in detail.

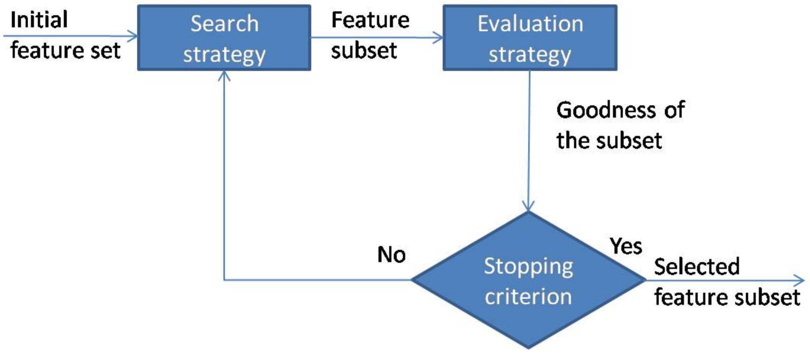


Figure 2.4: Feature selection process

2.4.2 Search strategy

The aim is to find the best feature subset according to the evaluation. If the initial feature set contains N features, it exists 2^N competing candidates. Exhaustive search is a solution to find the optimal subset, which evaluates all the candidate to decide. However it requires lots of computation, especially when the number of initial features is large (and more generally, it is not feasible in a reasonable amount of time). Some other methods without exhaustive search also have the ability to find the optimal subset, such as Brand and Bound (BB) algorithm. However, the applications of BB algorithm are limited, because it requires that the evaluation criterion is monotonic in the whole search process (which is highly rare in real applications and very difficult to know in advance).

Since these optimal search algorithms have some constraints on computation, numerous sub-optimal search techniques are considered. Heuristic-based search is only quadratic in terms of the number of features, which is simple and fast to implement. An heuristic-based search method may not always find the best feature subset, but it is guaranteed to find an optimal result in reasonable time. Typically, three directions of searching are categorized: forward selection (only adds feature to the subset); backward elimination (only deletions); and stepwise bi-directional search (both additions and deletions).

Hill-climbing search (also called greedy search or steepest ascent) is one of the simplest local search methods. It moves from the current node to the other nodes with the highest evaluation and terminate when no node improves the evaluation over the current node [KJ97]. One common disadvantage of hill-climbing search is that it is easy to terminate on a local maximum. Best first search algorithm is similar with greedy hill climbing search, but it is more robust as it allows backtracking in the search process. During the best first search procedure, the most promising node from the current feature subset is chosen. If the path given by a promising node is suddenly less promising, it can go back to choose another promising node and continue the search path from there. A stopping criterion is usually given in practice [Hal99]. In this work, we define that the search will backtrack when 5 nodes do not significantly improve the merit. In [KJ97], the experiments showed that, obviously, best-first search achieved better performance than hill-climbing both on artificial and real datasets.

2.4.3 Evaluation strategy

There are various approaches to evaluate the goodness of a selected feature subset. Different evaluation strategies could give different optimal subsets. Dash and Liu [DL97] divided the evaluation strategy based on five criteria: Distance Measure, Information Measure, Dependence Measure, Consistency Measure and Classifier Error Rate Measure. Various methods of feature selection are composed of different combinations of generation procedures and evaluation strategies. They presented a total of 15 combinations of 3 types of generation procedures and 5 types of evaluation criteria.

In this work, feature selection is an important and necessary process, which highly affects the performance of classification. We will describe two kinds of feature selection methods on the basis of two categories: one is on the intrinsic characteristics of attributes (filter), the other depends on the classification algorithm (wrapper).

2.4.3.1 Feature selection based on the intrinsic characteristics of attributes: filter

Filter methods are considered as the earliest method of feature selection in machine learning. They consider only the intrinsic characteristics of the data without considering any classifier [Hal99]. Figure 2.5 illustrates the selection procedure of filter method. Typically, a feature relevance score is computed to show the goodness of the subset, only high scoring features are kept and undesirable features are filtered out before the validation of the classifier. Then the selected feature subset is used as input of classifier. That is to say, filter methods are performed only once before classification.

Filter technique begins with univariate, such as Chi-Square [BS87] and Information Gain [YL03], which is fast and efficient but without considering feature-feature interactions [SIL07]. Then it extends to multivariate as CFS [Hal99] , Relief algorithm [AABC04] and Markov Blanket Filter [KS96]. Relief algorithm assigns a relevance weight to each feature and searches all the relevant features (both weak and strong), however this method is less efficient in deleting redundant features.

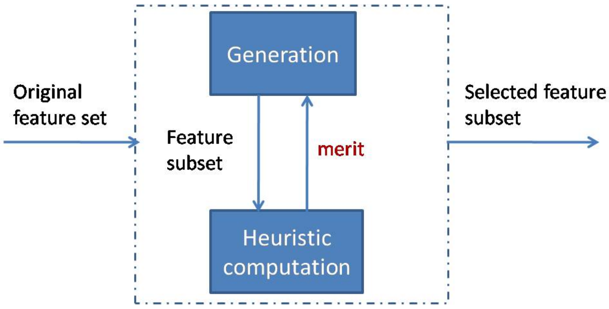


Figure 2.5: The filter method to feature subset selection

Correlation-based Feature Selection (CFS) approach is one of the filter methods, which ranks feature subsets according to the correlation based on the heuristic of 'merit' (goodness), which was described by M. A. Hall [Hal99] as the following expression:

$$M_s = \frac{k \cdot \overline{r_{cf}}}{\sqrt{k + k \cdot (k-1) \cdot \overline{r_{ff}}}}$$

where k is the number of features selected in the current subset, $\overline{r_{cf}}$ is the average of feature-class correlations, for each element $f \in S$ of our current subset, and $\overline{r_{ff}}$ is the average of feature-feature correlations for each pairwise of elements.

Numerous experiments in [Hal99] show that CFS is useful to remove the redundant and irrelevant features from learning data and can be used in common machine learning algorithms. In this work, this approach is employed to select the effective features from the numerous initial features. Best-first search is applied with CFS evaluation to search the set with the best value.

2.4.3.2 Feature selection based on the classification algorithm: wrapper

Wrapper methods have been initially described by John et al. [JKP94]. It selects feature subsets according to the performance of classification. The classification results are given, for instance, by correct classification rate or global error rate. Figure 2.6 illustrates the process of feature selection by wrapper methods. Wrapper techniques easily reach better results than filters due to the fact that they establish the special interaction between the training data and the learning algorithm. However, the execution time before obtaining the desired results could be very long. Indeed, the learning algorithm has to be repeated for both training and testing data in the whole selection period and need to re-run when the learning algorithm is changed.

Wrapper method uses a search algorithm to go through the whole combination of features. Different wrapper methods are given by various search strategies,

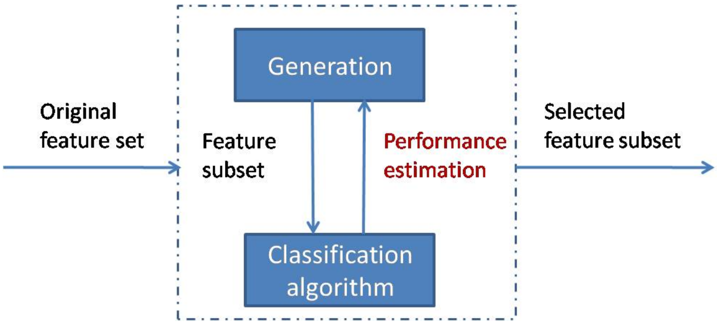


Figure 2.6: The wrapper method to feature subset selection

which include Sequential Forward Search (SFS) [PNK94], Sequential Backward Elimination (SBE) [PNK94], Genetic Algorithm [YH98], Estimation of Distribution Algorithm [IML$^+$01], etc. In this work, best first search algorithm is also chosen and accuracy is used to evaluate the goodness of a subset.

2.4.3.3 Discussion

The two approaches described above have their respective advantages and disadvantages. The CFS approaches have the main advantage of considering both feature-feature and feature-class correlations. The main advantage of wrapper methods is that they give better results, which is due to the fact that the classification algorithm is already specified and used to compute the merit [KJ97]. However, this specificity of wrapper methods also cause their drawbacks, as it provokes high risk of overfitting and makes it very time-consuming. By comparison, CFS methods have a lower computational complexity and are independent of classification procedure [SIL07].

2.5 Application of offline human recognition

2.5.1 Presentation of CASIA Database

CASIA Gait Database [CAS01] is a video database of 20 persons, walking at normal pace. Each person walks with six different orientations relatively to the video camera. Each image contains a unique person and the 20 persons do not change their clothes between trials. Six trials are presented for each of these 20 subjects: walking from right to left and from left to right, walking in front of the video camera (coming and leaving, two times in each direction), walking in a direction that is at 45° of the camera from the left and from the right. Figure 2.7 shows the 20 persons contained in the CASIA gait database. The whole number of images for the 20 persons (20 classes) is 19135 and the distribution of the samples in the different classes is given in Table 2.2. This table shows that the number of samples in each of the 20 classes is close (considering the average and standard deviation values). The dataset is, as a consequence, well balanced.

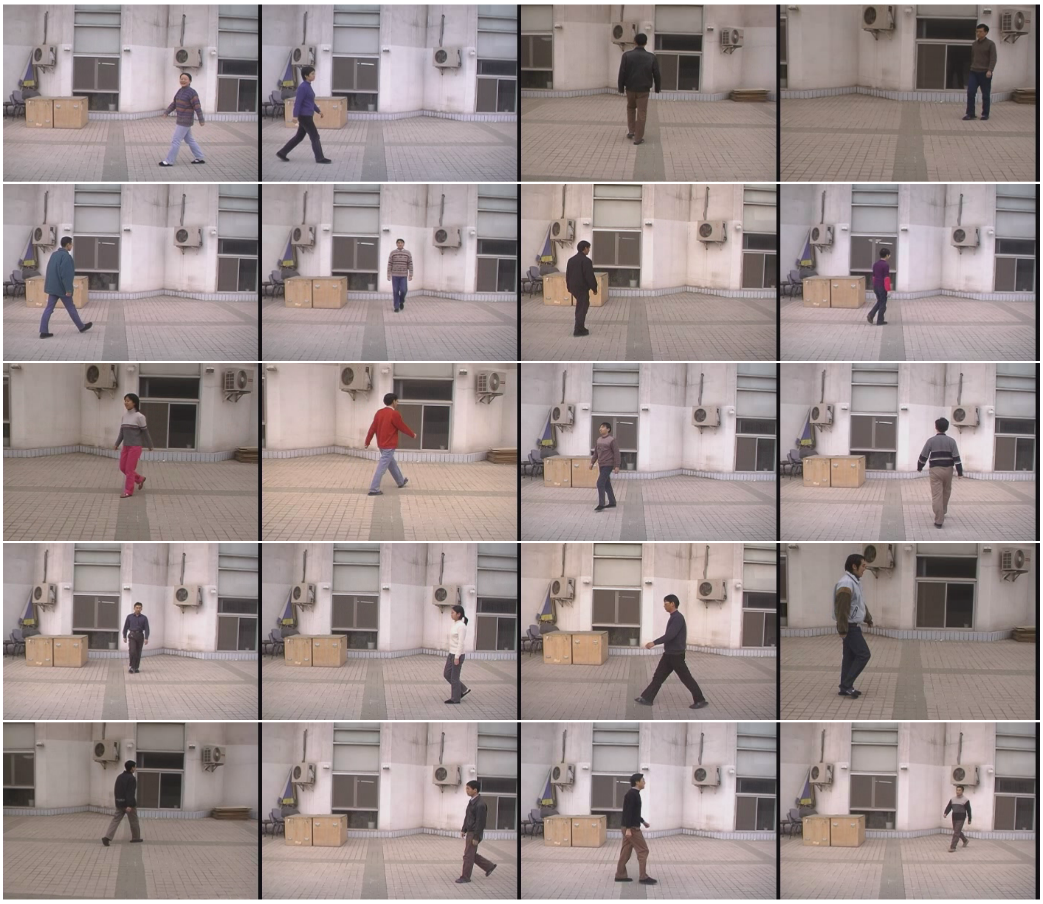


Figure 2.7: Twenty persons in the CASIA gait database [CAS01]

Background subtraction has been performed and the silhouette pictures have been extracted from the sequences. From the silhouette, we segment the body into three different parts based on height ratio: the head, the top part (the shoulder and chest) and the bottom part (the legs), which are shown in Figure 2.8. Such segmentation of the body into three parts has been chosen because these three parts are generally in different colors (from different clothes) and are considered as a natural way to represent a person [Ham11]. In this thesis, each part is processed separately and the corresponding features are computed independently. As a consequence, three different analyses are done, which are more accurate than considering the whole body as a unique part and averaging it. After extracting

Number of classes	20
Number of frames	19135
Average Cardinality	956.75
Std Cardinality	83.5243

Table 2.2: Distribution of the 20 classes within the 19135 images in CASIA gait database (Std cardinality is the standard derivation.)

the initial features of the 20 persons, we define different feature sets in order to investigate the comparison of computing time and classification results.

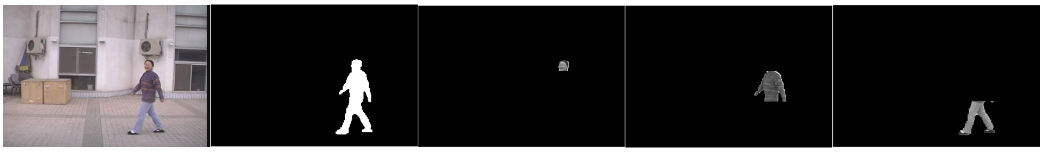


Figure 2.8: The silhouette picture and the three parts of the body

2.5.2 Extraction of initial features

In the experiments, a set of features composed of color and texture characteristics is used to define the 20 persons to recognize.

Color features with Red, Green and Blue (R, G and B) components of each frame are varying depending on several factors, such as illumination condition, surface reflectance and quality of the camera. In these conditions, normalization is necessary. In our work, the Grey-world Normalization (GN) has been used, which has been mentioned in the previous chapter and has shown good performance [TCKA$^+$10]. It gives normalized values R', G' and B' ($R' = \frac{R}{mean(R)}$, $G' = \frac{G}{mean(G)}$ and $B' = \frac{B}{mean(B)}$). It assumes that changes in the illuminating spectrum can be modelled by three constant factors applied to R, G and B, in order to obtain invariance and be more robust to illumination intensity variations [FSC98]. In [TCKA$^+$10], the experimental results have proven that the features with Grey-world Normalization obtained higher classification rate than the original data and the ones with other normalization methods.

As explained in the end of Section 2.5.1, we consider three different parts for each body. For each part, we compute the different color features, which consist of *Mean Value*, *Standard Deviation Value* for each color component and *Energy* in four beams of the histogram of the image. Take the *R* component of the head part as an example, the normalized value R'_H is defined as: $R'_H = \frac{R_H}{mean(R)}$. This leads to the extraction of 18 color-based features for each part of the body for three color components. In total, we have 54 color-based features for each person.

Haralick texture features are also used. They are based on the spatial grey level co-occurrence matrix of pixel values [HSD73]. These features describe the texture of the image by 13 statistics calculated from the co-occurrence matrix. In this work, after converting each body part (original image) into grey level, one matrix is obtained to represent this region, with values between 0 (black) and 255 (white). From the spatial grey level dependence matrix, 13 texture features of each part of the person are computed. They are listed in Table 2.3. For each person to describe, we obtain 39 texture-based features.

Combing these two kinds of features, we computed a total of 93 features for each person: 54 color-based features and 39 texture-related features, which are described in detail in Table A.1 in Appendix A.

Type of Feature	Description
Color	Mean Value for the normalized value of each part
	Standard Deviation for the normalized value of each part
	Histogram with 4 beams for the normalized value of each part
Texture	Energy Correlation Inertia Entropy Inverse Difference Moment Sum Average Sum Variance Sum Entropy Difference Average Difference Variance Difference Entropy Information Measure of Correlation 1 Information Measure of Correlation 2

Table 2.3: The initial features based on color and texture

2.5.3 Feature selection

Three feature selection methods previously described are chosen and compared: PCA, CFS and Wrapper (described in Section 2.4).

- **PCA:** In this work, when the sum of the variance is at least equal to 95% of the initial variance of the data, the process is interrupted and the subspace is considered as an optimal answer.

  As a result, 26 features are created, which are linear combinations of the 93 initial features. These 26 new data for each of the 19135 images constituted a new dataset (CASIA-PCA).

- **CFS:** Best-first search method is used in the feature space. It starts with an empty set and search forward by adding some features (all ordered by their merit). We have to specify the number of steps that stop the search for a specific path. This is defined when 5 nodes does not significantly improve the merit. At this point, it will backtrack to search in both direction from this subset. This algorithm gives a subset of 40 features, which contains the most representative features (the most correlated to the class) with the less possible redundancy. The features selected in the final subset (CASIA-CFS) are listed in Table 2.4, where, for instance, $5-color-head$ means that there are 5 color features chosen in the head part. It is worth noticing that the texture features are less represented in the final subset and the most important part of the subset is given by the bottom part of the body (the largest).

- **Wrapper:** It determines the best subset by considering *a priori* the classifier that we will use (in our case SVM). For the search method, it is the same, best-first, with the same parameters used by CFS. As a result, a new subset (CASIA-Wrapper) is created with a merit of 0.127 (the global error rate in %), which contains 16 features selected by Wrapper method, as presented in Table 2.4. Color features, especially their *Mean Values* and *Standard Deviation Values*, are well represented and texture features (*Entropy*) are selected in all parts of the body.

FeatureSet	FeatureNumber	Description
CASIA-Wholeset	93	initial color and texture features
CASIA-PCA	26	linear combinations of original features
CASIA-CFS	40	5-color-head 13-color-top 14-color-bottom 1-texture-head 4-texture-top 3-texture-bottom
CASIA-Wrapper	16	4-color-head 4-color-top 4-color-bottom 1-texture-head 2-texture-top 1-texture-bottom

Table 2.4: Four feature subsets for CASIA Database

In the end, four sets of features are prepared for the classification stage: CASIA-Wholeset (initial set of 93 features), CASIA-PCA, CASIA-CFS and CASIA-

Wrapper. In Appendix A, Table A.2 describes in detail the features selected in each subset.

For PCA method, 26 features are obtained instead of the initial 93 features with a 99.6% relevance (the average of ROC Area). However, the disadvantage of PCA method is that it looses the interpretation of the features, because the features selected by PCA are the combinations of the initial features. Another disadvantage is that the construction of the set selected by PCA highly depends on the initial data. In CASIA-CFS, most of the selected features are color-based. Comparing CASIA-CFS with CASIA-Wrapper feature sets, 11 color features are in both sets. In addition, when we carefully look at the values of the covariance matrix of PCA (that give us the importance of each feature in the linear combination used to the new vectors), we can notice that the ones that are selected by the other methods have the highest coefficients. It is obvious that color-based features are more useful in people classification in our system. Besides, Entropy features (part of texture features) could bring an interesting additional information.

2.5.4 Experimental results

In this work, firstly, one-against-one SVM classifier with RBF kernel is tested to classify 20 persons in CASIA Database, who are represented by four features subsets. In the experiments, the SVM classifier is tested with a stratified 10-fold cross validation. For each person, the number of the total images for several sequences and several directions is about 1000. For each fold, we randomly choose 90% images of each person used as the training samples, and all the rest are used for the testing. The experimental results based on four different feature subsets are shown in Table 2.5. The classification rates of all classes on the four feature subsets are satisfying, which are higher than 99%. This result proves that the 20 persons are well characterised by the extracted color and texture features. We notice that the classification results on three selected feature subsets are similar to one on CASIA-Wholeset. It shows that these selected features keep the main information and meet the requirements of this classification system. According to the comparison of three feature subsets, CFS gives the best results with 99.9%. The global classification rate on CASIA-Wholeset is also 99.9%. However, it is worth noticing the fact that the number of dimensions of CASIA-Wholeset and CASIA-CFS is 93 and 40, respectively. Obviously, CASIA-CFS has the advantage of computational cost.

By comparison, the classifier based on all-against-all SVM with RBF kernel is also tested in this human recognition system. Similarly, the proposed four feature subsets represented 20 persons in CASIA Database are treated as inputs of the classifier. Similarly, we also test the SVM classifier with 2-fold cross validation. Table 2.6 shows the satisfying experimental results, which are higher than 95% of

	Wholeset (%)	PCA (%)	CFS (%)	Wrapper (%)
C1	99.8	97.7	99.8	97.4
C2	100	99.8	99.8	99.8
C3	100	99.9	100	100
C4	99.2	98.6	99.2	98.8
C5	99.8	99.1	99.8	99.8
C6	99.6	98.6	99.5	97.5
C7	100	100	100	100
C8	100	99.9	100	100
C9	100	99.9	100	100
C10	100	100	100	100
C11	100	100	100	99.6
C12	100	98.5	99.9	99.3
C13	100	99.1	100	98.92
C14	99.6	98.2	99.8	97.6
C15	100	99.3	100	99.6
C16	99.9	99.6	99.9	99.3
C17	99.8	99.4	99.7	99.2
C18	99.6	96.9	99.8	98.3
C19	100	97.4	100	99.4
C20	100	100	100	100
Global(%)	99.9	99.1	99.9	99.2

Table 2.5: Recognition rate of one-against-one SVM based on the four proposed databases

global classification rate for the four feature subsets. The comparison between Table 2.5 and Table 2.6 shows that one-against-one SVM algorithm performs a litter better than all-against-all one in this offline human recognition system. Besides, it works much fast than all-against-all SVM. However, the classical one-against-one SVM classifier will bring difficulties if we want to achieve online recognition, in which the classifier needs to online update with the new data containing new information of persons or environments. As a consequence, all-against-all SVM algorithm will be used for online human recognition system in the next chapter.

2.6 Conclusion

This chapter has presented an offline human recognition system, in which SVM is chosen as the classifier. A review of SVM has been presented with the evolution from binary SVM to multi-category SVM. These kinds of multi-class SVM algorithm are described and discussed: one-against-rest, one-against-one and all-against-all. Features have to be extracted to represent persons and considered as the input of the classifier. A review of feature extraction and feature selection methods has been presented. Color, texture and sharp features (or their fusion features) are usually taken into account. In order to extract complete characteristics of humans, the dimensionality of feature set could be quite high. Inevitably, some

	Wholeset (%)	PCA (%)	CFS (%)	Wrapper (%)
C1	96.66	94.37	96.64	92.17
C2	99.99	99.99	99.99	99.87
C3	99.76	98.88	99.63	99.11
C4	98.23	94.11	97.24	90.40
C5	99.51	99.00	99.51	98.74
C6	97.50	94.73	95.89	87.76
C7	99.74	98.60	99.62	95.83
C8	100.00	99.24	99.76	99.76
C9	100.00	99.23	100.00	99.13
C10	100.00	100.00	100.00	100.00
C11	99.13	98.51	99.36	91.90
C12	98.51	95.98	97.02	93.91
C13	99.14	98.88	99.01	92.91
C14	98.38	96.89	98.63	93.28
C15	99.51	95.60	98.99	94.84
C16	99.12	99.11	98.87	95.73
C17	99.13	98.74	98.88	96.13
C18	96.89	94.46	97.00	91.49
C19	98.89	93.97	97.50	94.77
C20	99.99	100.00	100.00	97.27
Global(%)	99.00	97.51	98.68	95.25

Table 2.6: Recognition rate of all-against-all SVM based on the four proposed databases

of the data in the initial feature set are irrelevant or redundant. As a consequence, feature selection or dimension reduction is necessary and significant to the subsequent classification step, which can reduce original feature set, improve prediction accuracy and reduce computational time.

There are three main approaches to feature subset selection described in this work. PCA is a well-known method of dimensionality reduction, which also has ability for feature selection. CFS works only on the inner properties of the dataset (distribution of the values, correlation between features and with the class, etc.), without considering the classification algorithm. As a result, it is fast and has a low computational cost. However, wrapper methods use a criterion or an heuristic to evaluate the different subsets depending on the performance of a selected classifier. They easily reach good performance in classification but are prone to have a risk of overfitting and to have a high computational cost.

In the experiments, CASIA Gait Database is chosen, in which each of the 20 persons are segmented into three parts. Color and texture features are then extracted from each part of each person for representation. The whole feature set contains 93 dimensions. Three dimensionality reduction methods, PCA, CFS and wrapper, are executed to select the optimal feature subset. Finally, four feature subsets are given and called CASIA-Wholeset, CASIA-PCA, CASIA-CFS and CASIA-Wrapper with 93, 26, 40 and 16 dimensional features, respectively.

One-against-one SVM with RBF kernel is used for multi-category classification. The results show good classification performances with the proposed four feature subsets in off-line system. Results for each subset are moreover very similar.

In the next chapter, an online human recognition system will be presented. The used database (same as in this chapter) will be fixed with 20 known persons captured by a static camera. The human recognition system will automatically identify person in real-world environments with only few images for initial learning. Online learning algorithms will be discussed and employed for human recognition in real-world application. Similarly, the above four feature subsets will be considered as an input for this online classification.

Chapter 3

Online human recognition system

3.1 Introduction

Chapter 2 has presented an offline human recognition system and its application achieved high classification rate on CASIA Gait Database using SVM classifier. However, offline system can not well satisfy real-world applications. Indeed, it requires that the test stage begins only when the whole learning process has finished, as a consequence, class models are fixed in the test procedure. That is to say, an offline classifier model can not adaptively updated during the test procedure even if very important information is occurring.

In realistic environment, the information about the persons is time-varying, not only due to the different possible positions, clothes, poses and expressions; but also due to environmental conditions, such as illumination variation. Even a very huge static database of people's images can not express the whole set of possibilities [MTP03]. Furthermore, it is difficult to have a complete knowledge about a person in order to recognize him/her in video sequences, whatever its orientation, background condition, etc. How to efficiently update the classifier model with the variable conditions is a challenge.

In order to establish a more robust and practical classification system, it is necessary to achieve unknown (new) person classification. Unknown person indicates a person that is totally new to the system, for which there is no information in the initial learning stage. A satisfying classification system can identify a new person and then save it as a new class, which can be used in the succeeding classification. How to successfully detect and classify new persons is another challenge.

In this chapter, Section 3.2 presents and uses an incremental and decremental SVM classifier as a solution to deal with the online classification problem of non-stationary data. The algorithm is briefly introduced and results of experiments on CASIA Gait Database are given in an online setting. Afterwards, in order to address the problem of new person classification, a clustering algorithm is applied. In Section 3.4, SAKM algorithm is presented and the results of a group of experiments are discussed. Section 3.5 concludes this chapter.

3.2 Incremental SVM

3.2.1 Literature review

Compared to batch algorithms, online or incremental algorithms have the great advantageous of dealing with large and non-stationary data. Indeed, they have ability to add new information to an existing trained model. Online methods are particularly useful in situations that involve online streaming data [ASK08]. In 2009 [LL09], Liang and Li have proved that incremental SVM is suitable for large dynamic data and more efficient than batch SVMs as far as the computing time is concerned.

The work of Syed et al. in 1999 [SHKS99] is considered as one of the first SVMs allowing incremental learning. The learning machine is incrementally trained on new data using previous support vectors only. Based on the algorithm in [SHKS99], an extension called SV-L-incremental algorithm has been developed by Ruping [Rüp01], which adjusts the objective function by adding a new regularization term. However, these works give only approximate results. In 2001, Cauwenberghs and Poggio [CP01] designed an exact online incremental learning SVM, which update the decision function parameters when recursively adding or deleting points one at a time. In 2003, Diehl and Cauwenberghs [DC03] improved the previous work and generalized a framework for exact incremental SVM. Instead of randomly choose training data, Cheng [CJ11] collects initial training sample set by using *k*-mean clustering algorithm, the experimental result shown that this method is more effective and robust. An extension of single incremental/decremental algorithm is proposed by Karasyama, in [KT10]. Multiple incremental/decremental SVM is introduced to allow addition or deletion of multiple data samples, in which the conventional single parametric programming is extended to multi-parametric programming. The experiments have proven that this algorithm is useful to reduce the computational cost. In 2012 [KHST12], he proposed multiple incremental/decremental algorithm for instance-weight support vector machines (WSVM).

The above methods are mainly developed to deal with binary SVM classification. There are not many researches on incremental and decremental multi-class problem. In order to solve this problem, Boukharouba et al. [BBL09] proposed a multi-class incremental and decremental support vector classifier, named MID-SVM. The experimental results on synthetic and real-world dataset showed the feasibility and good performance of the proposed method. In this work, the classification algorithm in [BBL09] is re-implemented and used in a realistic environment and firstly applied to solve online human recognition problem.

3.2.2 MID-SVM

MID-SVM, a single sample incremental and decremental algorithm, is an exact online method that keeps the Karush-Kuhn-Tucker (KKT) conditions of one single optimization problem satisfied between multiple classes, when adiabatically adding or eliminating data.

MID-SVM algorithm is based on all-against-all multi-class classification approach, which has been described in Section 2.2.2.3. Let's simply present again the classification principal of all-against-all SVM. Consider a dataset T of N pairs (x_i, y_i), where $i = 1, \cdots, N$, $x_i \in R^d$ is the input data, $y_i \in \{1, \cdots, K\}$ is the output class label. The SVM classifier used for data classification is defined by:

$$x_i \in C_k; \quad k = \underset{j=1,\cdots,K}{\arg} \max f_j(x_i) \tag{3.1}$$

Each decision function f_j is expressed as: $f_j(x) = w_j^T \Phi(x) + b_j$, where $j = 1, \cdots, K$ and the function $\Phi(x)$ maps the original data x to a possible higher-dimensional space to solve non-linear problems, where the initial non-linear problem could be solved as a linear problem. In multi-category classification, as described in Figure 3.1, the margin between classes i and j is $2/||w_i - w_j||$. In order to get the largest margin between classes i and j, minimization of the sum of $||w_i - w_j||^2$ for all $i = 1, \ldots, K$ and $j = 1, \ldots, K$, $i \neq j$ is computed. Also, as described in [BB99], the regularization term $\frac{1}{2}\sum_{i=1}^{K} ||w_i||^2$ is added to the objective function. In addition, a loss function $\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l \in C_{ij}} \xi_l^{ij}$ is used to find the decision rule with the minimal number of errors, where $C_{ij} = C_i \cup C_j$ and the slack variable ξ_l^{ij} measures the degree of misclassification of the l^{th} training vector, related to the hyperplane between the class C_i and C_j. In this case, the proposed quadratic function is presented as follows:

$$\begin{aligned} \min_{w_i, b_i} \quad & \frac{1}{2}\sum_{i=1}^{K}\sum_{j=i+1}^{K} ||w_i - w_j||^2 + \frac{1}{2}\sum_{i=1}^{K} ||w_i||^2 + C\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l \in C_{ij}} \xi_l^{ij} \\ s.t. \quad & \forall x_l \in C_{ij}; \\ & y_l^{ij}\left[(w_i - w_j)^T \Phi(x_l) + (b_i - b_j)\right] - 1 + \xi_l^{ij} \geq 0; \\ & \xi_l^{ij} \geq 0; \quad i = 1, \cdots, K; \quad j = i+1, \cdots, K \end{aligned} \tag{3.2}$$

whree $y_l^{ij} = \begin{cases} 1 & if\, x_l \in C_i \\ -1 & if\, x_l \in C_j \end{cases}$

C (s.t. $C \geq 0$) is the trade-off term between the margin of hyperplane and the number of errors.

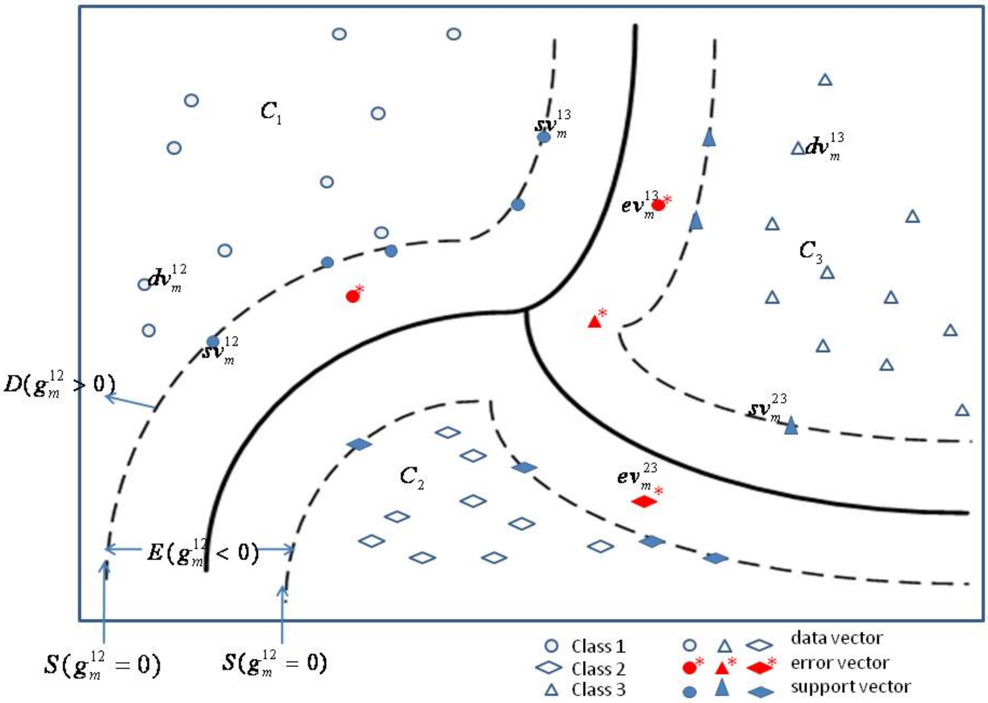


Figure 3.1: The boundaries and margins for three class classification problem [BBL09]

In order to minimize this objective function, which is a quadratic programming task, we solve it by Lagrange multipliers method. The Lagrange function L is defined by:

$$\begin{aligned} L = & \frac{1}{2}\sum_{i=1}^{K}\sum_{j=i+1}^{K}||w_i - w_j||^2 + \frac{1}{2}\sum_{i=1}^{K}||w_i||^2 + C\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l \in C_{ij}}\xi_l^{ij} - \sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l \in C_{ij}}\mu_l^{ij}\xi_l^{ij} \\ & - \sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l \in C_{ij}}\alpha_l^{ij}\left(y_l^{ij}[(w_i - w_j)^T\Phi(x_l) + (b_i - b_j)] - 1 + \xi_l^{ij}\right) \end{aligned} \tag{3.3}$$

where $\alpha_l^{ij} \geq 0$, $\mu_l^{ij} \geq 0$, $i \neq j$ are Lagrange coefficients.

The Lagrangian L has to be minimized with respect to w_i, b_i and ξ_l^{ij} and maximized with respect to α_l^{ij} and μ_l^{ij}. Then, at the saddle point, the derivation of the Lagrangian L is equal to zero. For all $i = 1,\dots,K$, we get the following equations: $\frac{\partial L}{\partial w_i} = 0$, $\frac{\partial L}{\partial b_i} = 0$ and $\frac{\partial L}{\partial \xi_l^{ij}} = 0$. We obtain [Bou11]:

$$w_i = \frac{1}{K+1} \sum_{\substack{j=1 \\ j \neq i}}^{K} \Big(\sum_{x_l \in C_i} \alpha_l^{ij} \Phi(x_l) - \sum_{x_l \in C_j} \alpha_l^{ij} \Phi(x_l) \Big) \tag{3.4}$$

$$\sum_{\substack{j=1 \\ j \neq i}}^{K} \Big(\sum_{x_l \in C_i} \alpha_l^{ij} - \sum_{x_l \in C_j} \alpha_l^{ij} \Big) = 0 \tag{3.5}$$

$$\alpha_l^{ij} + \mu_l^{ij} = C \tag{3.6}$$

Based on the kernel trick $\Phi(x_l)^T \Phi(x) = k(x_l, x)$ and the expression of w_i in Equation 3.4, the decision function can be expressed as:

$$\begin{aligned} f_i(x) &= w_i^T \Phi(x) + b_i \\ &= \left[\frac{1}{K+1} \sum_{\substack{j=1 \\ j \neq i}}^{K} \left(\sum_{x_l \in C_i} \alpha_l^{ij} \Phi(x_l) - \sum_{x_l \in C_j} \alpha_l^{ij} \Phi(x_l) \right) \right] \Phi(x) + b_i \end{aligned} \tag{3.7}$$

$$f_i(x) = \frac{1}{K+1} \sum_{\substack{j=1 \\ j \neq i}}^{K} \left(\sum_{x_l \in C_i} \alpha_l^{ij} k(x_l, x) - \sum_{x_l \in C_j} \alpha_l^{ij} k(x_l, x) \right) + b_i \tag{3.8}$$

After substituting the KKT condition into the primal Lagrangian L (Equation 3.3), we derive the dual Lagrangian function W of the decision function f (Equation 3.8). The KKT conditions at a point x_m ($x_m \in C_{ij}$, for all $i = 1, ..., K-1$, $j = i+1, ..., K$) can be expressed by computing the following gradient: $g_m^{ij} = \frac{\partial W}{\partial \alpha_m^{ij}}$ and $\frac{\partial W}{\partial b_i} = 0$.

The KKT conditions on the point $x_m \in C_{ij}$ divide data D into three categories according to the value of g_m^{ij} for all $i = 1, ..., K-1$, $j = i+1, ..., K$:

$$g_m^{ij} = \frac{\partial W}{\partial \alpha_m^{ij}} \begin{cases} > 0; & if \quad \alpha_m^{ij} = 0; & D(dv_m^{ij}) \\ = 0; & if \quad 0 < \alpha_m^{ij} < C; & S(sv_m^{ij}) \\ < 0; & if \quad \alpha_m^{ij} = C; & E(ev_m^{ij}) \end{cases} \tag{3.9}$$

Considering a three-class problem for example, as explained in Figure 3.1, support vectors (S) are on the boundary, error vectors (E) exceed the margin and data vectors (D) are inside the boundary.

The incremental learning manner of MID-SVM is the adiabatic increment [CP01], which is based on the following idea: when a new data x_c is received, we initially set the coefficients $\alpha_c^{pq} = 0$, where $p = 1, ..., K-1, q = p+1, ..., K$ and the parameters of the existing support vectors are adapted in order to keep all training samples to satisfy to the KKT conditions. In particular, when the new data x_c is added, the KKT conditions can be expressed differentially as:

$$\begin{aligned}\Delta g_m^{ij} \quad = \quad & y_m^{ij}\Bigg(\beta_c^{ij,pq}\Delta\alpha_c^{pq}K_{cm} + \sum_{x_l \in C_i}\Big(2\Delta\alpha_l^{ij} + \sum_{\substack{n=1\\n\neq i,j}}^{K}\Delta\alpha_l^{in}\Big)K_{lm} - \sum_{x_l \in C_j}\Big(2\Delta\alpha_l^{ij} \\ & + \sum_{\substack{n=1\\n\neq i,j}}^{K}\Delta\alpha_l^{jn}\Big)K_{lm} - \sum_{\substack{n=1\\n\neq i,j}}^{K}\sum_{x_l \in C_n}\Big(\Delta\alpha_l^{in} - \Delta\alpha_l^{jn}\Big)K_{lm} + \Big(\Delta b_i - \Delta b_j\Big)\Bigg)\end{aligned} \tag{3.10}$$

$$\gamma_c^{i,pq}\Delta\alpha_c^{pq} + \sum_{\substack{n=1\\n\neq i}}^{K}\Big(\sum_{x_l \in C_i}\Delta\alpha_l^{in} - \sum_{x_l \in C_j}\Delta\alpha_l^{in}\Big) = 0 \tag{3.11}$$

where $i = 1, ..., K-1, j = i+1, ..., K$, α_c^{pq} is the coefficient being incremented.

Coefficients $\beta^{ij,pq}, \gamma^{i,pq}$ are defined as:

$$\beta_c^{ij,pq} = \begin{cases} 2 & if(p,q) = (i,j)\ and\ x_c \in C_i \\ -2 & if(p,q) = (i,j)\ and\ x_c \in C_j \\ 1 & if\ p = i, q \neq j\ and x_c \in C_i \quad or\ p \neq i, q = j\ and x_c \notin C_{ij} \\ -1 & if\ p \neq i, q = j\ and x_c \in C_j \quad or\ p = i, q \neq j\ and x_c \notin C_{ij} \\ 0 & otherwise \end{cases} \tag{3.12}$$

$$\gamma_c^{i,pq} = \begin{cases} 1 & if\ x_c \in C_i \\ -1 & if\ x_c \notin C_i \end{cases} \tag{3.13}$$

Finally, the solution function is given as follows: (see [Bou11] for more detail.)

$$\begin{bmatrix} \Delta b \\ \Delta a \end{bmatrix}\Delta\alpha_c^{pq} = -\mathbf{RH}^{pq}\Delta\alpha_c^{pq} \tag{3.14}$$

where $b = [b_1, ..., b_K]^T$ and $a = [a_1^{12}, a_2^{12}, ...a_i^{ij}, a_j^{ij}, ..., a_{K-1}^{(K-1)K}, a_K^{(K-1)K}]^T$, a_i^{ij} contains the weights of all support vectors sv_n^{ij} that belong to the class C_i ($a_i^{ij} = [\alpha_1^{ij}, \ldots, \alpha_n^{ij}]$).

3.2.3 Increment and decrement procedures

3.2.3.1 Increment procedure

The increment procedure can be used when a new sample is added to the previous training data, to the class C_p, where $p \in [1, \ldots, K]$. To simplify the explanation of the incremental learning procedure, we take a two-class classification for example (class C1 and C2). The procedure is described in detail by Figure 3.2. When a new sample x_c is added to the incremental learning two-class classifier, depending on the value of g_c^{12} and α_c^{12}, the new data x_c is recognized as a support vector, error vector or data vector. The discrimination is based on the KKT conditions, illustrated in the first part of Figure 3.2. If x_c is classified as a support vector, the previous support vector set S as well as the corresponding classification boundaries and margins should be updated. While $g_c^{12} > 0$ and $\alpha_c^{12} < C$, we apply the increment of α_c^{12} and update (b, a) using Equation 3.14 at each iteration until x_c is classified in S, D or E. During this procedure, if the value of α_m^{12} becomes zero (x_m is originally a support vector in C1 or C2), x_m is removed from support vector set S to data vector set D and the corresponding classification boundaries and margins should be updated also.

According to multi-category classification, the increment procedure of the MID-SVM algorithm is illustrated by Algorithm 1 [MTP03, CP01].

3.2.3.2 Decrement procedure

When an old vector x_z has to be removed (from the training data) from the existed class C_e, where $e \in [1, \ldots, K]$, the decrement procedure is applied. Two-class classification is taken as an example, the decrement procedure is illustrated in Figure 3.3. If x_z belonged to the set of data vectors ($\alpha_z^{12} = 0$), it will be left out and the learning procedure will terminate without any boundary updating. If $g_z^{12} < -1$, x_z was an error vector. Otherwise, we decrement α_c^{12}: $\alpha_c^{12} = \alpha_c^{pq} - n$ and update the value of $[b, a]$ according to Equation 3.14. During this process, if $g_m^{12} < 0$ ($x_m \in D$), we terminate the decrement procedure and apply the increment procedure until x_m is classified as a support vector ($dv_m^{12} \rightarrow sv_m^{12}$). If S changes, update $\mathbf{R}$, $[b, a]$ and $\mathbf{H}^{pq}$ and boundary updates accordingly. Then return to the decrement procedure.

Similarly, the decrement procedure of MID-SVM algorithm is illustrated by Algorithm 2 [MTP03, CP01].

3.2.4 Migration of data between the three sets

When a new data is added or removed, the hyperplane of the SVM classifier will be updated if this training data is not classified into the data vector set (D), and

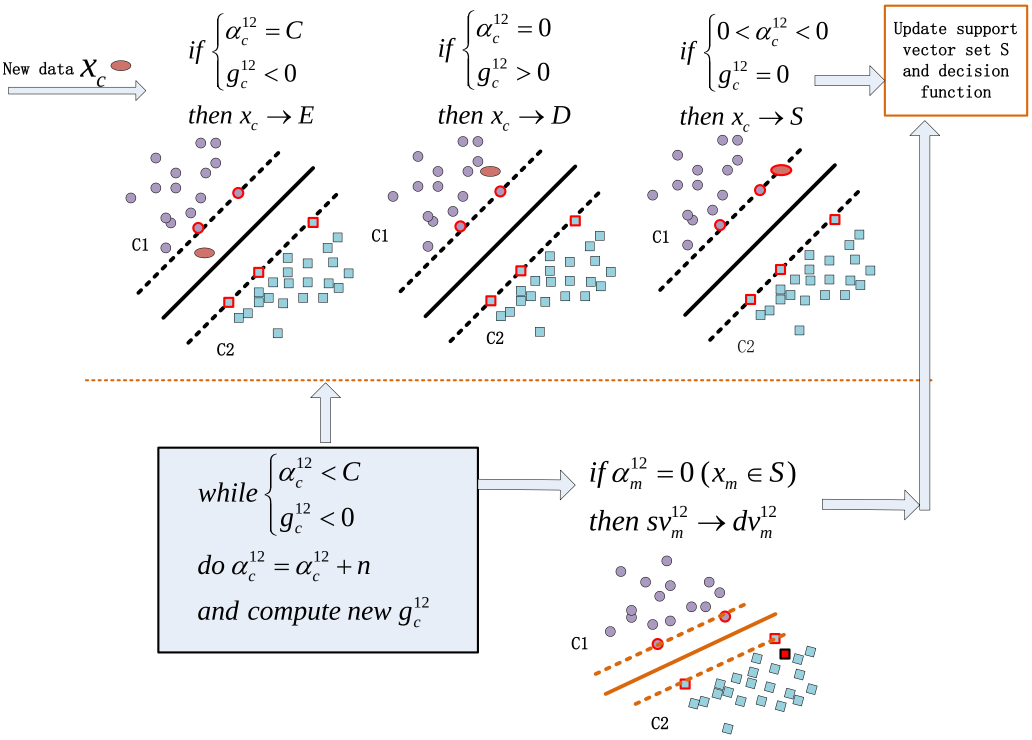

New data X_c
if $\begin{cases}\alpha_c^{12} = C \\ g_c^{12} < 0\end{cases}$
then $x_c \to E$
if $\begin{cases}\alpha_c^{12} = 0 \\ g_c^{12} > 0\end{cases}$
then $x_c \to D$
if $\begin{cases}0 < \alpha_c^{12} < 0 \\ g_c^{12} = 0\end{cases}$
then $x_c \to S$
Update support vector set S and decision function
C1
C2
C1
C2
C1
C2
while $\begin{cases}\alpha_c^{12} < C \\ g_c^{12} < 0\end{cases}$
do $\alpha_c^{12} = \alpha_c^{12} + n$
and compute new g_c^{12}
if $\alpha_m^{12} = 0\ (x_m \in S)$
then $sv_m^{12} \to dv_m^{12}$
C1
C2

Figure 3.2: Work flow of incremental learning procedure (considering a two-class classification for example)

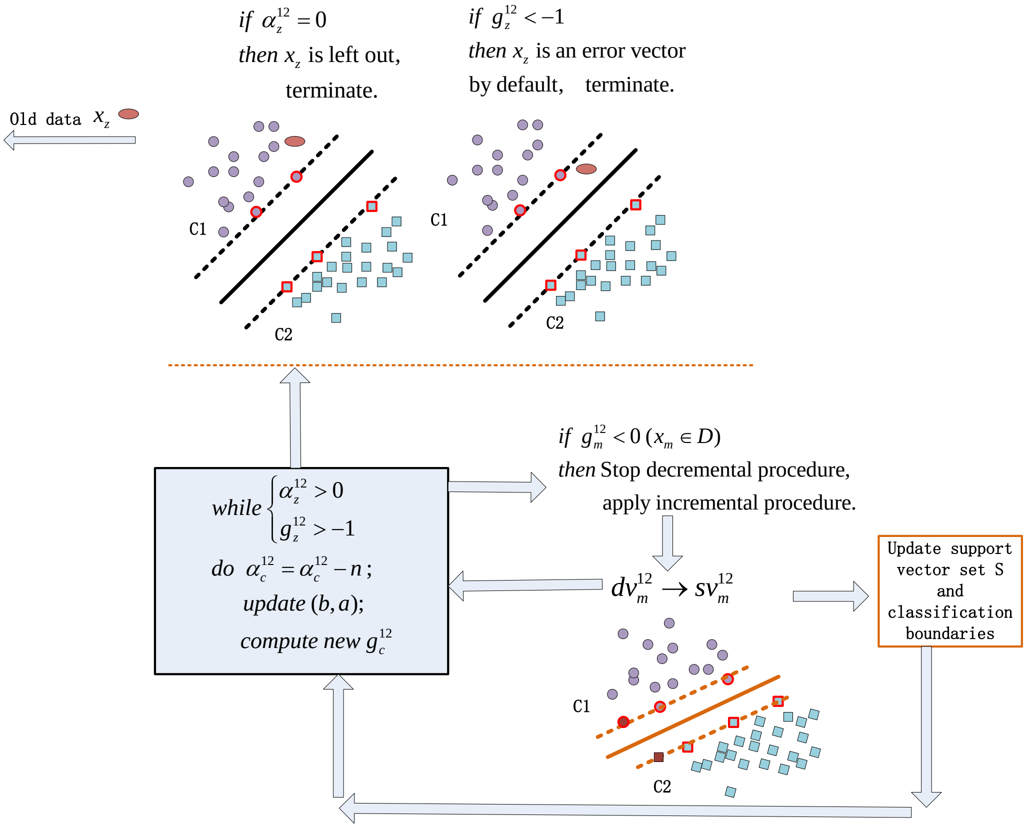

if $\alpha_z^{12} = 0$
then x_z is left out,
terminate.
if $g_z^{12} < -1$
then x_z is an error vector
by default, terminate.
Old data x_z
C1
C2
C1
C2
if $g_m^{12} < 0 \; (x_m \in D)$
then Stop decremental procedure,
apply incremental procedure.
while $\begin{cases} \alpha_z^{12} > 0 \\ g_z^{12} > -1 \end{cases}$
do $\alpha_c^{12} = \alpha_c^{12} - n$;
update (b, a);
compute new g_c^{12}
$dv_m^{12} \rightarrow sv_m^{12}$
Update support vector set S and classification boundaries
C1
C2

Figure 3.3: Work flow of decremental learning procedure (considering a two-class classification for example)

Algorithm 1: Increment Procedure of MID-SVM algorithm [MTP03, CP01].

```
$x_c \rightarrow C_p,\ p \in [1,\dots,K]$;
Set $n > 0$;
for $q = 1,\dots,K, q \neq p$ do
    $\alpha_c^{pq} \leftarrow 0$;
    $Z = [b,a]$;
    Compute $H^{pq}$;
    while $g_c^{pq} < 0$ and $\alpha_c^{pq} < C$ do
        $\alpha_c^{pq} = \alpha_c^{pq} + n$;
        $Z = Z - \mathbf{R} \cdot \mathbf{H}^{pq} \cdot n$;
        $g_k^{ij} = g_k^{ij} + \Delta g_k^{ij},\ i = 1,\dots,K-1,\ j = i+1,\dots,K, x_k \in C_{ij}$;
        if $\alpha_m^{ij}$ are close to zero then
            Corresponding support vectors move inside the margin
            $\left(sv_m^{ij} \rightarrow dv_m^{ij}\right)$;
            Update $\mathbf{R}, Z$ and $\mathbf{H}^{pq}$;
        end
    end
    if $\alpha_c^{pq} = C$ then $x_c$ is added to $C_p$ as an error vector;
    else $x_c$ is considered as a support vector.
end
```

then the previous vectors in different sets (S, E and D) could migrate from their current set to a neighbour set. Fig. 3.1 explains the geometrical interpretation of each set and from this figure we can infer the possible migrations as follows:

- From D to S (only in the decremental learning procedure): the data vector becomes a support vector. This case happens when the update value of g_m^{ij} for $x_m^{ij} \in D$ reaches 0 ($\mathrm{g}_\mathrm{m}^\mathrm{ij} > 0 \rightarrow \mathrm{g}_\mathrm{m}^\mathrm{ij} = 0$).

- From E to S: the error vector becomes a support vector. This case happens when the update value of g_m^{ij} for $x_m^{ij} \in E$ reaches 0 ($\mathrm{g}_\mathrm{m}^\mathrm{ij} < 0 \rightarrow \mathrm{g}_\mathrm{m}^\mathrm{ij} = 0$).

- From S to E: the previous support vector becomes an error vector. This case happens when α_m^{ij} ($x_m^{ij} \in S$) is equal to C ($0 < \alpha_\mathrm{m}^\mathrm{ij} < C \rightarrow \alpha_\mathrm{m}^\mathrm{ij} = C$).

- From S to D: the previous support vector becomes a data vector. This case happens when α_m^{ij} ($x_m^{ij} \in S$) is equal to 0 ($0 < \alpha_\mathrm{m}^\mathrm{ij} < C \rightarrow \alpha_\mathrm{m}^\mathrm{ij} = 0$).

Algorithm 2: Decremental Procedure of MID-SVM algorithm [MTP03, CP01].

```
$x_z \leftarrow C_p$, $p \in [1,\ldots,K]$;
Set $n > 0$;
for $q = 1,\ldots,K, q \neq p$ do
    $\alpha_z^{pq} \leftarrow$ weight of $x_z$;
    $Z = [b,a]$;
    Compute $H^{pq}$;
    while $g_z^{pq} > -1$ and $\alpha_z^{pq} > 0$ do
        $\alpha_z^{pq} = \alpha_z^{pq} - n$;
        $Z = Z + R \cdot H^{pq} \cdot n$;
        $g_k^{ij} = g_k^{ij} + \Delta g_k^{ij}$, $i = 1,\ldots,K-1, j = i+1,\ldots,K, x_k \in C_{ij}$;
        if $g_l^{ij} < 0$ is happened ($x_l \in D$) then ;
        The data $x_l$ is updated to be a support vector ($dv_l^{ij} \rightarrow sv_l^{ij}$) and incremental procedure is applied;
        else Return;
    end
    if $\alpha_z^{pq} = 0$ then $x_z$ is left out;
    else $x_z$ is by default as a training error.
end
```

3.2.5 Experimental results using incremental SVM

This section presents the experimental results of human recognition with incremental SVM classifier based on four feature subsets, which have been described in Section 2.5. Figure 3.4 illustrates the workflow of the data in our incremental system. CASIA Database is used, in which only the first 50 images of each class (about the first 5% samples of the whole dataset) are used for initialization and the remaining images are used for online learning (and to update the classifier). In this procedure, new frames are added one by one and the recognition system is updated step by step with an adaptive decision function. The difference between initial and online learning is on the labels considered. During the initialization step, the class labels of the added samples are correct, but in the online learning phase, the class labels are given by the current classifier model (first we classify the sample with the current SVM and update with this decision).

Table 3.1 presents the classification results of MID-SVM. The experimental results illustrate that the proposed incremental SVM is able to meet the demands of online multi-category classification and achieves satisfying performance. The results of the different classes vary among the four different feature sets. The

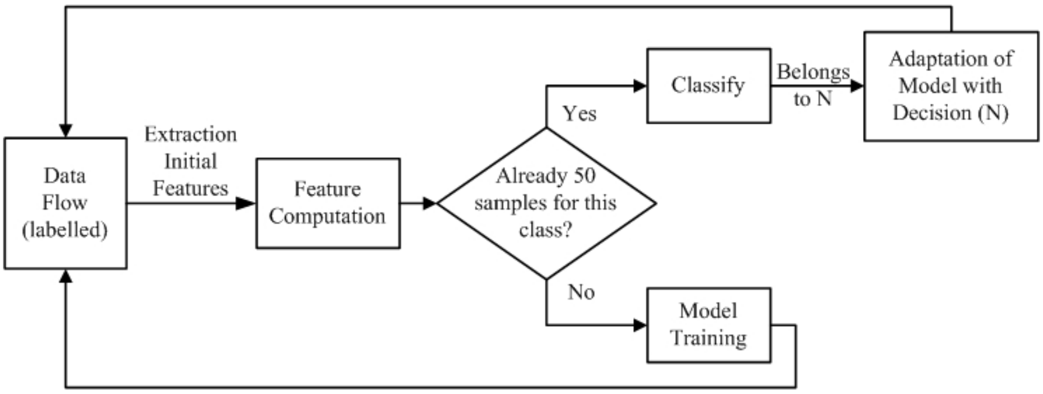


Figure 3.4: Incremental learning work flow

global recognition results for the four feature sets are encouraging (over 95%). The best result is given by CASIA-CFS subset, which is as high as 98.46%. Some classes are less correctly recognized, such as C1 and C18 with recognition rates below 93% in the four datasets. By step-by-step checking during the online learning procedure, we find that these wrong classifications are caused by some 'false' support vectors occurring in the procedure of online learning. These false support vectors could be derived from noise vectors and incorrectly classified as the support vectors of some classes during the online process. This could influence the global performance of classification by wrongly updating the decision function and inducing false results on the remaining classification.

As shown in Table 3.1, CASIA-PCA and the CASIA-Wrapper datasets have lower performance than the others. CASIA-CFS dataset gives the best performance with a 98.46% global recognition rate, which is similar with the result of the whole dataset. In CASIA-CFS dataset, more features are extracted compared to CASIA-PCA and CASIA-Wrapper, but it considerably reduces the number of features comparing to the initial feature set.

In Chapter 2, classical all-against-all SVM classifiers (without incremental learning) has been implemented for offline person recognition. The experimental results are satisfying and described in Table 2.6. The all-against-all SVM classifier based on MID-SVM algorithm in this chapter is developed and suitable for online classification. Compared to the results of these two algorithms (by the comparison between Table 2.6 and Table 3.1), the performance of MID-SVM algorithm is comparable with one of the classical SVM. However, we notice the different item of percentage of learning data, that varies from 90% in classical all-against-all SVM to 5% in MID-SVM.

	CASIA-Wholeset	CASIA-PCA	CASIA-CFS	CASIA-Wrapper
C1	92.16	91.07	92.59	91.83
C2	99.44	99.72	99.44	99.44
C3	100	98.74	100	99.58
C4	95.77	95.37	97.99	91.15
C5	99.10	99	99.10	99.10
C6	99.37	92.15	99.79	92.25
C7	100	93.65	100	99.08
C8	100	99.76	100	99.88
C9	100	100	100	100
C10	100	100	100	100
C11	99.78	99.23	99.78	96.92
C12	99.31	96.11	98.28	95.65
C13	98.98	94	99.39	86.98
C14	97.34	94.42	97.72	93.03
C15	100	91.69	100	94.16
C16	96.87	97.74	98.62	95.11
C17	98.23	98.11	98.23	97.99
C18	91.89	85.70	92	83.88
C19	97.42	85.77	96.56	93.37
C20	99.22	100	100	99.67
Global	98.21	95.6	98.46	95.39

Table 3.1: Recognition rates of MID-SVM based on the four proposed databases.

3.3 Discussion

MID-SVM algorithm is suitable for online application, which has shown satisfying performance in our human classification system. However, it has no ability to solve the novelty detection problem. It needs to learn samples (at least few samples) of each class in the initialization procedure. In real-world video surveillance, it is impossible to obtain training samples of all classes, even only a few data. It is evident that the proposed MID-SVM algorithm could not address the problem of new person classification. Unsupervised learning algorithms offer the possibility for new class recognition, such as clustering algorithm. In the next section, an adaptive clustering algorithm will be described and tested in our human classification system.

3.4 Self-Adaptive Kernel Machine

3.4.1 Literature review

One-class SVM was introduced in 2001 by Schökopf et al. [SPST$^+$01], which extended the SVM classifier to handle training with only positive information. The algorithm has achieved successes in classification problems [MY02a], abnormal detection [LHTX03] and novelty detection [GKVL06]. Then, one-class

SVM has also been extended to online models based on the incremental learning algorithm in [CP01], such as [TL03]. However, these algorithms have high computational cost. In 2004, Kivinen et al. [KSW04] proposed the Naöe Online Regularized Risk Minimization Algorithm (NORMA) to address the online classification problem. NORMA algorithm implements online learning in a Reproducing Kernel Hilbert Space (RKHS) and iteratively updates the decision function based on stochastic gradient descent. Nevertheless, NORMA algorithm is not suitable for dealing with the problem of learning in non-stationary environment [BLM05]. Self-Adaptive Kernel Machine (SAKM), proposed by Amadou-Boubacar [BLM05], combines the advantages of NORMA algorithm and one-class SVM. It is a kernel-based algorithm initially proposed for online clustering in a multi-class environment. In this work, SAKM algorithm is applied to achieve person classification with novelty identification.

3.4.2 Overall formulation of SAKM

SAKM is a kernel-based algorithm for online unsupervised clustering of non-stationary data. All data $x_i \in X$ are mapped by a function Φ to a Hilbert Space H and clusters are represented by the distributions of support vectors in Reproducing Kernel Hilbert Space (RKHS). As shown in Figure 3.5, with a RBF kernel, all data are mapped on a quadrant of circle inner-product Hilbert Space H, based on the properties of Hilbert Space: $\langle \Phi(x_i), \Phi(x_i) \rangle_H = 1$ and $\left\langle \Phi(x_i), \Phi(x_j) \right\rangle_H \leq 1$, where $i \neq j$. One-class SVM is used as estimator in SAKM. Lagrangian technique is applied to determine the optimal hyperplane and the boundary function f is defined as:

$$f(x) = \sum_i \alpha_i k(x, x_i) + b \tag{3.15}$$

where α_i are the coefficients and $k(.,.)$ is the kernel function, which replaces the dot product of the mapping function $k(x_i, x_j) = \left\langle \Phi(x_i), \Phi(x_j) \right\rangle)$, and b is the offset of the decision function. In SAKM, the RBF kernel function is applied, which is as $k(x_1, x_2) = \exp(-\frac{1}{2\sigma^2} \|x_1 - x_2\|^2), \quad \forall (x_1, x_2) \in \mathrm{X}^2$, where parameter σ is called bandwidth.

According to the characteristic of online clustering, the clusters and the boundary decision functions could be updated over the time based on the SAKM updating rules. Let us define a set Ψ^t of temporal clusters C_n^t and a set F^t of temporal boundary functions f_n^t at time t :

$$\begin{aligned} \Psi^t &= \{C_1^t, \cdots C_n^t, \cdots C_Z^t\} \\ F^t &= \{f_1^t, \cdots f_n^t, \cdots f_Z^t\} \end{aligned} \tag{3.16}$$

where Z is the number of clusters at time t.

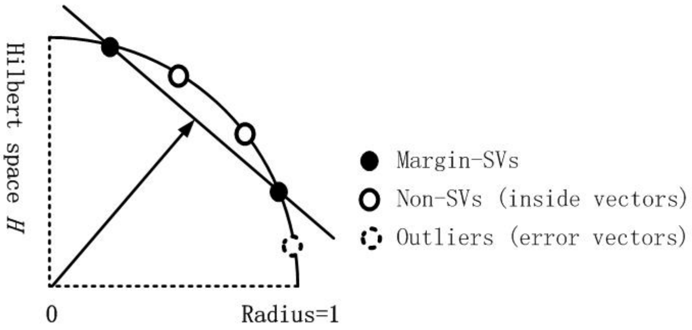


Figure 3.5: Data distribution of one-class in RKHS

At time t, the boundary function of the cluster f_n^t is decided by the Lagrangian parameters $\alpha_{i,n}^t$ of the support vectors $SV_{i,n}^t$ and the offset b_n^t.

$$f_n^t(x) = \sum_i \alpha_{i,n}^t k(x, SV_{i,n}^t) + b_n^t \tag{3.17}$$

3.4.3 Kernel-induced similarity measure

In data classification, an appropriate similarity measure is necessary to obtain reliable algorithms. According to SVM methods, one of the widely used measure is based on the kernel similarity function directly induced in the Hilbert space H. However, for non-stationary data in online multi-category context, the decision boundaries are variable over the time (the boundary at time t is based on the clusters determination at time $t-1$). As a result, kernel-induced similarity metric [Bor03] is used to evaluate the distance between the new data coming at time t and its nearest support vectors for all existing clusters known at time $t-1$.

In Hilbert space H, at time t, kernel-induced similarity metric is applied to compute the proximity level of a new data x_c and the support vector $SV_{i,n}$ of the cluster C_n. The winner support vector $SV_{win,n}$ is obtained by:

$$Win = \arg\min_i \|\Phi(x_c) - \Phi(SV_{i,n})\|_H \tag{3.18}$$

The distance between a data x_c and each cluster C_n is defined by kernel-induced similarity measure function $Dis_{c,n}$:

$$Dis_{c,n} = Dis(x_c, C_n) = \frac{\delta}{\sqrt{2}} \|\Phi(x_c) - \Phi(SV_{win,n})\|_H \tag{3.19}$$

where δ is a sign function as follows:

$$\delta = \begin{cases} 1 & if \quad f_n^t(x_c) < 0 \\ 0 & else \end{cases} \tag{3.20}$$

In Hilbert space, $\forall x_i \neq x_j \in \mathrm{X}$, $\langle \Phi(x_i), \Phi(x_i) \rangle_\Gamma = 1$, $\left\langle \Phi(x_i), \Phi(x_j) \right\rangle_\Gamma \leq 1$, when introducing the RBF kernel, Equation 3.19 is equal to:

$$\begin{aligned} Dis_{c,n} =& \delta \sqrt{1 - k(x_c, SV_{win,n})} \\ =& \delta \sqrt{1 - \exp(-\frac{1}{2\sigma^2} \|x_c - SV_{win,n}\|^2)} \end{aligned} \tag{3.21}$$

At time t, the set of the existing clusters is Ψ^t. When a data x_s is presented, according to this kernel-induced similarity measure function, the criterion of SAKM learning procedure is created following the equitation.

$$\Psi_{win} = \{C_n^t \in \Psi^t | Dis_{c,n} \leq Threshold\} \tag{3.22}$$

where $Threshold$ is a fixed acceptance threshold of kernel-induced similarity measure.

3.4.4 SAKM learning procedures for online clustering

Based on the evaluation of the kernel-induced similarity measure, SAKM process is composed of four steps: creation, adaptation, fusion and elimination.

- **Creation stage:** It occurs at the beginning of the learning or when a new cluster happens. The initialization of a cluster is done in this step.
- **Adaptation stage:** This procedure is based on stochastic gradient descent in Hilbert space. It updates the parameters (e.g., support vectors, hyperplanes, etc.) of the existing clusters.
- **Fusion stage:** In this procedure, clusters with similar information will be merged as one single cluster.
- **Elimination stage:** This procedure is used to delete the clusters with only few data. These data will be added in the test dataset and re-classify in a later step.

The above four stages are mainly based on SAKM decision rules. As a consequence, online learning procedures are called as follows:

$$\begin{cases} card(\Psi_{win}) = 0 & \rightarrow \quad Creation\ stage \\ card(\Psi_{win}) = 1 & \rightarrow \quad Adaptation\ stage \\ card(\Psi_{win}) \geq 2 & \rightarrow \quad Fusion\ stage \end{cases} \tag{3.23}$$

3.4.5 SAKM learning procedures for online semi-supervised classification

SAKM algorithm is originally proposed for online clustering. In this work, SAKM algorithm is used to solve the online human recognition problem, especially for new persons detection and classification. Firstly, some samples of familiar (known) persons are considered for the initialization stage to build an initial SAKM classifier model. Then this trained model is used to classify data coming from familiar and new persons. This classifier model can be adaptively updated using online learning procedure. This workflow is illustrated by Figure 3.6.

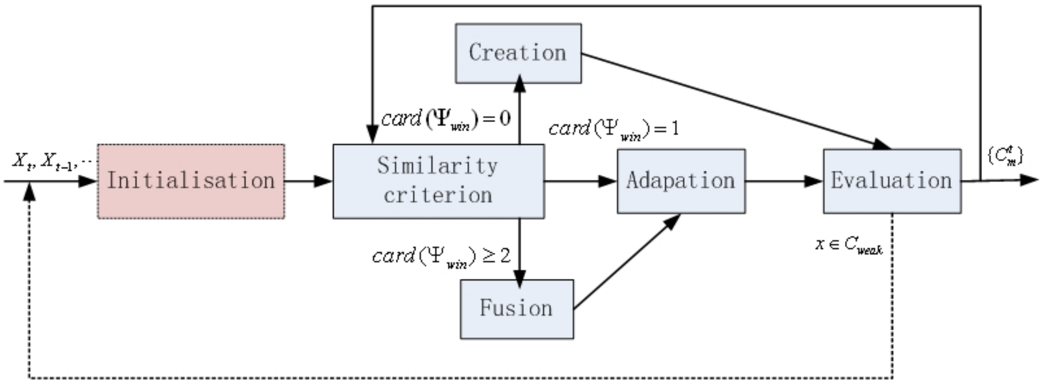


Figure 3.6: Workflow of SAKM for online classification

3.4.5.1 Initial learning procedure

Initialisation stage At time $t = 1$, the first data x_1 is the unique sample of cluster C_1, which is considered as a support vector $SV_{1,1} = x_1$. The coefficients of C_1 boundary function f_1 are initialized as: $\alpha_{1,1} = 1$, and $b_1 = \eta(1-\upsilon)$, where η is a learning rate and υ is the margin fraction of support and outliers vectors.

Adaptation stage The adaptation of SAKM algorithm is similar to NORMA, since it is based on the method of stochastic gradient descent. In the adaptation stage, the samples of the training classes are added to update the parameters (α_n, b_n) of cluster C_n. Gradient descent technique in RKHS is applied to update the parameters α_n of the boundary function f_n:

$$\left\{ \begin{array}{l} \alpha_{i,n}^{t+1} = (1-\eta)\alpha_{i,n}^{t}, \quad for \quad i < t \\ \left\{ \begin{array}{l} \alpha_{i,n}^{t+1} = \eta, \quad if \quad f_{i,n}^{t}(x_s) < 0 \\ \alpha_{i,n}^{t+1} = 0, \quad others \end{array} \right. \end{array} \right. \tag{3.24}$$

The obtained parameter α_n is normalized at each step and the offset b_n is given by hyperplane Equation 3.17 at a chosen support vector $SV_{c,n}$.

$$f_n^t(SV_{c,n}) = 0 \quad \Leftrightarrow \quad b_n^t = \sum_i \alpha_{i,n}^t k(SV_{c,n}, SV_{i,n}^t) \tag{3.25}$$

As a consequence, the parameters of boundary decision function of the winner cluster (α_n, b_n) are updated iteratively as follows:

$$\begin{cases} \alpha_{i,n}^{t+1} \quad \leftarrow \quad \alpha_{i,n}^{t+1} \Big/ \sum\limits_i \alpha_{i,n}^{t+1} \\ b_n^{t+1} = \sum\limits_i \alpha_{i,n}^{t+1} k(SV_{c,n}, SV_{i,n}^t), \quad t > 1 \end{cases} \tag{3.26}$$

At the end of the training stage, we obtain the two previously described sets Ψ^t and F^t (Equation 3.13).

3.4.5.2 Online learning procedure

Creation stage When $card(\Psi_{win}) = 0$ in Equation 3.23, it illustrates that the classifier, considering the current model, detects that the sample should belong to a new class. A new cluster C_{Z+1} is created and the sets Ψ and F are updated correspondingly.

$$\begin{array}{l} \Psi^{t+1} = \Psi^t \cup \{C_{Z+1}^t\} \\ F^{t+1} = F^t \cup \{f_{Z+1}^t\} \end{array} \tag{3.27}$$

Adaptation stage As shown in Equation 3.23 ($card(\Psi_{win}) = 1$), the new data is close enough to only one cluster C_n based on the evaluation of kernel-induced similarity measure function. The parameters (α_n, b_n) of cluster C_n are updated with this new data. The adaptation rule is the same as the one used in the training stage, using Equation 3.24 and 3.26.

Fusion stage When $card(\Psi_{win}) \geq 2$ in Equation 3.23, it shows that there are more than one winner cluster. That is to say, the new data is shared by two or several clusters. Let's keep all winner clusters and their boundary functions in sets Ψ_{win} and F_{win}, respectively, where $\Psi_{win} = \{\bigcup\limits_{win} C_{win}\}$ and $F_{win} = \{\bigcup\limits_{win} f_{win}\}$. This stage aims to fuse those clusters into a unique one (named as cluster C_{fus}):

$$C_{fus} = \{x \in \Psi_{win} | f_{fus} \geq 0\} \tag{3.28}$$

Then the sets of clusters and their boundary functions are updated as:

$$\begin{array}{l} \Psi = (\Psi - \Psi_{win}) \cup \{C_{fus}\} \\ F = (F - F_{win}) \cup \{f_{fus}\} \end{array} \tag{3.29}$$

In the end of the adaptation stage, the parameters of SAKM network are updated correspondingly if some clusters are merged.

Elimination stage In online learning, noisy data possibly occur, which may create many clusters with only few data. In this procedure, the cardinality of each cluster is compared with a threshold Thr_d. Let us define a set $\Psi_{weak} = \{\bigcup_{weak} C_{weak}\}$, where the clusters C_{weak} contain less than Thr_d data. The cluster set Ψ and the boundary function set F are modified as:

$$\begin{aligned} \Psi &= \Psi - \Psi_{weak} \\ F &= F - F_{weak} \end{aligned} \tag{3.30}$$

The data in Ψ_{weak} are then added in the online learning dataset, which will be re-tested in the latter classification procedure after a fixed period of time.

In summary, the whole processing of online SAKM multi-class classification is illustrated by Algorithm 3.

Algorithm 3: Online classification with SAKM

```
Set initialization data $T_N$ and online learning data $X$;
Set all parameters and thresholds;
Train data $T_N$ and build initial SAKM network;
while X is not empty do
    Evaluate kernel similarity function: $Dis_{s,n}$;
    Evaluate kernel similarity criterion: $\Psi^{win}$;
    if $card(\Psi^{win} = 0)$ then Creation procedure;
    else if $card(\Psi^{win} = 1)$ then Adaptation procedure;
    else $card(\Psi^{win} \geq 2)$Fusion procedure;
    if then the number of the clusters is less than thresholds then
        Elimination procedure
    end
end
```

3.4.6 Experimental results of SAKM

3.4.6.1 SAKM for semi-supervised classification

In order to test the performance of SAKM algorithm on novelty detection and classification of persons, a group of experiments are presented in online environment, in which the four previously created feature subsets from CASIA Database are used. In each experiment, only 100 images (about 10%) of three classes ($P1$, $P2$, $P3$) are used for initialization. The online learning data that will be tested are composed of the other data of the three learnt classes and all the data of the four new classes ($P4$, $P5$, $P6$, $P7$). These experimental results are shown in Table 3. For three learnt classes, the classification rates are higher than 95% on four feature subsets. The other four new classes are successfully detected and characterized with higher than 84% classification rate. In the end of the learning procedure, there are more than one cluster for one class, as shown in the last column of Table 3. It is worth noticing that the number of clusters is under the influence of the number of features of the training data, which is illustrated by the comparison of the second and the last column of Table 3.

It is necessary to fuse the different clusters which represent the same persons (classes). Here, the labels of initial learning data are used to merge the different clusters, which are hard to automatically update with the online learning procedure of SAKM. That is why the fuse procedure of the above experiments did not work well. In future work, how to automatically fuse clusters and achieve the goal to have one cluster for one person will have to be addressed.

	Nb features	P1 (%)	P2 (%)	P3 (%)	P4 (%)	P5 (%)	P6 (%)	P7 (%)	Nb Clusters
CASIA-Wholeset	93	100	100	97.48	97.53	95.34	90	100	23
CASIA-PCA	26	100	100	95.27	98.02	88.34	84.33	100	16
CASIA-CFS	40	100	100	99.26	100	96.64	90.83	100	16
CASIA-Wrapper	16	100	100	97.69	100	95.96	89.17	100	9

Table 3.2: Results of online classification with SAKM algorithm (three classes for initial learning and seven classes for online classification).

3.4.6.2 SAKM for online clustering

Several experiments are performed to test the performance of SAKM. The difference with the above experiments is that there is no initial learning procedure in clustering. The process is totally unsupervised, which starts from a new data in online learning dataset X and ends when X is empty. In the first instance, SAKM algorithm is tested on a group of persons only based on color features. We expect to categorize number of persons into several groups, which are based on the color of their clothes. As shown in Figure 3.7, several persons (P1, P4, P5, P7, P13, P16 and P20) in CASIA database are chosen and could be divided into three main groups based on the color of the person's top clothes (Black, White and Blue). The feature subset of the following experiments is only composed of the color features of the top. The first experiment (N1) is dedicated to online clustering of three classes (P4, P13 and P20). The classification accuracy achieves 100% of each class, as shown in the third column of Table 3.3. In experiment N2, two more classes (P5, P7) are added into the clustering dataset, which belong to the group Black and should be classified in the same group as P4. The fourth column of Table 3.3 illustrates that all images in both P5 and P7 are totally classified in the correct group. Similarly, in experiment N3 and N4, two more new classes (P1 and P16) are taken into account (without P5 in experiment N3). The classes included in the groups Black and Blue obtained higher classification accuracy. However, the accuracies of classes in group White are lower. It is worth noticing that the color of the clothes (top) has some differences between the class P13 and P16, as shown in Figure 3.7. Globally, the classification accuracies in Table 3.3 show good performance of SAKM on online clustering based on the basic color features.

Two other experiments (N5 and N6) are designed to test the performance of SAKM on online clustering in a more complex situation: one group contains only one single class. The dataset of experiment N5 is composed of all images in P1, P4, P7, P13, P16 and P20. After changing the initial parameters, the same algorithm (SAKM) is used. All classes are correctly distinguished and the recognition accuracies are shown in Table 3.3. In experiment N6, class P5 is added and SAKM algorithm with the same parameters of ones in experiment N5 is used. Finally, we get four groups instead of the expected seven groups (P4, P5 and P7 are grouped in one cluster, which are colored in Table 3.3). The result of clustering on SAKM algorithm is changing with the online learning data. Because the color features of P4, P5 and P7 are very similar, it is prone to recognize them into the same group and keep all support vectors into this group during online processing. In order to solve this problem, we considered two more experiments.

The dataset of the last two experiments is composed of both color and texture features of the top part. The online clustering data of the experiment N7 are the same of the ones in N5. Based on SAKM algorithm, N7 satisfied to detect each class into a single group. However, if we compare the results of the experiment N5 and N7, the classification accuracies of experiment N5 (based on only color features) are higher than ones of N7, especially for class P13. By checking the confusion matrix, we find that a part of images of P13 are badly classified into the group of P4. That is maybe due to the influence of the texture features. The experiment N8 is similar to N6, but it uses the color and texture features. The results are shown in the last column of Table 3.3, which are not as satisfying as the results of N6, but the experiment N8 satisfied to isolate P7 into a single group. According to these comparisons (N5/N7 and N6/N8), different kinds of features could be helpful to correctly distinguish different classes, especially when some classes have similar characteristics of one kind of feature. Several kinds of features may influence the performance of clustering when one kind of feature could be used to successfully detect each class. Except a good classification algorithm, an efficient feature set is significant and has big influence on the performance of the whole system.

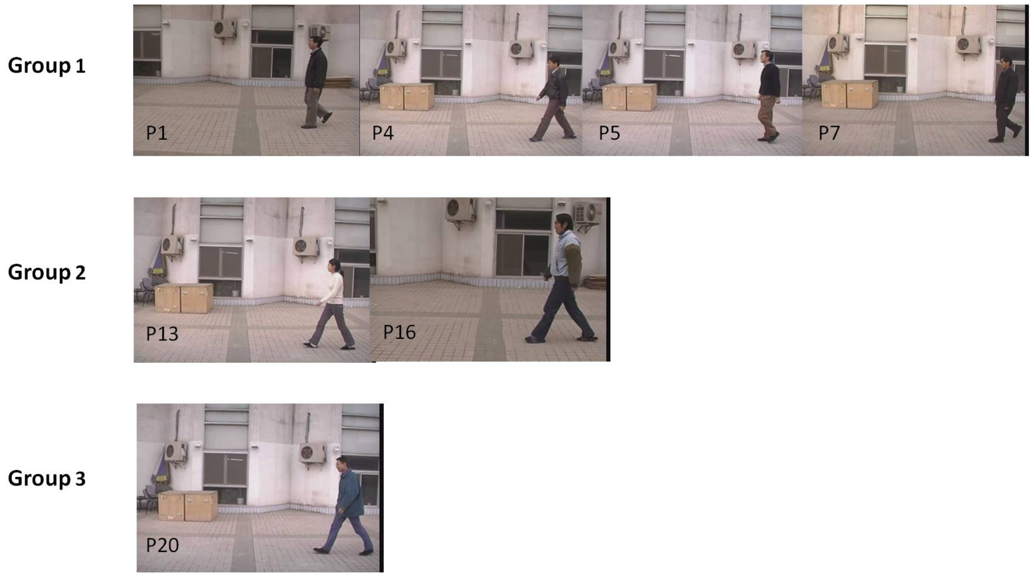


Figure 3.7: Three clusters composed of several persons for onling clustering

Classes	Groups	N1 Accuracy (%)	N2 Accuracy (%)	N3 Accuracy (%)	N4 Accuracy (%)	N5 Accuracy (%)	N6 Accuracy (%)	N7 Accuracy (%)	N8 Accuracy (%)
Features		color	color	color	color	color	color	color+texture	color+texture
P1	Group 1	-	-	95.96	99.79	100	100	97.62	97.62
P4		100	100	100	100	100	100	97.9	98.09
P5		-	100	-	99.43	-	100	-	97.34
P7		-	100	100	100	100	99.9	94.06	94.06
P13	Group 2	100	100	99.7	93.42	100	99.9	**77.37**	**77.56**
P16		-	-	91.79	99.05	99.29	99.52	96.9	97.14
P20	Group 3	100	100	99.15	99.79	100	100	95.45	96.83

Table 3.3: Results of different experiments on online clustering based on SAKM

As a consequence, the experimental results have shown that SAKM algorithm has the ability to differentiate and classify new classes in online multi-category application. However, there are still some problems not yet solved. How to improve the robustness of SAKM algorithm to overcome the problems in realistic application is still a hard task.

3.5 Conclusion

This chapter has presented an online multi-class classification system, in which two online algorithms are separately applied on CASIA database.

MID-SVM algorithm proposed by Boukharouba et al. [BBL09] is firstly used for online human recognition in video surveillance. Compared to classical SVM, MID-SVM is more suitable for the practical situation of surveillance system, as it has the ability for online learning by adaptively updating the classifier model with new information. Experimental results on the four predetermined feature sets have shown the satisfying performance of MID-SVM. It achieves higher than 95% global rate to classify 20 persons in an online setting.

However, MID-SVM algorithm is limited to supervised learning, which can not be applied for novelty classification. SAKM as a clustering approach is applied to detect new persons and managed to achieve online person classification. The experimental results on CASIA database have shown that SAKM algorithm has the ability to detect and classify new classes. However, the performance of SAKM algorithm is not perfect in real-world application. To select efficient features and to improve the performance of SAKM is still a challenge.

The aim of the proposed human recognition system is to achieve automatically person discovery and identification in real-world environments. In this chapter, the experimental results on CASIA Database have demonstrated the satisfying performances of the proposed system based on MID-SVM and SAKM algorithm. However, compared to the conditions in CASIA Database, real-world conditions are more complicated, such as abrupt variation illumination, shadows, etc. In order to evaluate the performance of the proposed system, one more complicated database will be created and used in the next chapter. According to this database, new features will be selected and discussed. Classification related to two different applications will then be tested.

Chapter 4

Applications of video surveillance system

4.1 Introduction

The previous chapters have presented offline and online human recognition systems. The experimental results have shown the satisfying performances of online algorithms on several feature sets for human recognition. This chapter will focus on a realistic human recognition system, which will contain two specific applications. In this system, all experiments will be done based on our new database, which was acquired by two video cameras located in our laboratory.

As it turned out, the goals of video surveillance are different according to the different applications. In [HJLQ11], their aim is to detect the abnormal pedestrian crossing from video processing. In [GSH06, TCKA$^+$10], the goal is to address the person re-identification problem from several cameras with various position and illumination conditions. In [SL13], the goal is to perform person recognition in living environment, in which the changes of person's clothes and posture are considered. In this chapter, two types of video surveillance applications will be presented. One is for online human re-identification in closed indoor environment, such as a company or a security organization. The other is for person discovery in public areas, such as a garden or a train station. In the first application, the number of persons is fixed and the initial information for each person could be obtained and used to initialize the classification model. The goal is to answer this question: is there someone in the video sequence and who is it? The traditional classification methods require quantity of training images to learn a complete model. Besides, they can not incorporate any additional information from new incoming data. It is a challenge to automatically identify persons with a limited number of video frames used for initial learning and reach respectable performance with thousands of images acquired in various conditions (lighting, clothes, pose, etc.). Besides, how to re-identify person from network cameras is another challenging task. Since in online setting, there could have totally different views, backgrounds or lighting conditions.

In public area, it is very hard to achieve person classification based on supervised learning method. As we do not know who will appear in the video surveillance stream and we can not get initial information before he/she appears. How to identify new persons is a current problem. Besides, the system has to have the ability to learn and update with the information of these new persons. The final goal is to achieve automatic person identification in public area without any constraint.

In Section 4.2, we describe the overall architecture of the online video surveillance system and introduce our newly acquired database. Section 4.3 explains how to extract silhouettes of persons from video frames. Appearance features composed by color and texture are then extracted and selected in Section 4.4. Besides, a comparison of RGB and HSV colors will be performed in this section. Experimental results of the proposed methods performed in several applications are described in detail in Section 4.5 and Section 4.6. Section 4.7 concludes this chapter.

4.2 Overall real-time video surveillance system

Before going into details on the experiments for the two applications, we will present the overall proposed real-time video surveillance system. After the presentation of the architecture of the system, a new database is described, which has been created by us and considered as the main experimental data for the following applications.

4.2.1 Overall architecture

The overall architecture of this video surveillance system including silhouette extraction, foreground analysis and classification is depicted in Figure 4.1.

It consists of three main modules:

- The first module consists of background subtraction and silhouette extraction from the video sequences, which will be described in Section 4.3.

- The task of the second module is to find the suitable features to represent the person segmented from the background. It consists of initial feature extraction and feature selection procedures.

- The last module is the classification task, using either MID-SVM or SAKM algorithms depending on the considered applications, which will be described in detail in Section 4.5.

4.2.2 Experimental set-up

The proposed video surveillance system is expected to work in real-world environment. A video database named "IA Database" has been created using two fixed surveillance cameras, Cam1 and Cam2, installed in the corridors of our laboratory, as shown on Figure 4.2.

Cam1 is installed at the top of the hall and is only exposed to natural illumination. It contains 22 persons with a random walk at a normal pace. Each image of the video sequences contains only one person. Some persons walk while carrying objects (such as bag, paper, laptop or a cup of coffee), and no one changed their clothes in the different video sequences. Besides, most of the participants walk with four different orientations: walking in front of and back to the camera (entering and leaving in the hall), walking from right to left and from left to right through the passageway. However, some persons walk more randomly. Sample images of one person seen by Cam1, chosen randomly, are shown on Figure 4.3.

Cam2 is installed at the top of one corridor in our department and works with a lighting illumination (which is also slightly affected by sunlight from glass door). It also contains the same 22 persons as seen by Cam1 with a random walk at normal pace. All persons wear the same clothes as in Cam1. Each person has two video sequences: entering into and leaving from the corridor (walking towards to and back to the camera). Each image of the video sequences contains only one person. Taken the same person shown in Figure 4.3 for example, sample images obtained by Cam2 are illustrated in the first column of Figure 4.4.

The number of the images for each person captured by Cam1 and Cam2 is unbalanced, which is presented in Table 4.1.

IA	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12	P13	P14	P15	P16	P17	P18	P19	P20	P21	P22
Cam1	74	89	109	180	111	94	124	99	93	42	111	76	78	101	52	102	142	21	112	106	93	90
Cam2	60	64	69	76	66	64	56	69	71	65	38	41	60	52	53	64	63	26	59	57	27	59

Table 4.1: Number of the samples in Cam1 and Cam2 for each person

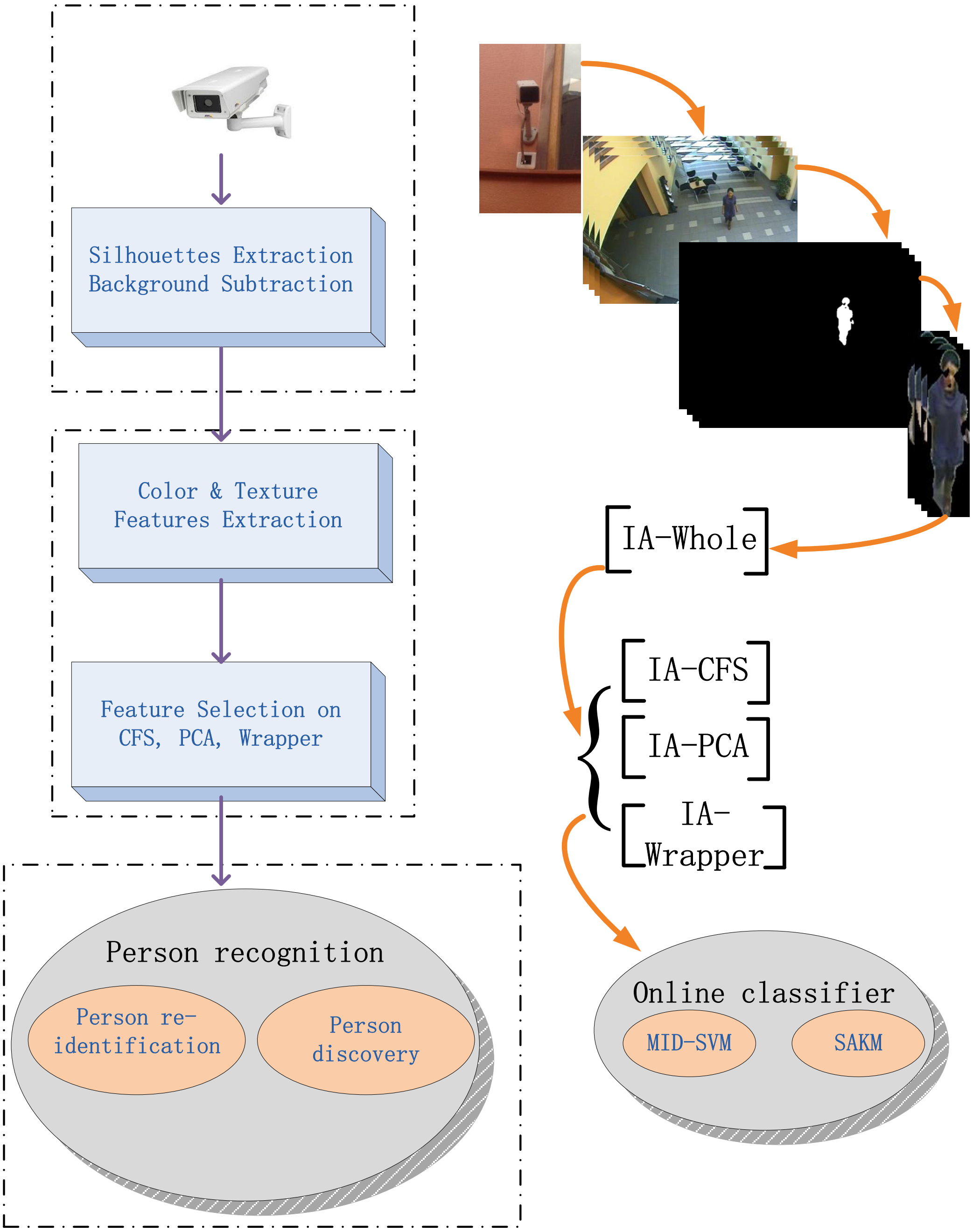


Figure 4.1: Overall architecture of real video surveillance system

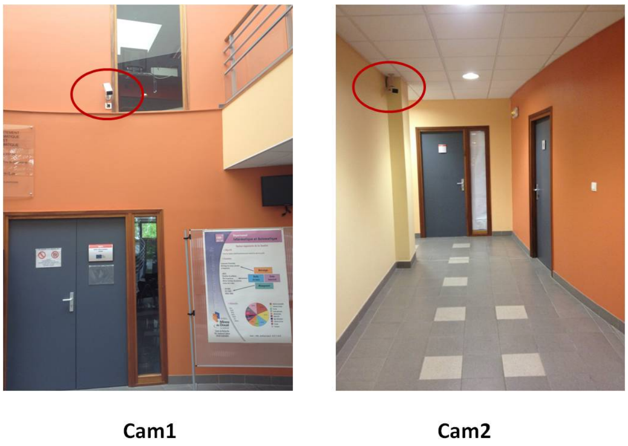


Figure 4.2: Cam1 and Cam2 installed in the lab

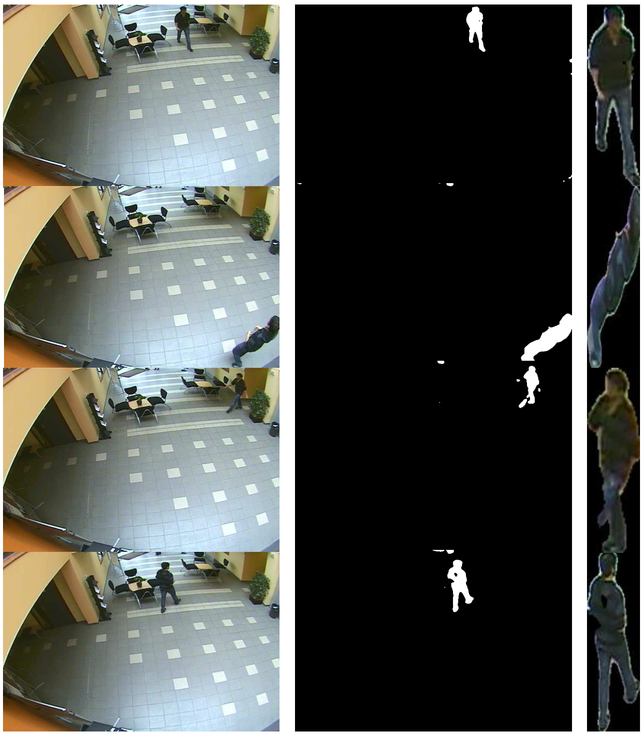


Figure 4.3: Image examples of one person seen by Cam1

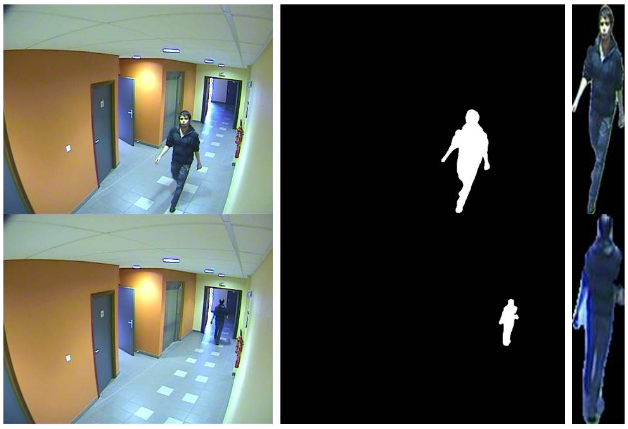


Figure 4.4: Image examples of one person seen by Cam2

4.3 Silhouette extraction

As mentioned in the overall architecture, the first task is to detect the moving objects and to extract the person's silhouette from the background. A very classical method is based on background subtraction techniques. In this work, background subtraction is based on the combined features of Local Binary Patterns (LBP) and photometric invariant color measurements. LBP operator is stable to illumination changes and has advantage in computational simplicity [HP06]. Photometric color features are extracted after transforming the images in the Hue Saturation Value (HSV) space. This color parametrization is more robust to illumination changes and reduces the effect of shadows (disambiguation) that both decrease the accuracy of the silhouette extraction.

Given a pixel x of the image I, the definition of LBP operator can be expressed as follows:

$$LBP_{P,R}(x) = \sum_{i=0}^{P-1} s(I(x) - I(x_i) + n)2^i \tag{4.1}$$

where $I(x)$ and $I(x_i)$ are respectively the grey-level values of the central pixel and P neighbour pixels on a circle of radius R. n is a added noise parameter, which could make LBP operator more static[YO07]. The binary function $s(x)$ is defined as:

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & x < 0 \end{cases}$$

One example of LBP operator is described in Figure 4.5.

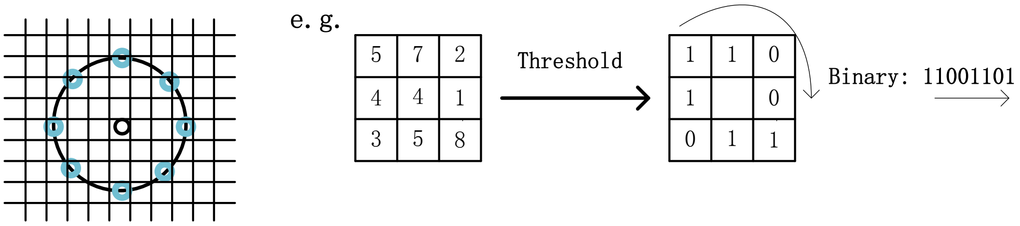


Figure 4.5: Example of the LBP operator

For each pixel, a model M based on local texture and HSV color features is set up, which contains 4 components:

$$M = \{I, I^{\max}, I^{\min}, LBP\} \tag{4.2}$$

where

- I is the average value of the HSV vector (H,S,V).
- $I^{\max}$ and $I^{\min}$ are respectively the estimated the maximal and minimal HSV values of all pixels in this model;
- LBP is the average value of $LBPs$;

The surveillance camera can automatically start to save video sequences when a person enters in the surveillance area and stop when the person leaves. The first frame of each video sequences is the background image, which is used to set up the background model according to Equation 4.2. During the acquisition, persons randomly walk in the hall of the laboratory or across the corridor, where are the surveillance areas, so each created video sequence is short, in which illumination changes are rare. As a consequence, the background model of each sequence is unchanged.

Foreground detection is based on the measurement of the texture and the color distance D [YO07], which could be expressed as the following equation:

$$D(M(x)) = \lambda D_T(LBP(x), LBP(x_c)) + (1-\lambda) D_C(I(x), I(x_c)) \tag{4.3}$$

where $M(x)$ is the statistical model of the pixel x, x_c is an input pixel and $\lambda \in [0,1]$, λ is the ratio indicating the weight of texture and color distance.

The texture distance D_T is calculated as:

$$D_T(LBP(x), LBP(x_c)) = \begin{cases} 0 & \text{if } |LBP(x) - LBP(x_c)| \leq Thr_T \\ 1 & \text{otherwise} \end{cases} \tag{4.4}$$

where Thr_T is a threshold, $Thr_T \in [0,1)$.

The color distance D_C is related to two distances and is defined as follows:

$$D_C(I(x), I(x_c)) = \max\left(D_{angle}(I(x), I(x_c)), D_{range}(I(x), I(x_c))\right) \tag{4.5}$$

The angle distance D_{angle} is computed by:

$$D_{angle}(I(x), I(x_c)) = 1 - e^{-\max(0, \theta - \theta_n)} \tag{4.6}$$

where θ is the angle of $I(x)$ and $I(x_c)$, θ_n is the angle of $I(x_c)$ and $I'(x_c)$, where $I'(x_c) = I(x_c) + I(n)$, n is the noise [YO07].

The range distance D_{range} is defined as:

$$D_{range}(I(x), I(x_c)) = \begin{cases} 0 & if\, I^{low} \leq I(x_c) \leq I^{high} \\ 1 & otherwise \end{cases} \tag{4.7}$$

where I^{low} and I^{high} are shadow and highlight factors, respectively, which are defined as: $I^{low} = \alpha I^{\min}$, and $I^{high} = \min(\beta I^{\max}, \frac{I^{\min}}{\alpha})$. In general, $\alpha \in [0.4, 0.7]$ and $\beta \in [1.1, 1.5]$.

Foreground detection is done according to the above-mentioned combined distance $D(M(x))$. To filter out noise, the bilateral Gaussian filter [PD06] is used, which is defined as: $D^b(M(x)) = \frac{1}{W^b} \sum G_{\sigma_s}(\|x - x_c\|) G_{\sigma_r}(\left\|I_g(x) - I_g(x_c)\right\|) D(M(x))$, where W^b is a normalizing constant, G_σ is a Gaussian kernel, I_p denotes the gray-level image, σ_s and σ_r respectively control the distance that each neighbor decreases in the image plane (the spatial domain S) and on the intensity axis (the range domain R). Finally, when $D^b(M(x)) > Thr_{bg}$, the pixel is considered as a foreground pixel, where Thr_{bg} is the detection threshold.

The following section will present the second module in online person classification, which is the extraction of person appearance associated with three methods of feature selection.

4.4 Feature extraction and selection

In IA database, persons are randomly walking and their orientations compared to the camera are also random, which cause the fact that there are only person's back images in some video sequences. As a consequence, as noticed in Chapter 1 about the different kinds of feature acquisition, appearance-based features are the best ones to represent the samples in comparison with face and gait features, which are obviously less robust here. Color and texture features of person's clothes and their visual body are extracted. The extracted components are similar to ones used in CASIA Gait Database: Mean Value, Standard Deviation and Histogram with

4 beams for grey-world normalization R', G', B'; as well as 13 Haralick Texture features. Similarly, color and texture features are individually extracted from three parts (head, top and bottom) of each person. Finally 93 features (54 color features combined with 39 texture features) are obtained to represent each person.

As described in Section 2.4, an efficient feature selection approach is necessary in order to reduce the computational cost. Three feature selection approaches have been applied: PCA, CFS and Wrapper, which have been described in detail in Chapter 2. Based on these three feature selection methods, the above-mentioned 93 initial features are respectively selected to obtain the best efficient feature subsets. As a consequence, in addition to IA-Whole (the initial feature set), three feature subsets are created for IA Database: IA-PCA, IA-CFS and IA-Wrapper, which are described in Table 4.2 and the details of selected features of each subset are described in Table A.3 in Appendix A.

FeatureSet	FeatureNumber	Description
IA- Wholeset	93	initial color and texture features
IA- PCA	22	linear combinations of original features
IA- CFS	24	2-color-head 9-color-top 7-color-bottom 2-texture-head 3-texture-top 1-texture-bottom
IA- Wrapper	17	2-color-head 5-color-top 5-color-bottom 1-texture-head 3-texture-top 1-texture-bottom

Table 4.2: Four feature subsets for IA Database (RGB color combined with Haralick texture features extracted from three parts of each person)

The following sections will detail the experiments on two different applications based on the proposed video surveillance system. The first focus on person re-identification and the other on new person discovery.

4.5 Person re-identification

In this section, the aim of our video surveillance system is to classify, online, persons in a closed environment, such as in a company, a private club or a security service organization. Before considering several cameras, we will focus on only one. The system should have the ability to do classification with only a few samples for initial learning and adaptively update with the change of the test images. Then, the task will be to re-identify persons appearing in two different cameras, which needs the system that has ability to adapt to different environment (background, illumination conditions, etc.). The following two series of experiments will test the performances of the proposed system, described on Figure 4.1, depending on the mentioned goals.

4.5.1 Online person classification

In this section, a set of experiments is done to test the performance of MID-SVM on online person classification, in which the experimental samples are based on the images captures by Cam1. At the beginning, only a few images of all persons in Cam1 are used to initialize the MID-SVM classifier. The goal of the classifier is to recognize the person who appears in Cam1.

The first experiment (named Experiment 1) is done, for each one of the 22 persons, using only the first 10 % video frames of each for initial learning and the other frames are used for online classification using MID-SVM algorithm. The experimental results on the four feature subsets are described in Table 4.3. The global classification rates on the four feature subsets are similar and higher than 83%. The IA-CFS feature subset obtains the best results with 88% global classification rate. The three feature subsets with selected features achieve better performances than the initial features (IA-Whole). Even though IA-Whole feature set contains much more features than the others, the results prove that the selected features which keep the main discriminative characteristics are more robust. The recognition rates of each class which are lower than 80% are marked in bold text. Several persons are recognized with a very low classification rate, such as P4, P5, P6 and P16. Such unsatisfying results could be due to the features (which maybe do not well represent these persons) or an important difference between the initial learning samples and the online classification ones.

In order to explain the reason of such low recognition rate and improve the performance of the person recognition system, two complementary experiments are tested. In Experiment 1, the first 10% samples of each person are chosen for initial learning, which only include a small part of states of person and environment, such as a person walking towards the camera in a constant illumination condition. The remaining (the other 90% samples) contains much more changes of persons and environments. In Experiment 2, we randomly choose the 10% images of each person instead of the first 10% ones, the remains of images are used for online classification. MID-SVM algorithm is also used and all parameters are the same to the ones in Experiment 1. It also shows that the results between subjects are more stable in one of the case. The comparison of experimental results between

	IA-Wholeset	IA-PCA	IA-CFS	IA-Wrapper
σ	9	5	4	3
P1	89.68	91.47	91.51	90.47
P2	85.58	87.94	89.77	84.35
P3	84.86	84.39	**74.72**	**76.56**
P4	**57.73**	**73.90**	85.73	**65.62**
P5	**65.31**	**74.02**	**71.23**	**71.41**
P6	**56.94**	**63.27**	**66.57**	**70.35**
P7	**71.87**	**74.83**	**75.59**	86.46
P8	94.74	92.68	97.61	84.46
P9	88.30	86.77	94.20	93.78
P10	88.13	89.01	97.50	91.11
P11	90.12	88.45	89.43	88.49
P12	92.36	88.20	98.57	98.54
P13	97.91	97.84	99.47	96.81
P14	85.01	81.27	92.18	88.09
P15	96.07	96.51	98.84	96.20
P16	**52.83**	**66.32**	**70.35**	**66.27**
P17	91.00	92.08	91.16	90.14
P18	96.84	96.07	100.00	99.95
P19	93.82	89.62	97.14	97.13
P20	87.17	82.29	82.07	**70.38**
P21	82.84	84.73	93.77	84.74
P22	92.62	91.78	91.96	87.86
Global	83.72	85.15	88.61	85.42

Table 4.3: (Experiment 1) Classification rates of MID-SVM algorithm for the 22 persons with an optimised σ value ($step = 0.1$)

Experiment 1 and 2 are described by Figure 4.6, which shows that in Experiment 2 the global classification rates are much higher, as well as the classification rate of each person with each feature subset has been improved. As a consequence, the proposed four feature subsets contain the discriminative attributes of persons and are useful in such human recognition system.

Another complementary experiment (Experiment 3) is done to test the performance with different sizes of the initial learning sets. Similar to Experiment 1, the first part of samples of each person are chosen for initial learning. The ratio of initial learning samples ranges from 10% to 90%. All parameters of MID-SVM classifier are the same to the ones in Experiment 1. The relation between initial learning sample size and classification accuracy is indicated by Figure 4.7. When the number of initial learning samples raises, the global classification rates of the four feature subsets correspondingly increase. It is easy to understand that the initial learning samples contain more possible changes of persons and environmental conditions when the ratio is increased. Besides, we notice that the performances of the four feature subsets are closed when the ratio achieves 70%, which shows that the superiority of selected feature subsets decreases with the initial learning sample size increasing. Note that the global classification rates are higher than 93% when only 10% samples of each person are randomly chosen for initial learning in Experiment 2, shown in the first item of Figure 4.6. However, when the initial learning samples are chosen in order, the comparable classification rates are achieved when the ratio is larger than 70%.

4.5.2 Online person re-identification

The proposed surveillance system is expected to re-identify persons from network cameras which work in different environmental conditions. The flowchart of person re-identification is illustrated by Figure 4.8, in which the MID-SVM algorithm is used for classification. The experiment (Experiment 4) is done using IA Database, using the two video cameras. It is noticed in Figure 4.3 and Figure 4.4 that the color appearances (color information) of the candidate vary with the camera. When one person enters the hall, Cam1 acquires and shares the video sequences by network, which are used to online initialize the MID-SVM classifier. Then when this person walks through the corridor where Cam2 is installed, the classifier updates with the video sequences from Cam2 and recognize this person. It also happens that one person walks through the corridor firstly and then leave the hall of the department. These two situations are tested in the following experiments.

First, all video sequences in Cam1 are chosen for initial learning and all video sequences in Cam2 are used for learning and determined updating. All the images are segmented based on the silhouette extraction approach and color and texture

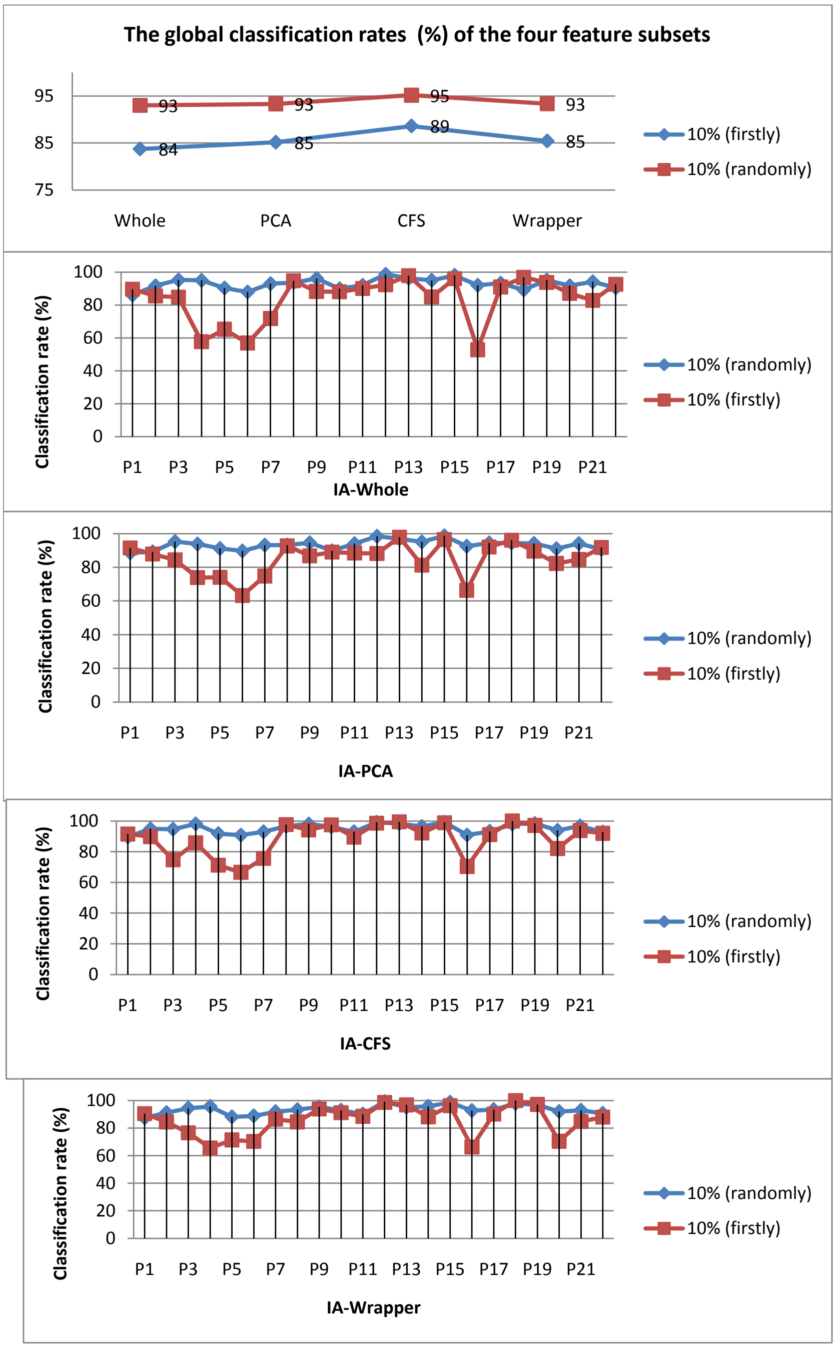


Figure 4.6: (Experiment 2) The comparison of classification rates in Experiment 1 and 2

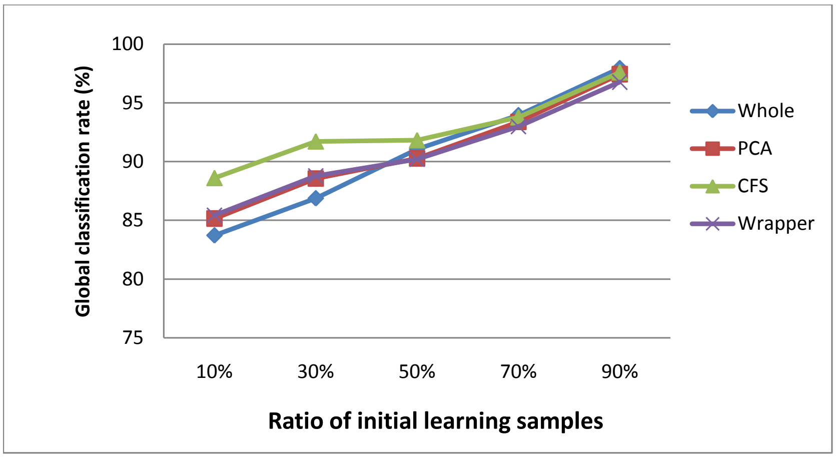


Figure 4.7: (Experiment 3) The relation between the initial learning sample size and classification accuracy

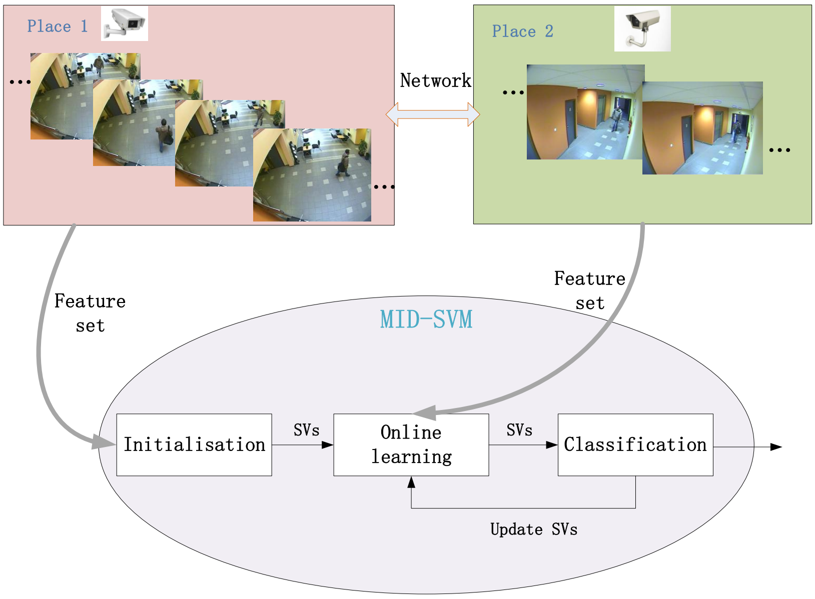


Figure 4.8: The flowchart of person re-identification

features are extracted. Both initial learning procedure and online classification stage are based on MID-SVM algorithm presented in Chapter 3.

Table 4.4 shows the classification performance of MID-SVM on each person. The experimental results shown that three selected feature subsets performed very well (with higher than 97% global classification rates). By comparison, IA-Whole obtained a little worse result with 94% of global classification rate. It is worth noticing that except the class P18, the other classes are successfully re-identified by MID-SVM algorithm for all the four feature subsets. We notice that the total number of the samples of class P18 captured by Cam1 and Cam2 is 21 and 26, respectively, as shown on Table 4.1. By checking in detail, the person (class) P18 carries a white paper in all video frames. In Cam1, all samples of class P18 used for initial learning are frontal view images, however in Cam2 the ones used for online classification contain both frontal and back view images (the number is 10 and 16, respectively). The example images of class P18 captured by Cam1 and Cam2 are shown in Figure 4.9.

	IA-Wholeset	IA-PCA	IA-CFS	IA-Wrapper
σ	10	5	5	5
P1	98.58	99.84	99.78	99.47
P2	98.06	99.70	99.81	99.97
P3	98.16	99.09	98.41	99.67
P4	94.78	99.04	99.69	99.26
P5	90.66	97.28	97.66	99.12
P6	93.92	97.79	98.47	98.75
P7	89.35	94.65	96.24	99.60
P8	94.95	99.36	96.23	97.09
P9	96.60	98.44	98.31	99.04
P10	98.31	99.73	98.78	98.47
P11	95.19	98.06	99.52	100.00
P12	82.90	93.80	95.91	91.83
P13	98.46	99.28	99.29	99.42
P14	95.42	99.97	99.52	98.65
P15	95.27	95.86	98.26	98.91
P16	98.23	99.77	99.96	98.70
P17	97.84	99.20	99.64	99.35
P18	**65.99**	81.19	**66.94**	**64.25**
P19	97.35	99.60	99.44	99.91
P20	94.91	98.72	99.52	99.83
P21	99.91	99.88	99.96	99.86
P22	96.79	99.06	98.27	99.30
Global	94.17	97.70	97.26	97.29

Table 4.4: (Experiment 4) Classification rates of MID-SVM algorithm on person re-identification for 22 persons (Samples in Cam1 for initial learning, $step = 0.1$).

In order to show the performance on re-identification of the proposed algorithm, an additional experiment (Experiment 5) has been done by exchanging the initial learning samples and the online learning samples. The images captured by

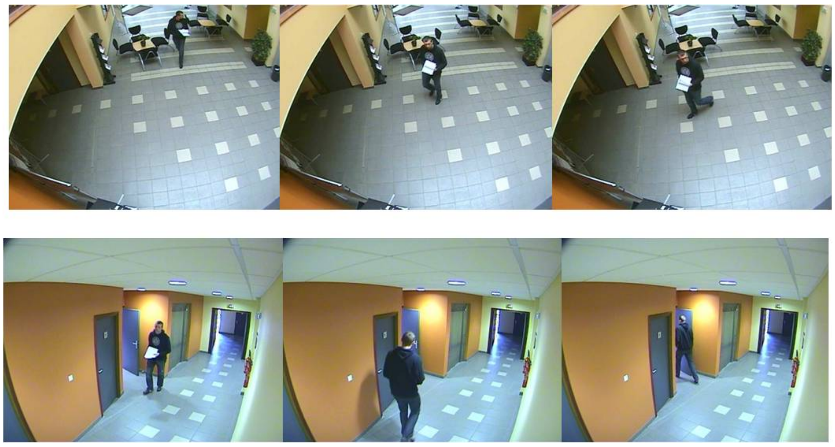


Figure 4.9: Example images of class P18 captured by Cam1 (upper part) and Cam2 (lower part)

Cam2 are used for initial learning and the images of Cam1 are used for online classification. As shown in Table 4.1, the number of initial learning samples is less than the one for online learning. The experimental results are described in Table 4.5. The global classification rates on the four proposed feature subsets are higher than 96%, which are a little higher than the results of the previous experiment (Experiment 4). The difference results between Experiment 4 and 5 may be due to the intrinsic attributes of the online classification data. In Experiment 4, Cam2 data are used for online recognition, however, Cam1 data are used in Experiment 5. Then we add another experiment (Experiment 6) to test if Cam2 data are more hard to be classified. Similar to Experiment 1 and 2, 10% samples of Cam2 data are selected for initial learning sequentially and randomly, respectively. The remaining samples are for online classification. MID-SVM algorithm is used on both initial learning and online classification processes. The experimental results are illustrated by Figure 4.10. The best performance is obtained by IA-CFS when initial learning data are selected randomly. Comparing the results of Experiment 2 and 6, shown by Figure 4.6 and 4.10, the classification rates of online recognition on Cam1 data are higher than ones on Cam2 data. It proves that Cam2 data are harder to classify, which may be due to the illumination conditions.

4.5.3 Studies on new feature subsets

Since two cameras are installed in the high place of the building, top-view images in video sequences are inevitable. The top-view images are obtained when the person is below the area of surveillance video, in which maybe only the head and

	IA-Wholeset	IA-PCA	IA-CFS	IA-Wrapper
σ	10	5	5	5
P1	98.33	99.76	99.98	99.99
P2	95.01	98.45	99.10	99.81
P3	97.79	98.24	99.02	99.69
P4	96.96	95.68	99.88	98.75
P5	98.49	99.59	99.52	99.84
P6	97.93	99.41	98.81	99.10
P7	88.79	95.54	96.99	93.09
P8	99.40	99.97	99.98	99.71
P9	93.64	96.24	99.10	99.10
P10	99.87	99.90	99.87	99.74
P11	95.00	97.50	96.90	98.96
P12	99.99	99.98	99.35	99.62
P13	99.97	98.72	100	99.94
P14	99.99	100.00	100.00	100.00
P15	94.54	100.00	99.93	99.68
P16	99.67	99.64	99.58	99.45
P17	98.09	99.56	98.67	98.73
P18	99.41	100.00	99.98	99.84
P19	**79.66**	96.12	95.26	90.48
P20	99.53	99.45	100.00	99.29
P21	96.11	98.15	93.64	97.45
P22	96.30	95.86	98.43	96.62
Global	96.54	98.56	98.81	98.58

Table 4.5: (Experiment 5) Classification rates of MID-SVM algorithm on person re-identification for 22 persons (Samples in Cam2 for initial learning, $step = 0.1$).

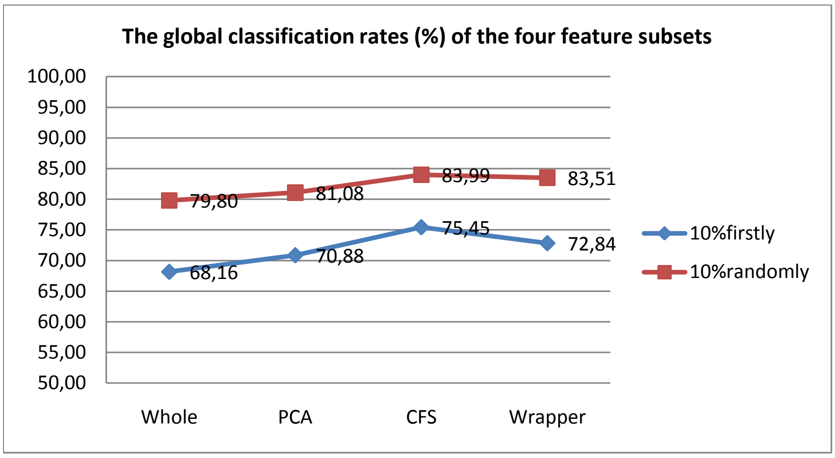


Figure 4.10: (Experiment 6) Global classification rates of person classification on Cam2 samples

a part of clothe are visible. We considered to process each silhouette as a unique part, then select color and texture features from it.

Since the proposed human recognition system is working in real-life environment, we try to extract color features which are more robust to light changes and in which it is easier to remove shadows. HSV color space could be a better choice than RGB, which separates color components from intensity. Similarly, color features: Mean Value, Standard Deviation and Histogram with 4 beams for grey-world normalization H', S', V'; as well as 13 Haralick Texture features are extracted from the whole body of each video frame. Finally 93 features (54 HSV color features combined with 39 texture features) are obtained to represent each person.

In order to analyse the performance of HSV color and RGB color features, a comparison experiment (Experiment 7) has been done. First of all, another initial feature set is created by extracting RGB color and Haralick texture features from the unique part of persons. Taken Cam1 data for example, 10 video frames of each person are chosen for initial learning and the others are for online recognition. The initial features are based on RGB and HSV color space respectively. The proposed MID-SVM algorithm is applied for this online classification. The experiments are tested with varying values of kernel parameter σ to find the best performance. The results are illustrated in Figure 4.11, which show that features based on HSV always achieve higher classification rates with the value of σ changing. From the experiments, we get the best performances of HSV color space with 97.69% global classification rate when σ is equal to 3 and RGB with 95.41% global rate when σ is 10. Figure 4.12 illustrates the classification rate of each person when RGB and HSV get the best performance respectively. The results show that almost each person is classified better when HSV colors are used. To investigate more HSV color space, color and texture are extracted and set up the initial feature subset named as IA-Whole*. Then three selected feature subsets are created: IA-PCA*, IA-CFS* and IA-Wrapper*, which are described in Table 4.6 and the details of selected features of each subset are described in Table A.3 in Appendix A.

4.5.3.1 Person classification

Experiment 8 aims to test the performance of online person classification on new feature subsets, which contain HSV color and Haralick Texture features of each person whole part. Similarly to Experiment 1, MID-SVM algorithm is applied to online classified Cam1 data and the first 10% samples of each person are used for initial learning the others are used for online classification. The experimental results on the four feature subsets are described in Table 4.7. The global classification rates are similar and higher than 95% on the four feature subsets. IA-Whole*

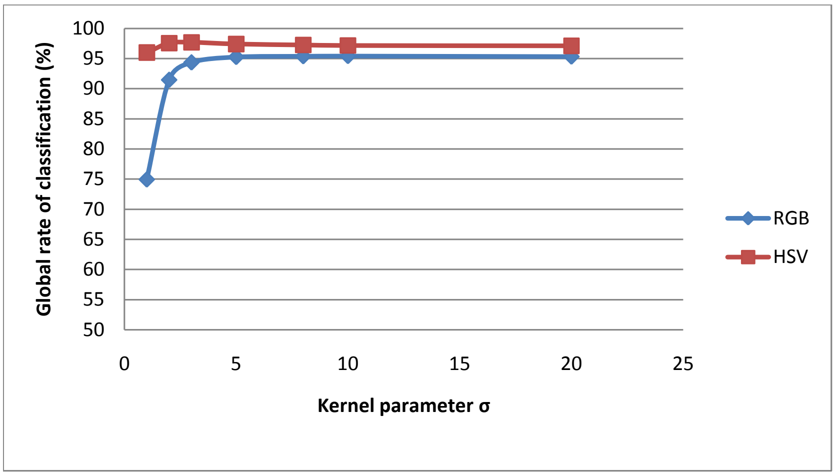


Figure 4.11: (Experiment 7) Performance comparison of RGB-based and HSV-based on varying values of kernel parameter σ

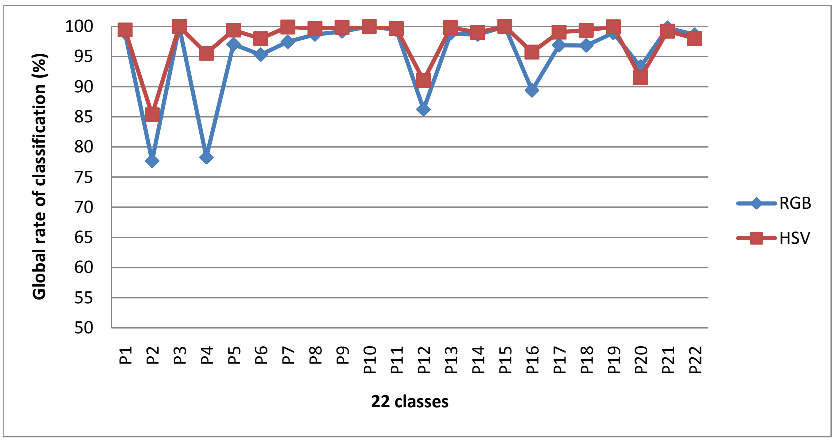


Figure 4.12: (Experiment 7) Comparison of classification rates for the 22 persons between RGB-based and HSV-based with an optimised σ value

FeatureSet	FeatureNumber	Description
IA-Wholeset*	31	initial color and texture features
IA-PCA*	9	linear combinations of original features
IA-CFS*	11	std_ s std_ v hist_ h_ beam4 hist_ h_ beam1 correlation entropy inverse_ diff sum_ var sum_ entropy info_ corr_ 1 info_ corr_ 2
IA-Wrapper*	7	mean_ h std_h mean_s mean_v inertia sum_avg diff_ var

Table 4.6: The new (HSV-based) features selected for 22 persons in IA Database

subset obtained the best result in this experiment. Even if the other three subsets (IA-PCA*, IA-CFS* and IA-Wrapper*) show hardly worse performances than IA-Whole, they contain less features and could save computational time. The recognition rates of each class which are lower than 80% are marked in bold text. Except several persons with low classification rate, such as P2 in IA-CFS subset and P4 in IA-Wrapper subset, the others are well classified. Compare to the result of Experiment 1, shown in Table 4.3, the classification rates in Experiment 7 are higher. It proves that the features extracted from the whole body based on HSV color combined Haralick texture can better represent persons captured by Cam1.

4.5.3.2 Person re-identification

Similarly to Experiment 5, in Experiment 9, all Cam2 data are used for initial learning and all Cam1 data are for online re-identification by the MID-SVM classifier. The experimental results on four proposed data subsets with varying values of kernel parameter σ are described in Figure 4.13. IA-Wrapper* subset obtained the best result, IA-PCA* subset performed as well as IA-Wrapper*, however, IA-Whole* subset obtained the lowest classification rate. It is because person images captured by two different cameras with totally different background, illumination condition, selected features are comparatively more strong and robust. Then we

	IA-Wholeset*	IA-PCA*	IA-CFS*	IA-Wrapper*
σ	3	2	2	1
P1	99.41	99.15	99.38	96.58
P2	85.35	86.78	**77.82**	92.35
P3	99.99	99.3	96.97	99.25
P4	95.54	81.1	94.3	**70.74**
P5	99.35	98.09	98.5	98.76
P6	97.95	94.25	97.93	91.68
P7	99.89	99.18	99.16	98.75
P8	99.65	99.42	98.25	95.22
P9	99.84	98.47	99.65	99.89
P10	100	99.97	100	98.12
P11	99.64	95.52	99.74	95.46
P12	91.02	93.14	82.73	99.93
P13	99.8	97.61	99.92	97.93
P14	98.97	99.61	92.14	98.45
P15	100	99.46	100	99.82
P16	95.73	90.26	93.41	89.97
P17	99.03	98.5	96.3	96.75
P18	99.36	98.49	98.4	98.68
P19	99.92	99.83	99.6	99.62
P20	91.5	91.76	97.01	92.99
P21	99.21	99.32	93.16	91.4
P22	97.98	98.57	95.96	95.6
Global	97.69	96.08	95.92	95.32

Table 4.7: (Experiment 8) Classification rates of MID-SVM algorithm on person re-identification (Cam1 data for initial learning, $step = 0.1$).

compare the results of Experiment 5 and 9, considering the comparison of Table 4.5 and 4.8, the new feature subsets do not achieve better results. It shows that the segmentation of the body in three different parts can improve the robustness of features.

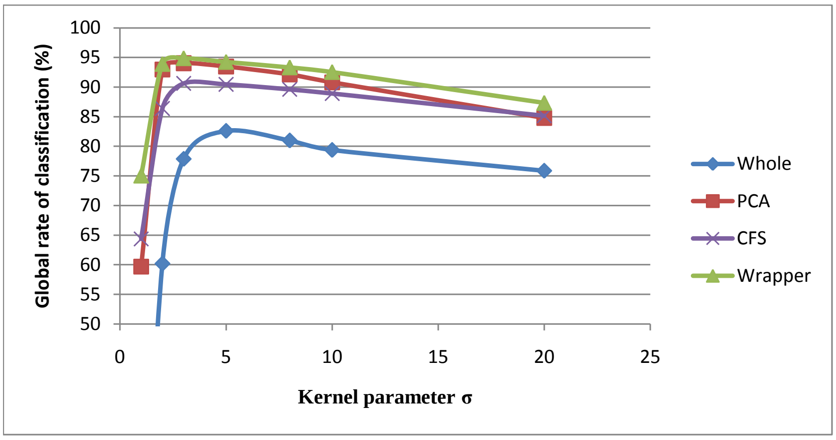


Figure 4.13: (Experiment 9) Performance comparison of the four feature subsets on person re-identification with varying values of σ

4.6 Person discovery

In a public area, it is hard to obtain the information of all random persons. It seems that the supervised classification methods are rarely possible to achieve the goal of new persons identification. SAKM algorithm presented in Section 3.4 has the ability to detect and classify new persons in an online setting. The flowchart of person discovery based on SAKM algorithm is described by Figure 4.14.

In this section, an experiment (Experiment 10) is done to test the performance of SAKM algorithm on online person clustering in the hall of our department. In these experiments, no initial information of any person is learnt before online clustering.

The Cam1 images of 22 persons are chosen in this experiment. The new feature subsets (IA-Whole*, IA-PCA*, IA-CFS* and IA-Wrapper*) are considered as the inputs of SAKM classifier. The experimental results of the four feature subsets are described in Table 4.9. The global recognition rates of all subsets are higher than 90%. It is worth noticing that the value of Thr is chosen differently for IA-Whole* subset, (Thr is an acceptance threshold of kernel-induced similarity measure). In the beginning, IA-Whole* subset was also tested with the

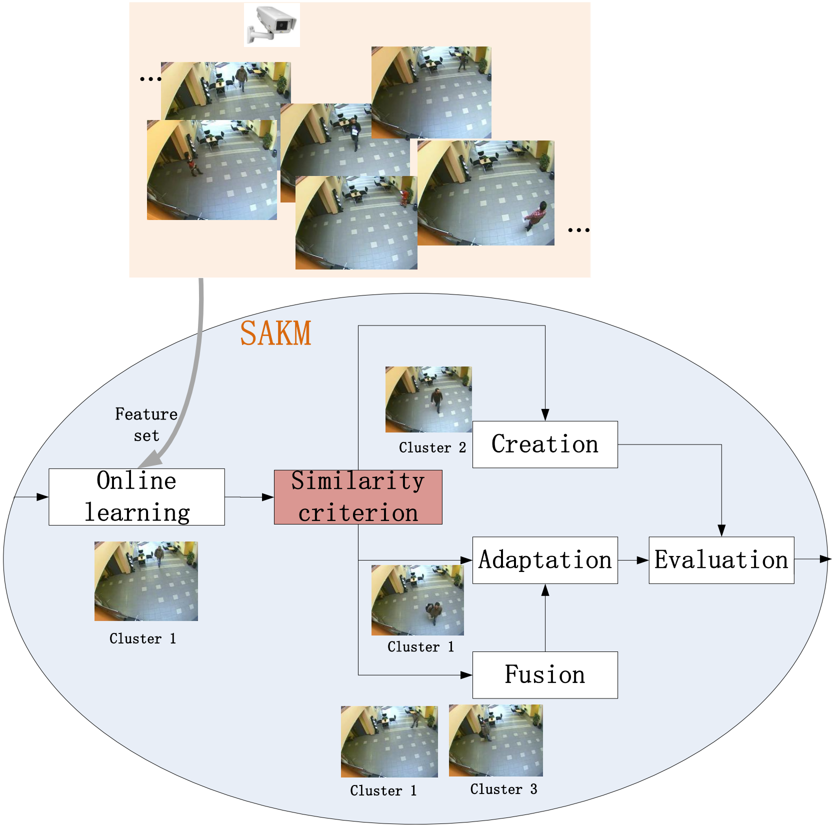


Figure 4.14: The flowchart of person discovery

	IA-Wholeset	IA-PCA	IA-CFS	IA-Wrapper
σ	5	3	3	3
P1	92.55	98.85	99.7	95.05
P2	97.12	98.52	95.39	99.44
P3	**63.01**	89.59	**63.44**	95.88
P4	**71.63**	90.4	89.84	89.96
P5	97.95	97.83	99.52	99.99
P6	**62.91**	91.72	86.41	93.78
P7	83.88	95.45	97.89	94.94
P8	99.81	99.99	99.97	100
P9	**33.77**	98.86	45.03	99.57
P10	92.32	96.4	98.92	95.25
P11	**56.02**	78.18	83.69	80.94
P12	89.44	68.14	99.89	93.84
P13	99.97	98.72	100	99.94
P14	**50.02**	99.43	**57.09**	91.49
P15	88.06	89.06	96.57	80.38
P16	98.93	99.3	99.15	99.85
P17	92.94	99.08	96.86	99.66
P18	93.42	92.99	99.71	83.2
P19	96.22	97.28	99.16	98.85
P20	85.07	92.5	99.78	95.46
P21	86.46	96.46	99.38	99.54
P22	85.42	100	86.29	99.99
Global	82.59	94.03	90.62	94.87

Table 4.8: (Experiment 9) Classification rates of MID-SVM algorithm on person re-identification (Cam2 data for initial learning, $step = 0.5$).

same value of Thr ($Thr = 0.86$), but the performance is much worse with 80.27% global rate and 22 classes are recognized into 37 clusters. When changing the suitable value of the threshold Thr, it performs better, however, it gives more clusters for 22 classes than the other three subsets. As proved in Section 3.4, the number of obtained clusters is under the influence of the number of the features of online clustering data. The classification performance for each person varies a lot, as shown in Table 4.9, there are two extremes of recognition rate (0% and 100%), such as P18 and P21. It is because that online clustering procedure based on SAKM algorithm updates step by step according to the realistic data. During the process of clustering, some wrong support vectors are inevitably created, the wrong would decrease if these wrong support vectors are updated as non support vectors and decision function is corresponding changed; on the contrary, the wrong would continuously increase and bring wide influence of the recognition rate. Especially, the number of P18 is only 21, which is much less than the others. When one wrong support vector happens, there could be not enough updating step to correct the wrong decision function. By checking the confusion matrix of IA-Wrapper subset, all images of P21 are totally wrong classified as P20. The kernel similarity distance in Hilbert space of P21 could be too close with one of P20.

There are some images of P20 and P21 in Figure 4.15, which also look similar.

	IA-Wholeset*	IA-CFS*	IA-PCA*	IA-Wrapper*
Thr	0.98	0.86	0.86	0.86
P1	98.65	100	100	100
P2	98.88	100	100	100
P3	100	100	100	100
P4	99.44	100	100	97.78
P5	100	99.1	100	100
P6	92.55	91.49	96.81	100
P7	97.58	100	100	100
P8	100	100	100	100
P9	100	97.85	100	100
P10	100	100	100	100
P11	93.69	98.2	100	100
P12	94.74	82.89	97.37	100
P13	85.9	83.33	100	100
P14	99.01	100	100	100
P15	100	96.15	98.08	100
P16	100	100	100	100
P17	91.55	98.59	100	100
P18	95.24	**0**	100	100
P19	94.64	100	100	100
P20	86.79	96.23	100	100
P21	98.92	90.32	100	**0**
P22	86.67	86.67	97.78	100
Global Rate	96.10	91.75	99.55	95.35
Nb Clusters	28	23	25	22

Table 4.9: (Experiment 10) Recognition rate of SAKM algorithm on person discovery and recognition

4.7 Conclusion

Video surveillance systems are widely used in various domain with multiple goals. In this chapter, the real-time video surveillance system has focused on the automatic person classification. We have presented two kinds of applications of the proposed real-time video surveillance system: person re-identification and person discovery. In the first application, it is possible to obtain some information of all candidates to initialize the classification model . However, in the second one, we can not obtain any initial information (even a few video frames) of all persons who possibly appear in the area monitored by the video cameras.

In this chapter, all experimental video files have been captured by two cameras which work in two different environmental conditions. Then candidates images are segmented by the silhouette extraction method. Firstly, all person images are

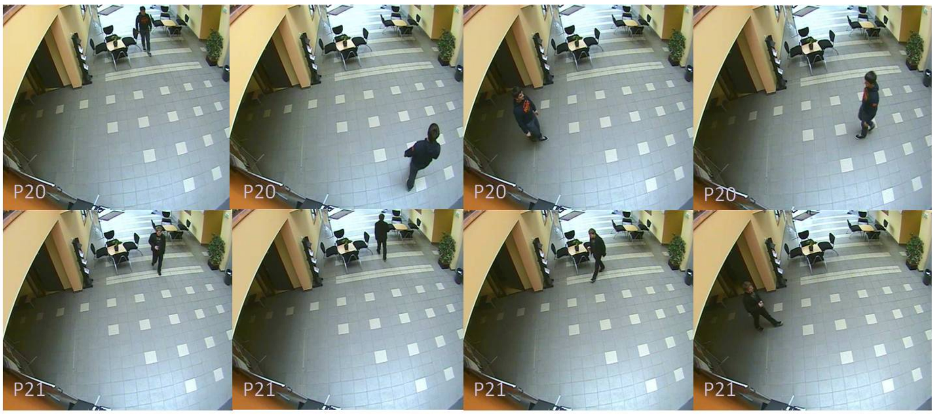


Figure 4.15: Example images of P20 and P21 captured by Cam1

segmented into three parts and initial features are extracted from each part separately, which we have done the same treatment for CASIA Database in Chapter 2. As a consequence, based on the three feature selection methods (CFS, PCA and Wrapper), four feature subsets are obtained: IA-Whole, IA-CFS, IA-PCA and IA-Wrapper. Then a new feature subset is created for comparison, which are extracted HSV color and Haralick texture features from the whole body. The comparison of the experimental results shows that the segmentation of the body in three parts can improve the robustness of the features.

According to the application of person re-identification, two series of experiments have been done. The first is to test the performance of MID-SVM algorithm on online classification in the proposed system. A set of experiments have been done and compared, the results have shown that the classification accuracy varies with the size and the order of the initial learning samples. Another set of experiments have proven the good performance on person re-identification based on MID-SVM algorithm. The application for unknown number of classes has focused on online human clustering. The SAKM algorithm is used for person discovery in online setting. All the four proposed feature subsets have achieved satisfying recognition rates (more than 90%).

The results different experiments have shown the satisfying performances of MID-SVM and SAKM algorithms applied in the proposed system. The proposed system has the ability to achieve not only online person re-identification but also person discovery, which could be used in closed indoor environment or public area.

Conclusion and perspectives

Conclusion of the thesis

Human classification is a very important and active domain of research, which have been widely employed in numerous applications, especially in security surveillance and forensics. According to the specific tasks, lots of works have been developed to address different problems of such system.

The objective of this thesis is to design a complete online human recognition system for automatic person classification and new person discovery based on the basic color and texture features extracted from person appearance. To validate the performance of the proposed system, two databases have been used: CASIA Database (publicly available) and IA Database (newly created). The latter is more complicated, as it has been acquired with two cameras, including more changes in illumination, pose and location. To achieve our objective, each module of human recognition system have been improved. Figure 4.16 shows the outline of our appearance-based human recognition system.

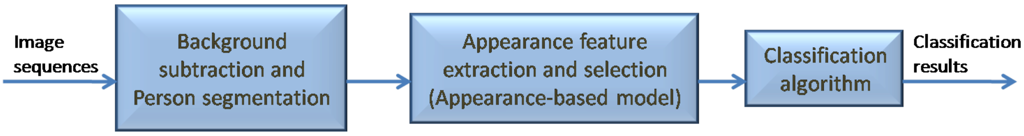


Image sequences
Background subtraction and Person segmentation
Appearance feature extraction and selection (Appearance-based model)
Classification algorithm
Classification results

Figure 4.16: General architecture of appearance-based human recognition system

The first module consists of background subtraction and silhouette extraction for the entering image sequences. To address these matters, we create a feature model combining LBP operator and photometric invariant HSV color features, which is more robust to illumination changes. Foreground detection is based on the measurement of the texture and the color distance in this model. In order to get more accurate appearance information of the persons, we propose a segmentation method, in which three parts of the body (head, top and bottom) have been considered and processed independently. In Chapter 4, we compared the performances for 22 persons recognition on IA Database when we process the image modelling the whole body or three segmented parts. The experimental results have proved the better performance of the proposed method consisting of three-parts segmentation.

The second module is composed of extracting initial features and selecting discriminative features. Considering the goal of our system and the specificity of the databases that we used, we decided to extract human features from the appearance. The proposed initial feature set consists of basic RGB color and Haralick texture features extracted from the three segmented parts of each person. To reduce the effect of illumination changes, grey-world normalization is applied for RGB color features. For each image, the initial feature set contains 54 normalized color features and 39 Haralick texture features, which inevitably include irrelevant or redundant features. To obtain more efficient and discriminative ones, we propose and compare three feature selection methods (PCA, CFS and Wrapper). Four feature subsets are finally created (the initial feature set and the other three obtained by different feature selection methods), which are the inputs of the following classifiers. The experiments of two classical multi-category SVM algorithms (one-against-one SVM and all-against-all SVM) in Chapter 2 have achieved good classification rates on all the four feature subsets. It proves that the proposed color combined with texture features success in person representation and the three feature selection methods could effectively keep the important discriminable information while reducing dimensionality. In the comparison of all experimental results, we notice that the features selected by CFS method are more robust than the others.

The final module of the proposed system is online human classification, in which MID-SVM (classification algorithm) and SAKM (clustering algorithm) have been employed in the proposed online system. To our knowledge, such dynamical classification algorithms have never been used for human recognition in real-world video surveillance. The proposed MID-SVM algorithm has the ability for online learning by adaptively updating the classification model with new information. In Chapter 3, only 5% of the images of each class are used for initial learning, the others are used for online learning and adapting the classification model. The experimental results on CASIA Database have shown the satisfying performance of MID-SVM by achieving higher than 95% of global recognition rate to classify 20 persons in an online setting. Since MID-SVM algorithm is limited to supervised learning with fixed number of classes, it can not be applied for novelty classification. We introduce SAKM algorithm, as a clustering approach, to detect online new persons and classify them. The experimental results on CASIA database in Chapter 3 have shown that SAKM algorithm has the ability to detect and classify new classes.

In Chapter 4, two kinds of applications (person re-identification and new person discovery) of the proposed real-time human recognition system are discussed.

In the application of person re-identification, the number of classes (known persons) are fixed. In IA Database, video sequences have been captured by two cameras that are installed in different locations and work in different illumination conditions. Using MID-SVM algorithm, all image sequences acquired by Cam1 are for initial learning and the ones obtained by Cam2 are used for online recognition. The experimental results have proven the good performance of person re-identification with greater than 96% global classification rate for 22 persons. In the application of person discovery, the number of classes is unknown and no initial information could be obtained on all the persons who possibly appear in the area monitored by the video cameras. The unsupervised learning SAKM algorithm is used to detect new person and achieve person clustering in online setting. All the four proposed feature subsets have succeed in person discovery and recognize with more than 90% accuracy rate.

As a consequence, the proposed system is a complete human recognition system, which contains automatic background subtraction, person representation based on extracted features and online classification. It has the ability not only to detect persons from video sequences and automatically recognize them, but also to re-identify persons from different video sequences captured by more than one cameras in various environmental conditions. Besides, especially, it also has the ability to differentiate new persons and classify them in real-life environments.

Next, future works will be discussed according to the different aspects that can be further improved to achieve an optimal human recognition system.

Perspectives

There could be several aspects in which the proposed human recognition system can be improved. Extensions of this work can be considered from feature selection to classification and are described in the next parts.

- The proposed body segmentation method is very useful in CASIA Database, however, it is not helpful for all images in IA Database. The segmentation method separates the body into three parts (head, top and bottom) based on the ratio of each part of each person. In IA Database, the camera is installed in the top, some image are captured when the person is below the camera, in which the ratios of the top and the bottom change to be smaller. The size of the whole body and the ratio of each body part are changing with the distance between the person and the camera. However, the size of each image captured by the same camera is invariable. A parameter R_{size} could be added, where $R_{size} = {}^{size\ of\ the\ body}\!/_{size\ of\ the\ image}$. When this parameter is larger than a threshold, the ratio of each body part will be accordingly adjusted. The segmentation could be more efficient using this self-adaptive ratio method.

- The proposed MID-SVM algorithm succeeded to classify persons in an online setting using CASIA and IA Database. However, there could be much more than 22 persons in more complicated applications. In this thesis, based on MID-SVM algorithm, the dynamic classifier needs to decide if it has to update when each single data is added or removed from the training dataset. The computation time will raise with the number of persons increasing. One solution to reduce the total computational cost is to update the model after multiple data points are added or removed, which is similar to the method in [KT10]. However, the choice of the interval of data is a new problem. Another angle to work on is the optimization of MID-SVM algorithm to obtain a faster process.

- In the experiments of person discovery based on SAKM algorithm, some different classes are easy to be clustered in one group, especially when they are wearing similar clothes. Another problem is that the samples of the same person will be separated into several clusters. More parameters or more thresholds of SAKM classifier could be tested to overcome these problems. One solution could to add another criterion on the angle between the incoming data and the previous support vectors [BL08]. The decision rules will then be based on the the similarity distance and the angle value.

- In future works, we wish to give more effective and robust features to characterize persons. Facial attributes are still very important and discriminable in visual surveillance, even though face can not be seen in some video sequences in challenging conditions. Weight-based facial features could be defined: the weight is 0 when no face is detected, otherwise the weight is greater than 0. We could use the features combined with appearance and weight-based facial features. This method could solve the problems of different persons wearing similar clothes or the same person re-identification after changing clothes.

Appendix A

Description of features

Head	Color	h_ mean_ r	Top	Color	t_ mean_ r	Bottom	Color	b_ mean_ r
		h_ std_ r			t_ std_ r			b_ std_ r
		h_ mean_ g			t_ mean_ g			b_ mean_ g
		h_ std_ g			t_ std_ g			b_ std_ g
		h_ mean_ b			t_ mean_ b			b_ mean_ b
		h_ std_ b			t_ std_ b			b_ std_ b
		h_ hist_ r_ beam1			t_ hist_ r_ beam1			b_ hist_ r_ beam1
		h_ hist_ r_ beam2			t_ hist_ r_ beam2			b_ hist_ r_ beam2
		h_ hist_ r_ beam3			t_ hist_ r_ beam3			b_ hist_ r_ beam3
		h_ hist_ r_ beam4			t_ hist_ r_ beam4			b_ hist_ r_ beam4
		h_ hist_ g_ beam1			t_ hist_ g_ beam1			b_ hist_ g_ beam1
		h_ hist_ g_ beam2			t_ hist_ g_ beam2			b_ hist_ g_ beam2
		h_ hist_ g_ beam3			t_ hist_ g_ beam3			b_ hist_ g_ beam3
		h_ hist_ g_ beam4			t_ hist_ g_ beam4			b_ hist_ g_ beam4
		h_ hist_ b_ beam1			t_ hist_ b_ beam1			b_ hist_ b_ beam1
		h_ hist_ b_ beam2			t_ hist_ b_ beam2			b_ hist_ b_ beam2
		h_ hist_ b_ beam3			t_ hist_ b_ beam3			b_ hist_ b_ beam3
		h_ hist_ b_ beam4			t_ hist_ b_ beam4			b_ hist_ b_ beam4
	Texture	h_ energy		Texture	t_ energy		Texture	b_ energy
		h_ correlation			t_ correlation			b_ correlation
		h_ inertia			t_ inertia			b_ inertia
		h_ entropy			t_ entropy			b_ entropy
		h_ inverse_ diff			t_ inverse_ diff			b_ inverse_ diff
		h_ sum_ avg			t_ sum_ avg			b_ sum_ avg
		h_ sum_ var			t_ sum_ var			b_ sum_ var
		h_ sum_ entropy			t_ sum_ entropy			b_ sum_ entropy
		h_ diff_ avg			t_ diff_ avg			b_ diff_ avg
		h_ diff_ var			t_ diff_ var			b_ diff_ var
		h_ diff_ entropy			t_ diff_ entropy			b_ diff_ entropy
		h_ info_ corr_ 1			t_ info_ corr_ 1			b_ info_ corr_ 1
		h_ info_ corr_ 2			t_ info_ corr_ 2			b_ info_ corr_ 2

Table A.1: 93 initial features extracted for CASIA Database and IA Database (RGB, 3Parts)

FeatureSet	FeatureNumber	Description		
CASIA-Wholeset	93	initial color and texture features		
CASIA-PCA	26	linear combinations of original features		
CASIA-CFS	40	h_mean_r h_std_r h_mean_b h_hist_r_beam2 h_hist_b_beam3 h_sum_avg	t_mean_r t_std_r t_mean_g t_mean_b t_std_b t_hist_r_beam1 t_hist_r_beam2 t_hist_r_beam3 t_hist_r_beam4 t_hist_g_beam1 t_hist_g_beam2 t_hist_b_beam1 t_hist_b_beam2 t_entropy t_sum_avg t_sum_var t_diff_entropy	b_mean_r b_std_r b_mean_g b_std_g b_mean_b b_hist_r_beam1 b_hist_r_beam2 b_hist_r_beam3 b_hist_r_beam4 b_hist_g_beam1 b_hist_g_beam2 b_hist_b_beam1 b_hist_b_beam2 b_hist_b_beam3 b_entropy b_sum_var b_diff_entropy
CASIA-Wrapper	16	h_mean_r h_std_r h_mean_g h_mean_b h_entropy	t_mean_r t_std_r t_mean_b t_std_b t_energy t_entropy	b_mean_r b_std_r b_std_g b_mean_b b_entropy

Table A.2: Features selected for CASIA Database (RGB, 3Parts)

FeatureSet	FeatureNumber	Description		
IA-Wholeset	93	initial color and texture features		
IA-PCA	22	linear combinations of original features		
IA-CFS	24	h_mean_r h_mean_g h_std_g h_diff_entropy	t_mean_r t_st_r t_mean_g t_std_g t_mean_b t_hist_r_beam1 t_hist_r_beam2 t_hist_g_beam1 t_hist_g_beam2 t_sum_avg t_sum_var t_info_corr_1	b_mean_r b_mean_g b_std_g b_mean_b b_std_b b_hist_r_beam1 b_hist_g_beam1 b_sum_avg
IA-Wrapper	17	h_mean_g h_std_b h_diff_entropy	t_mean_r t_std_r t_mean_g t_std_g t_std_b t_correlation t_sum_entropy t_diff_entropy	b_mean_r b_std_r b_mean_g b_std_g b_mean_b b_correlation

Table A.3: Features selected for IA Database (RGB, 3Parts)

Appendix B

Application of an incremental SVM algorithm for on-line human recognition from video surveillance using texture and color features. Neurocomputing, 2014

Neurocomputing 126 (2014) 132–140

Contents lists available at ScienceDirect

Neurocomputing

journal homepage: www.elsevier.com/locate/neucom

Application of an incremental SVM algorithm for on-line human recognition from video surveillance using texture and color features

Yanyun Lu [a,b], Khaled Boukharouba [a,b], Jacques Boonært [a,b], Anthony Fleury [a,b,*], Stéphane Lecœuche [a,b]

[a] Univ Lille Nord de France, F-59000 Lille, France
[b] EMDouai, IA, F-59500 Douai, France

ARTICLE INFO

Article history:
Received 5 January 2012
Received in revised form
3 May 2012
Accepted 20 August 2012
Available online 18 June 2013

Keywords:
Video surveillance
Human recognition
Incremental support vector machine
On-line multiclass classification

ABSTRACT

The goal of this paper is to present a new on-line human recognition system, which is able to classify persons with adaptive abilities using an incremental classifier. The proposed incremental SVM is fast, as its training phase relies on only a few images and it uses the mathematical properties of SVM to update only the needed parts. In our system, first of all, feature extraction and selection are implemented, based on color and texture features (appearance of the person). Then the incremental SVM classifier is introduced to recognize a person from a set of 20 persons in CASIA Gait Database. The proposed incremental classifier is updated step by step when a new frame including a person is presented. With this technique, we achieved a correct classification rate of 98.46%, knowing only 5% of the dataset at the beginning of the experiment. A comparison with a non-incremental technique reaches recognition rate of 99% on the same database. Extended analyses have been carried out and showed that the proposed method can be adapted to on-line setting.

1. Introduction

Nowadays, video surveillance is more and more considered as a solution to security enhancing and is, in this context, widely used in transports and public areas. Human recognition from video sequences and person tracking in a network of cameras is a key ability for such systems [1]. A significant amount of researches have been carried out in the field of human recognition, which are not only based on biometric features (face, gait, iris, etc.) [2–5], but also take into account non-biometric features (appearance) [6–8], especially in the application of pedestrian detection and multi-camera systems [9]. Appearance is defined by the person's visible clothing and body parts and it can be easily obtained after background subtraction (to isolate the person in the image). For a short period, the appearance of a person is expected to be invariant (same orientation with respect to the camera, same illumination, etc.). When considering longer-term periods of time, appearance can vary, especially in a network of cameras, in which the illumination is different or person changes the orientation. Even a very huge static database of people images cannot express the whole set of possibilities. On-line learning classifier with adaptive abilities could be a way to tackle this problem by exploiting the previous knowledge and updating the results from new conditions (environment, position of the person, etc.). It will, as a consequence, address this problem of short period of validity and use the system in the desired conditions.

In this work, Support Vector Machine (SVM) is implemented to settle the multi-category classification problem. Considering the generalization error, all-versus-all method (AVA) [10] is used, which solves a single quadratic optimization problem of size $(k-1)\cdot n$ in the k-class case (k is the number of classes and n is the number of samples). However, the classical SVM techniques are off-line and rely on the fact that the learning and testing phases are completely separated in the system. These methods are trained on a specific dataset and then tested in a real-world environment without any further learning. However, in our case, dealing with the human recognition in an open environment, the classes (persons) properties are dynamic and time-varying. On-line methods are particularly useful in the situations that involve on-line streaming data [11]. In 2009, Liang and Li have proved that incremental SVM is suitable for large dynamic data and more efficient than batch SVMs on the computing time [12]. Considering theses facts, an on-line model with incremental learning SVM as a solution is implemented in our system.

* Corresponding author at: Département Informatique et Automatique, École des Mines de Douai, 941, Rue Charles Bourseul, C.S. 10838, 59508 Douai, France. Tel.: +33 3 27 71 2381; fax: +33 3 27 71 2917/2980.
E-mail addresses: yanyun.lu@mines-douai.fr (Y. Lu), khaled.boukharouba@mines-douai.fr (K. Boukharouba), jacques.boonaert@mines-douai.fr (J. Boonært), anthony.fleury@mines-douai.fr (A. Fleury), stephane.lecoeuche@mines-douai.fr (S. Lecœuche).

0925-2312/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.neucom.2012.08.071

The work of Syed et al. [13,14] is considered as one of the first SVM with incremental learning. This work has been then extended and developed, such as with the SV-L-incremental algorithm [15] and NORMA algorithm [16]. However the work in [13] gives only approximate results. In 2001, Cauwenberghs and Poggio designed an exact on-line algorithm of incremental learning SVM, which updates the decision function parameters when adding or deleting one vector at a time [17]. In 2003, Diehl and Cauwenberghs improved the previous work and presented a framework for exact SVM incremental learning, adaptation and optimization, in order to simplify the model selection by perturbing the SVM solution when changing kernel parameters and doing regularization [18]. Most of these techniques allow only binary classification.

In order to tackle the problem of on-line multi-category classification, Boukharouba et al. proposed an incremental multi-class support vector classifier and the experiments showed that it could provide accurate results [19]. The classification algorithm in this paper reimplements this algorithm and applies it to solve human recognition problems in a network of cameras.

The main contribution of this paper is the application of incremental SVM techniques to the problem of identification/re-identification of persons in video-surveillance images, based only on appearance parameters. A large number of appearance parameters have been computed on these images, using a separation between three parts of the body to be closer to the description that one can do of a person. The system starts from the extraction of the previous selected features and it ends with the incremental classification of these images. Tests have been run on a collection of almost 20 000 images representing 20 different persons.

This paper is organized as follows: in Section 2, we present the global organization of the proposed human recognition system and introduce the experimental database. Section 3 describes the initial feature extraction and compares the performances of three different methods of feature selection. Then, Section 4 presents in detail the proposed on-line incremental multi-category SVM method. Section 5 shows the experimental results of the proposed method based on CASIA Gait Database. Finally, Section 6 concludes the paper and outlines our future works on this topic.

2. Human recognition in video frames

In this work, an on-line multi-class SVM algorithm is presented and applied to set up a surveillance system. Part of a first sequence of images is used to initialize the classifier to recognize the persons. Then the remaining images are used to test and update the classifier. Since we never have a complete information to represent and recognize a person in the learning step, incremental techniques are adapted to our problem. As the algorithm is incremental, each decision taken for a new frame will be used to update the SVM classifier. After receiving video frames, feature extraction is firstly implemented as a preprocessing stage, which has a strong influence on the quality of the recognition. As a consequence, the first part of this paper is dedicated to comparison of some feature selection methods. Then, incremental SVM was used as a multi-category classifier and its performance was compared to the one of the classical SVM (without incremental learning). The structure of our system is described by Fig. 1.

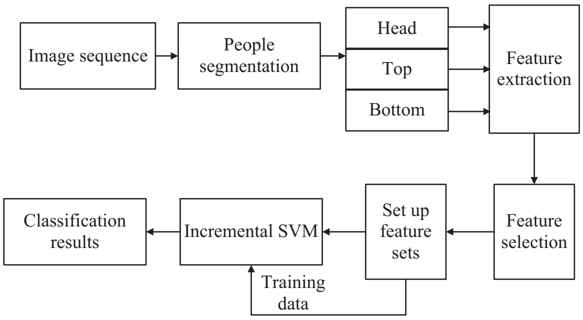


Fig. 1. The structure of proposed human recognition system.

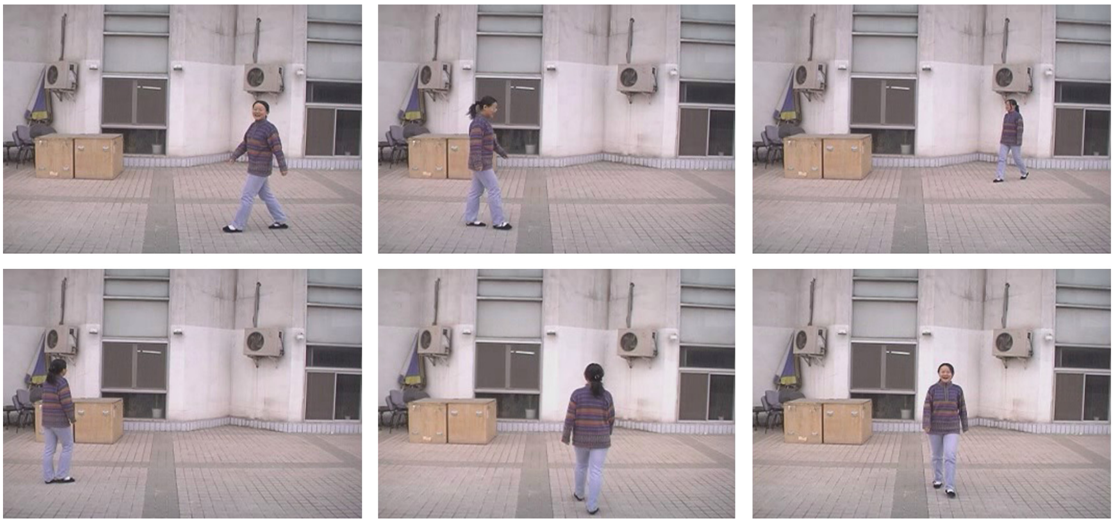


Fig. 2. One person with six actions in CASIA gait database [20].

Before collecting new realistic environmental data in our system, CASIA Gait Database [20] has been used. It is a video database of 20 persons, walking at normal pace. Each person walks with six different orientations relative to the video camera. Each image contains a unique person and the 20 persons did not change their clothes between trials. Six trials are presented for each of the 20 subjects: walking from right to left and from left to right, walking in front of the video camera (coming and leaving), walking in a direction that is at 45° of the camera from the left and from the right. Fig. 2 shows the six different actions that are repeated twice for each of the 20 persons. The whole number of images for the 20 classes is 19135 and the distribution of the samples in the different classes is described in Table 1. This table shows that the number of the samples in each of the 20 classes is almost the same (considering the average and standard deviation values).

CASIA Gait Database is first designed to recognize gait parameters in video images. However, for our application, and even if we are not interested in biometric features (as gait), the use of this dataset is relevant. From this dataset, we have, for the different person, a set of video with different orientation, possible different illumination, but no variation of clothes. However, as for every people, depending on the orientation, the statistics on the clothes can be different. This dataset allows to test our algorithm in the similar condition as in the case of video-surveillance, in which we have people walking and changing in the environment (camera, illumination), but no variation of clothing between images. It allows to test the ability of the system to be adapted according to the different views of a person in movement. That is the reason why, before collecting and creating our own and important database, we tested with this one, which is freely and publicly available. Furthermore all the results of the article can be verified and reproduced.

Table 1
The distribution of 20 classes within the 19 135 images in CASIA gait database.

Number of classes	20
Number of frame	19 135
Average cardinality	956.75
Std cardinality	83.5243

In CASIA Gait Database, background subtraction has been performed and the silhouette pictures have been extracted from the sequences. From the silhouette, we segment the body in three different parts: the head, the top part (the shoulders and the chest) and the bottom part (the legs). These three parts are shown in Fig. 3. Such segmentation of the body into three parts has been chosen because these three parts are generally of different colors (from different clothes) and are considered as a natural way to represent a person (when someone describe the appearance of a person). Gasser et al. [21] also used such segmentation to recognize people by a video camera. However, this previous work only used the average value of the three parts. In our system, each part is processed separately and the features are computed for these three parts independently. As a consequence, we have three different analyses that are more accurate than considering the whole body as a unique part. After extracting the initial features of 20 persons, we define different feature sets in order to investigate the comparison of computing time and classification results.

3. Feature extraction and selection

3.1. Extraction of the initial feature set

Object classification needs some attributes (features) to model the object to be recognized. Appropriate features can correctly represent the object and easily differentiate the classes. Most of the known and used features to define human beings are based on face, gait, body sharp and appearance [7,8,5,9]. Since the appearance of a person is made up of clothes and visible parts, color features are easy to obtain. Besides, color features are based on the general characteristics of the pixels and invariant to translation, rotation and not sensitive to scale under correct normalization conditions. As a consequence, color features are extracted, combined with texture features, to define the persons to recognize.

Color features with Red, Green and Blue (R, G and B) components of each frame captured by camera are varying depending on several factors, such as illumination condition, surface reflectance and quality of the camera. As a consequence, normalization is necessary. In this paper, the grey-world normalization (GN) has been used and gives the normalization R′, G′ and B′. It assumes that

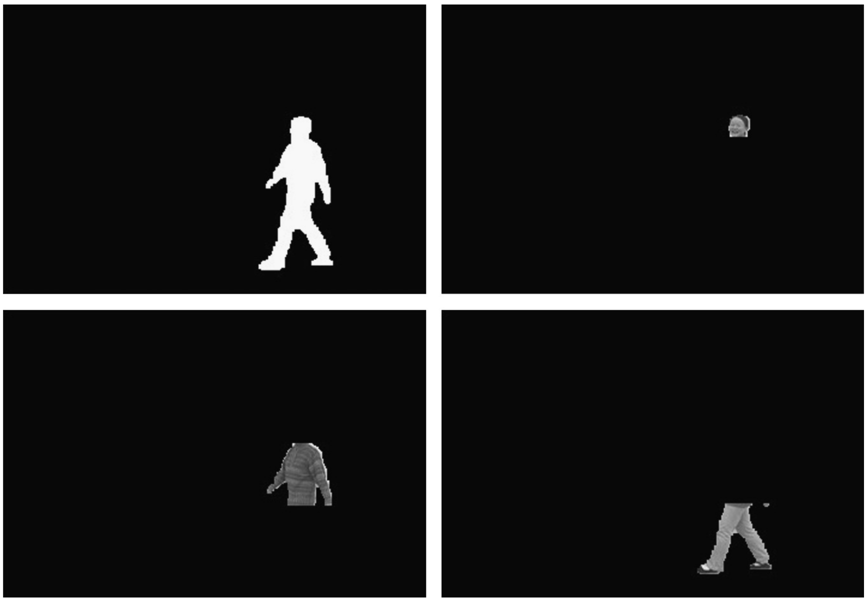


Fig. 3. The silhouette picture and the three parts of the body that have been considered.

Table 2
The initial features based on color and texture

Type of feature	Description
Color	Mean value for R′, for G′ and for B′ Standard deviation for R′, for G′ and for B′ Histogram with 4 beams for R′, for G′ and for B′
Texture	Energy Correlation Inertia Entropy Inverse difference moment Sum average Sum variance Sum entropy Difference average Difference variance Difference entropy Information measure of correlation 1 Information measure of correlation 2

changes in the illuminating spectrum can be modelled by three constant factors applied to R, G and B, in order to obtain invariance and be more robust to illumination intensity variations [22].

As explained in the end of Section 2, we considered three different parts for each body. For each part, we compute the different color features, which consist of *Mean Value*, *Standard Deviation Value* for each color component and *Energy* in four beams of the histogram of the image. This leads to the extraction of 18 color-based features for each part of the body of three color components. That is to say, we have 54 color-based features for each person.

Texture features based on the spatial co-occurrence of pixel values have been previously defined by Haralick [23]. With this method, thirteen features have been given. Then one matrix is obtained to represent regions, with values between 0 (black) and 255 (white) after converting each body part in grey levels. From the Spatial Grey Level Dependence Matrix, 13 features of each part of the body are computed. They are listed in Table 2. For each person to describe, we obtain 39 texture-based features.

Based on color and texture features, we computed a total of 93 features for each person: 54 color-based features and 39 texture-related features. The features are named firstly with the position (*h* for head, *t* for top and *b* for bottom), and secondly with its description (*mean* for average value, *std* for standard deviation, *hist_beam*1 or *hist_beam*2 for the histograms) and finally with the color if it applies (*r* for red, *g* for green and *b* for blue). For instance, the *Mean Value* of the top in the red component is named as *t_mean_r*.

3.2. Feature selection

To represent persons in the recognition system, we could use a high dimensional data that can lead to high discrimination performances in classification. However, high dimensional data are difficult to interpret and may raise dimensionality curse problems. In order to avoid useless dimensions of the training data, and as a consequence, reduce the computing time, many algorithms with supervised or unsupervised learning are designed to reduce dimensionality to its minimum, still keeping the high performances obtained using the original dataset.

Based on how is constructed the search for the optimal feature set, feature selection methods are mainly divided into three categories: filter methods (open-loop methods), wrapper methods (closed-loop methods) and embedded methods (closed-loop methods, which also can be seen as part of wrapper methods) [24,25]. Filter methods work on the data without considering the classification algorithm. The evaluation of the subset (with a criterion or an heuristic), as a consequence, depends only on the inner properties of the dataset (distribution of the values, correlation between features and with the class, etc.). Wrapper methods also use a criterion or an heuristic to evaluate the different subset but on the contrary to the filter methods, this heuristic depends on the performance of a selected classifier, on the current dataset with the chosen features. In that case, not only the properties of the data are considered but also the classifier and its performances with the selected features.

A large number of feature subset selection methods exist. We chose some algorithms that seemed to us relevant examples of methods to compare them and their efficiency in our problem. The first selected method is PCA. This method is based on the properties and distribution of the data to determine a new feature space (for which each dimension is a linear combination of attributes) in which the data are efficiently represented (in terms of independence between dimensions). This method is part of the filter methods. The second one, Correlation-based Feature Selection, is also a filter method but uses an heuristic to determine the better and the smaller set. In this second method, this determination is based on the correlation (between attributes and with the class), that are also inner properties of the class, but on the contrary to PCA, CFS will select a subset of attribute and not create a new space from the whole set of attributes. Finally, after these two filter-based approaches, we tested wrapper-based feature selection.

3.2.1. PCA

Feature selection based on PCA aims at reducing the number of dimension without losing the main information by projecting the data on a new orthogonal basis [26]. In our work, when the sum of the variance is at least equal to 95% of the initial variance of the data, we stop and consider the subspace as optimal answer to our problem.

As a result, 26 features are created, which are linear combinations of the 93 initial features. These 26 new data for each of the 19 135 images constituted a new database (CASIA-PCA).

3.2.2. Correlation-based feature selection

Correlation-based feature selection (CFS) is a simple filter algorithm which ranks feature subsets according to the correlation based on the heuristic of “merit”, which was described by Hall [27] as the following expression:

$$M_s = \frac{k \cdot \overline{r_{cf}}}{\sqrt{k + k \cdot (k{-}1) \cdot \overline{r_{ff}}}}$$

where k is the number of features selected in the current subset, $\overline{r_{cf}}$ is the mean feature-class correlation, for each element $f \in S$ of our current subset, and $\overline{r_{ff}}$ is the mean feature-feature correlation for each pairwise of elements. From this heuristic, the search method begins with the empty set and adds some features, one at a time, in order to find efficiently the feature set that possesses the best value. Best first method is applied to search the set with the best merit value.

For our initial set of features, the algorithm gives us a subset of 40 features that are the most representative features with the less possible redundancy. The features selected in the final subset (CASIA-CFS) are listed in Table 3, where for instance 5−*color*−*head* means that there are 5 color features chosen in the head part. The texture features are less represented in the final subset and the most important part of the subset is given by the bottom part of the body.

3.2.3. Wrapper

Wrapper method has been initially described by John et al. [24]. Similar to CFS, wrapper method uses a search algorithm to go through the whole combination of features. But it computes the

Table 3
The features selected by each data set.

FeatureSet	FeatureNumber	Description
CASIA-Wholeset	93	Initial color and texture features
CASIA-PCA	26	Linear combinations of original features
CASIA-CFS	40	5-color-head 13-color-top 14-color-bottom 1-texture-head 14-texture-top 3-texture-bottom
CASIA-Wrapper	16	4-color-head 4-color-top 4-color-bottom 1-texture-head 2-texture-top 1-texture-bottom

merit of a subset according to the results of the classification (given, for instance, by global error rate) of the dataset with the targeted algorithm. As a consequence, the execution time before obtaining the desired results could be huge (because of the necessity for each tested subset of training and testing). However, the advantage of this method is that it can give better results as the classification algorithm is already specified and used to compute the merit.

Over the 93 features, 16 features for each person have been selected by this method, as presented in Table 3. As a result, a new dataset (CASIA-Wrapper) is created. Color features, especially their *Mean Values* and *Standard Deviation Values*, are well represented and texture features (*Entropy*) are selected in all the parts of the body.

3.3. Discussion on the selected features

As described in the above subsections, four sets of features were prepared for the classification stage: CASIA-Wholeset (initial set of 93 features), CASIA-PCA, CASIA-CFS and CASIA-Wrapper. Table 3 gives the features selected by each set.

For PCA method, we obtained 26 features instead of the initial 93 features with a 99.6% relevance (the average of ROC Area). However, the disadvantage of PCA method is that it loses the interpretation of the features, because the features selected by PCA are the combinations of the initial features. In CASIA-CFS, most of the selected features are color-based. And comparing CASIA-CFS with CASIA-Wrapper feature sets, 11 features are in both sets: *h_mean_r*, *h_std_r*, *h_mean_b*, *t_mean_r*, *t_std_r*, *t_mean_b*, *t_std_b*, *b_mean_r*, *b_std_r*, *b_std_g*, *b_mean_b*. In addition, when we carefully look at the values of the covariance matrix of PCA (that gives us the importance of each feature in the linear combination creating the new vectors), we can notice that the ones that are selected by the other methods are selected with higher coefficients. It is obvious that color-based features are more useful in people classification in our system. Entropy features of texture-based are usually selected and give the most useful information to human classification in all feature sets.

In Section 5, the discussion will be given on the classification results obtained by four different feature sets.

4. On-line multi-category SVM with incremental learning

In Section 3, the preprocessing step of the human recognition system (feature extraction and selection) has been presented. In this section, we mainly describe the implementation of incremental SVM in an on-line human recognition system and explain how the system effectively update the parameters of the classifier when a new frame is presented and classified.

Any incremental learning algorithm should satisfy the following conditions [28]: (1) it has ability to learn additional information brought by new data; (2) it should preserve knowledge of the previous training data; (3) it has ability to create new classes with new data; (4) it should not require access to the original data, which are used to train the existing classifier. An on-line model should have the ability to be used during the learning step and update with the information brought by new data. The proposed incremental SVM algorithm is satisfied with these conditions, it can be depicted as follow: in the case of multi-category classification, when a new data is added, the incremental algorithm adapts the decision function in a finite number of steps until all the samples in the existing training set satisfies the Karush–Kuhn–Tucker (KKT) conditions.

In this section, we redefine and explain the main principle of incremental SVM and some important details on the functioning of the algorithm presented in [19]. This paper applies this algorithm to a new problem (identification of persons) and does not change the way the algorithm is defined.

4.1. Multi-category SVM and the KKT conditions

Let us consider a training dataset T of N pairs (x_i, y_i), where $i = 1, \dots, N$, $x_i \in R^d$ is the input data, $y_i \in \{1, \dots, K\}$ is the output class label, $K \geq 2$. The SVM classifier used for data classification is defined by

$$x_i \in C_k; k = \underset{j=1,\dots,K}{\arg} \max f_j(x_i) \tag{1}$$

Each decision function f_i is expressed as

$$f_i(x) = w_i^T \Phi(x) + b_i \tag{2}$$

where function $\Phi(x)$ maps the original data x_i to a higher-dimensional space to solve non-linear problems. In multi-category classification, the margin between classes i and j is $2/\|w_i - w_j\|$. In order to get the largest margin between classes i and j, minimization of the sum of $\|w_i - w_j\|^2$ for all $i, j = 1, \dots, K$ is computed. Also, as described in [10], the regularization term $\frac{1}{2}\sum_{i=1}^{K} \|w_i\|^2$ is added to the objective function. In addition, a loss function $\sum_{i=1}^{K} \sum_{j=i+1}^{K} \sum_{x_l \in C_{ij}} \xi_l^{ij}$ is used to find the decision rule with the minimal number of errors in the inseparable case, where the slack variable ξ_l^{ij} measures the degree of misclassification of the lth training vector, related to the hyper-plan ij. So, the proposed quadratic function is as the follows:

$$\begin{aligned} \min_{w_i, b_i} \quad & \frac{1}{2} \sum_{i=1}^{K} \sum_{j=i+1}^{K} \|w_i - w_j\|^2 + \frac{1}{2} \sum_{i=1}^{K} \|w_i\|^2 + C \sum_{i=1}^{K} \sum_{j=i+1}^{K} \sum_{x_l \in C_{ij}} \xi_l^{ij} \\ \text{s.t.} \quad & \forall x_l \in C_{ij}; \\ & y_l^{ij}[(w_i - w_j)^T \Phi(x_l) + (b_i - b_j)] - 1 + \xi_l^{ij} \geq 0; \\ & \xi_l^{ij} \geq 0; \quad i = 1, \dots, K; \; j = i+1, \dots, K \end{aligned} \tag{3}$$

where $C \geq 0$ trades off the term that controls the number of outliers. A larger C corresponds to assign a higher penalty to errors.

The goal is to minimize this objective function, which is a quadratic programming task. We solve it by Lagrange multipliers method. The Lagrange function L is defined by

$$\begin{aligned} L = & \frac{1}{2} \sum_{i=1}^{K} \sum_{j=i+1}^{K} \|w_i - w_j\|^2 + \frac{1}{2} \sum_{i=1}^{K} \|w_i\|^2 \\ & + C \sum_{i=1}^{K} \sum_{j=i+1}^{K} \sum_{x_l \in C_{ij}} \xi_l^{ij} - \sum_{i=1}^{K} \sum_{j=i+1}^{K} \sum_{x_l \in C_{ij}} \mu_l^{ij} \xi_l^{ij} \end{aligned}$$

$$-\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l \in C_{ij}} \alpha_l^{ij}(y_l^{ij}[(w_i - w_j)^T \Phi(x_l) + (b_i - b_j)] - 1 + \xi_l^{ij}) \tag{4}$$

where $\alpha_l^{ij} \geq 0$, $\mu_l^{ij} \geq 0$, $i \neq j$ are Lagrange coefficients.

The Lagrangian L has to be minimized with respect to w_i, b_i and ξ_l^{ij} and maximized with respect to α_l^{ij} and μ_l^{ij}. Then in the saddle point the derivation of the Lagrangian L is equal to zero, for all $i=1, \ldots, K$, we compute the following gradient: $\partial L/\partial w_i = 0$, $\partial L/\partial b_i = 0$ and $\partial L/\partial \xi_l^{ij} = 0$. In consequence, we get

$$w_i = \frac{1}{K+1}\sum_{\substack{j=1\\ j\neq i}}^{K}\left(\sum_{x_l \in C_i} \alpha_l^{ij}\Phi(x_l) - \sum_{x_l \in C_j} \alpha_l^{ij}\Phi(x_l)\right) \tag{5}$$

$$\sum_{\substack{j=1\\ j\neq i}}^{K}\left(\sum_{x_l \in C_i} \alpha_l^{ij} - \sum_{x_l \in C_j} \alpha_l^{ij}\right) = 0 \tag{6}$$

$$\alpha_l^{ij} + \mu_l^{ij} = C \tag{7}$$

Then by replacing w_i by its expression (Eq. (5)), the problem of optimization of L is transformed to a minimization of dual formulation W as shown in [19].

4.2. Incremental algorithm

The main idea of incremental learning SVM is to train a SVM with a partition of the dataset, reserve only the support vectors at each training step and create the training set for the next step with these vectors. Syed et al. showed that the decision function of an SVM depends only on its support vectors, that is to say that it will achieve the same results between using the whole dataset and only the support vectors [14]. The key of incremental algorithm is to preserve the KKT conditions on all existing training data while adiabatically adding a new vector.

The KKT conditions on the point $x_m \in C_{ij}$ divide data D into three categories according to the value of g_m^{ij} for all $i=1, \ldots, K$, $j=i+1, \ldots, K$

$$g_m^{ij} = \frac{\partial W}{\partial \alpha_m^{ij}} \begin{cases} >0; & \text{if } \alpha_m^{ij}=0; & D(dv_m^{ij}) \\ =0; & \text{if } 0<\alpha_m^{ij}<C; & S(sv_m^{ij}) \\ <0; & \text{if } \alpha_m^{ij}=C; & E(ev_m^{ij}) \end{cases} \tag{8}$$

As explained in Fig. 4, support vectors (S) are on the boundary, error vectors (E) exceed the margin and data vectors (D) are inside the boundary.

When a new data x_c is added, we initially set these coefficient $\alpha_c^{pq}=0$, $p=1, \ldots, K, q=p+1, \ldots, K$ change value incrementally and the parameters of the existing support vectors are adapted in order to keep satisfied to the KKT conditions. In particular, the adaptation of g_m^{ij} when new data x_c is added can be expressed differentially as

$$\Delta g_m^{ij} = y_m^{ij}\left(\beta^{ij,pq}\Delta\alpha_c^{pq}K_{cm} + \sum_{x_l \in C_i}\left(2\Delta\alpha_l^{ij} + \sum_{\substack{n=1\\ n\neq i,j}}^{K}\Delta\alpha_l^{in}\right)K_{lm} - \sum_{x_l \in C_j}\right.$$
$$\left.\left(2\Delta\alpha_l^{ij} + \sum_{\substack{n=1\\ n\neq i,j}}^{K}\Delta\alpha_l^{nj}\right)K_{lm} - \sum_{\substack{n=1\\ n\neq i,j}}^{K}\sum_{x_l \in C_n}(\Delta\alpha_l^{in} - \Delta\alpha_l^{nj})K_{lm} + (\Delta b_i - \Delta b_j)\right) \tag{9}$$

$$\gamma^{i,pq}\Delta\alpha_c^{pq} + \sum_{\substack{n=1\\ n\neq i}}^{K}\left(\sum_{x_l \in C_i}\Delta\alpha_l^{in} - \sum_{x_l \in C_j}\Delta\alpha_l^{in}\right) = 0 \tag{10}$$

where $i=1, \ldots, K$, $j=i+1, \ldots, K$, α_c^{pq} is the coefficient being incremented and K is kernel function, so that $K_{lm}=K(x_l, x_m)=\Phi(x_l)^T \Phi(x_m)$. Coefficients $\beta^{ij,pq}, \gamma^{i,pq}$ are defined in [19].

From Eq. (8), for all the support vectors sv_m^{ij}, we get $g_m^{ij}(sv_m^{ij})=0$, then $\Delta g_m^{ij}(sv_m^{ij})=0$. Therefore Eqs. (9) and (10) can be written as the following matrix equation: (as described in detail in [19])

$$\begin{bmatrix} \Delta b \\ \Delta\alpha \end{bmatrix} = -RH^{pq}\Delta\alpha_c^{pq} \tag{11}$$

where $b=[b_1, \ldots, b_K]$ and $\alpha=[\alpha_1^{12}, \alpha_2^{12}, \ldots \alpha_i^{ij}, \alpha_j^{ij}, \ldots, \alpha_{K-1}^{(K-1)K}, \alpha_K^{(K-1)K}]$, α_i^{ij} expresses the weights of support vectors sv_n^{ij} that belong to the class C_i.

This last equation will be used to update the decision functions expressed first in Eq. (2).

4.3. Adding a new vector

When a new sample x_c is added, depending on the value of g_c^{pq}, if $g_c^{pq}>0$, x_c is not a support vector or error vector, we add it to the set D and terminate; else we apply the increment of α_c^{pq} and update all the coefficients for the vectors in S_p, where S_p is the support vector set of the previous step. When $g_c^{pq}=0$, x_c is considered as a support vector and added to the set S. At each incremental step, the value of b and a in Eq. (11) are updated. As a consequence, the matrix R, the data, the number of support vector in set S and the decision function will be updated. x_c can also be an error vector if $\alpha_c^{pq}=C$ and add it to the set E. Otherwise, if the value of $\Delta\alpha_c^{pq}$ is too small to cause the data move across S, E and D, the largest possible increment α_c^{pq} is determined by the book-keeping procedure [17].

4.4. Migration of data between the three sets

When a new data is added, the hyperplane of the SVM classifier is updated, and then the vectors of different sets (S, E and D with $T=S \cup E \cup D$) could migrate from their current set to a neighbor set. Fig. 4 explains the geometrical interpretation of each set and from this figure we can infer the possible migrations as follows:

- From D or E to S: the data vector or error vector becomes a support vector. This case happens when the update value of g_m^{ij} for $x_m^{ij} \in D$ reaches 0.
- From S to E: the previous support vector becomes an error vector. This case is detected when α_m^{ij} is equal to C.
- From S to D: the previous support vector becomes a data vector. This case is detected when α_m^{ij} is equal to 0.

4.5. Implementation of the incremental SVM algorithm

The details of the algorithm are described in [19], and the algorithm has been implemented in the Weka Software [29], as a package.

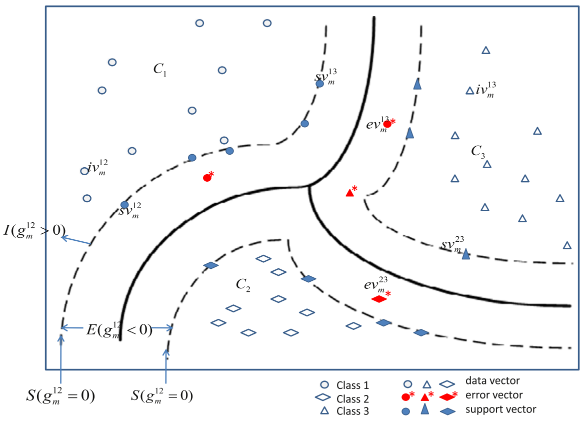


Fig. 4. Three sets obtained from training samples, considering three classes.

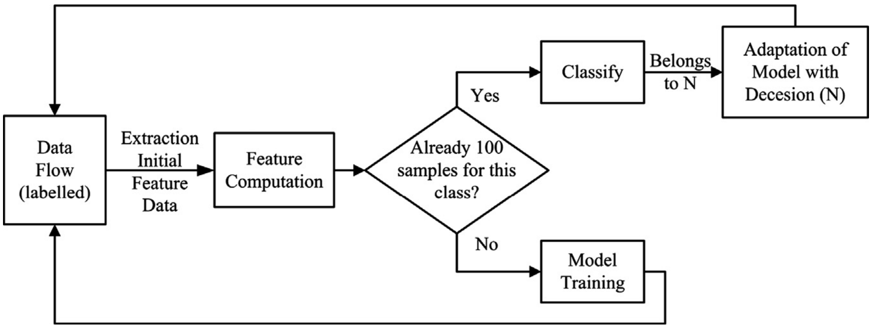


Fig. 5. Incremental learning work flow.

Table 4
Recognition rates of SVM with incremental learning based on the four proposed databases ($\sigma = 1$, $C = 19$, $step = 10^{-3}$).

	Wholeset	PCA	CFS	Wrapper
C1	**92.16**	**91.07**	**92.59**	**91.83**
C2	99.44	99.72	99.44	99.44
C3	100	98.74	100	99.58
C4	95.77	95.37	97.99	**91.15**
C5	99.10	99	99.10	99.10
C6	99.37	**92.15**	99.79	**92.25**
C7	100	**93.65**	100	99.08
C8	100	99.76	100	99.88
C9	100	100	100	100
C10	100	100	100	100
C11	99.78	99.23	99.78	96.92
C12	99.31	96.11	98.28	95.65
C13	98.98	**94**	99.39	**86.98**
C14	97.34	**94.42**	97.72	**93.03**
C15	100	**91.69**	100	**94.16**
C16	96.87	97.74	98.62	95.11
C17	98.23	98.11	98.23	97.99
C18	**91.89**	**85.70**	**92**	**83.88**
C19	97.42	**85.77**	96.56	**93.37**
C20	99.22	100	100	99.67
Global	98.21	95.6	98.46	95.39

Table 5
Recognition rate of SVM without incremental learning based on the four proposed databases.

	Wholeset	PCA	CFS	Wrapper	Wrapper 5%
C1	99.8	97.7	99.8	97.4	64.3
C2	100	99.8	99.8	99.8	99.4
C3	100	99.9	100	100	97.2
C4	99.2	98.6	99.2	98.8	82
C5	99.8	99.1	99.8	99.8	92.7
C6	99.6	98.6	99.5	97.5	60.1
C7	100	100	100	100	77.6
C8	100	99.9	100	100	91.6
C9	100	99.9	100	100	98.7
C10	100	100	100	100	99.9
C11	100	100	100	99.6	72.1
C12	100	98.5	99.9	99.3	82.6
C13	100	99.1	100	98.9	83.2
C14	99.6	98.2	99.8	97.6	76.2
C15	100	99.3	100	99.6	78.2
C16	99.9	99.6	99.9	99.3	92
C17	99.8	99.4	99.7	99.2	93
C18	99.6	96.9	99.8	98.3	63.2
C19	100	97.4	100	99.4	75.2
C20	100	100	100	100	95
Global	99.9	99.1	99.9	99.2	83.7

5. Experimentation

This section shows the experimental results of human recognition with incremental SVM classifier based on CASIA Gait Database. In addition, a comparison experiment is performed using classical SVM based on the same database.

5.1. *Results of the proposed method*

Fig. 5 illustrates the workflow of the data in our incremental system. Only 50 images of each class (5% of the whole dataset) are used for training and the remaining images are used for testing (and updating the classifier). Both training and testing phase use incremental learning, in which new frame is added one by one and the recognition system is updated step by step with an adaptive decision function. The difference between training and testing phase is that during the training step, the class labels of the added samples are correct. In the testing phase, the class labels are given by the classification of SVM with the current class model and so are accurate only if the classification went well.

Table 4 shows the results of classification of incremental SVM. As expected, the CASIA-PCA dataset has lower performances than the others. CASIA-Wrapper has also lower results. Because the number of features selected by CASIA-Wrapper is the smallest, as a consequence, CASIA-Wrapper lost more information of the initial features than the others. In the CASIA-CFS one, more features are extracted, but it reduces considerably the number of features comparing to the initial feature set. CASIA-CFS dataset gives the best performances, which are similar with the results of the whole dataset. But CASIA-CFS reduces the processing time, due to the reduction of the number of the features.

The results of the different classes vary among the four different feature sets. The global recognition results for the four feature sets are encouraging (higher than 95%). The best set is given by CASIA-CFS with a 98.46% global recognition rate. Some classes are less correctly recognized, such as C1 and C18 with recognition rates below 93% in the four datasets. The results below 95% are shown in bold typesetting in Table 4. The experimental results show that the proposed incremental SVM is able to meet the demands of on-line multi-category classification and achieves satisfying classification rate. Next parts will compare these results with the classical SVM using training/testing set split.

5.2. *Results of the comparison experiment with classical SVM*

We performed a comparison experiment using a classical SVM algorithm based on the same feature datasets in the same experimental conditions, to check if the lower results of some classes are caused by inner properties of these classes. The RBF kernel has also been used in classical SVM with the same kernel parameter and the same value for *C*.

In the comparative experiments, the classical SVM classifier is tested with a stratified 10-folds cross-validation (randomly chosen). Table 5 reports the comparative results of classification recognition rate of non-incremental learning for the four different feature sets. In these sets, the results are almost identical and the mean of recognition rate of all classes achieves more than 99%. Some classes, which got lower recognition rates in incremental SVM, achieve good performances in the classical SVM. That is to say, lower results of incremental learning are not due to inner properties of the classes, but are due to a lack of knowledge at some point of the learning process that turns some examples of some class into examples of other classes.

The two tests protocols (with and without incremental learning) are different, however, it would have been difficult to achieve two interesting tests with the same protocol. The results with a non-incremental algorithm are used as a reference to show if incremental SVM can achieve similar satisfying results. As we have no formal comparison of the two techniques, these initial comparative results give some clues on the remaining work to improve our incremental procedure.

To extend these results and check that incremental learning is bringing new insights, we also performed another experiment. This experiment is presented in the last column of Table 5. For this experiment, the same number of images, which are used in the initial stage of the incremental learning experiment, have been used to train the classical SVM. The remaining images have been used for testing. We can see that considering a low number of images in the training dataset make that the results decrease largely comparing to the other columns for which the number of available data were higher. That demonstrates that some classes, even if the clothes of each person does not change, are not so easily distinguishable considering only a small sequence of images. As a consequence, use of incremental learning techniques is justified for such application in which, even if the subject appearance is not supposed to change, will present some variation and drifts in the class model because of the different orientation of the person comparing to the camera or because of the changes of illumination.

5.3. Discussion

In the classical SVM, all the training data are available in the learning process and the recognition system could be tested after the end of learning process. This learning phase aims to minimize the error (with the slack variables) to determine the boundaries for classification. However, even in this case, we have some errors that are introduced by the classifier, but these vectors are ignored (they are known as error vectors and they do not belong to a class). When performing with incremental learning, we update the margins with the result of the classification. As a consequence, if a frame is incorrectly classified, it will be considered as a part of a wrong class. When retraining, this frame possibly could become a support vector of this wrong class. Without the whole information of all training data at the beginning, if such frame is presented just after the initialization process, it could migrate to the set S of support vectors instead of the set E of error vectors. And this support vector will update the decision function and include a slow movement of the separation between the considered classes. The wrong result by this false support vector will then have an impact on the remaining classifications. For instance, if a new frame, which is close to this class with a wrong support vector, is presented and again incorrectly classified, the inaccuracy will increase. In a noise-free case, [19] proved that at the end of the process, the support vectors are the same (and so are the boundaries). However, as in our case, some of the vectors that are used for training can be misclassified, we will get a different support vector set comparing to the classical SVM.

In our implementation of the algorithm, some tracking materials are added for the support vectors in order to verify if some of them are erroneously classified. We have checked that some vectors are indeed mistaken for support vectors in some classes, even if they are not good descriptors of these classes (which cause the misclassification). For some of the classes (C1 and C18 in all the datasets and for instance C4 in CASIA-Wrapper), these support vectors even stay in the final support vectors set. For some other classes, these vectors become support vectors and then have been eliminated (to become error or data vectors) after some steps.

6. Conclusion and future work

In this paper, we have presented an on-line human recognition system from video surveillance images, which is an extension and an application of the work in [19] about incremental SVM. The proposed recognition method is from a very limited training set and has showed good performances of the tests on a real database.

Incremental SVM technique is firstly introduced to perform person recognition in a real-world application. In order to overcome the problems that the classes (persons) are not completely known at the beginning of the process and the classical SVM for this task requires a huge database with too many samples for each person, incremental learning algorithm is implemented, in which learning process of human recognition is from just a few training images. Incremental learning algorithm is fast and update the recognition model according to the different expositions or orientations of the persons (assuming that the drift is moderated between the learning phases). In addition, it overcomes the drawback of the appearance-based features (short period of validity). Therefore, incremental learning algorithm is more suitable for the practical situation of on-line surveillance system.

Second, according to our specific application, the most efficient feature set is determined. Since the analysis of three different parts of the body are more accurate than the one of the whole body, segmentation of each body into three parts (Head, Top, and Bottom) has been implemented firstly. Considering these three parts of each silhouette, color and texture features are extracted in the video sequences and the Wholeset database with 93 features is formed. Then three feature selection methods are compared to reduce the feature space and obtain the optimal set. These feature selection methods are chosen to represent the two different methods (filter and wrapper). Finally, four sets of database are obtained. The most satisfied result is based on CFS, which consists of 40 features for each image.

Third, incremental SVM is tested with the four different feature sets and is compared to the classical SVM with non-incremental learning. The experimental results show the recognition rates of the classical SVM are satisfying with all feature databases. We also showed that incremental SVM satisfies the condition of on-line setting and the performance with the CASIA-CFS dataset is good with more than 98% of global accuracy rate. These results are compared with non-incremental results, to show that it performs well if we have a sufficient knowledge of the classes but not if we consider the reduced number of training samples that is needed for the better results of incremental SVM.

In the future work, more attention should be paid to the misclassified vectors for the updating of the decision function. Rüping proposed an algorithm named SV-L-incremental algorithm [15], which added a coefficient (comparable to the slack variable) to old support vectors and with the consequence to reduce the weights of the new data when updating. The future work includes to find a metric able to indicate us a probability to be wrong in classification and to decide for each vector if we will retrain or not, using this new information.

Finally, the last part of our future work is to try to create new classes from data that are collected. For the moment, we only can classify and update the recognition model with already known person. Novelty detection and class creation will be a part of the design of the system that suits for video surveillance applications requirements. For that, we have to consider another criterion that will determine the non-inclusion of the data to any of the known class and decide to split to create a new class.

Acknowledgments

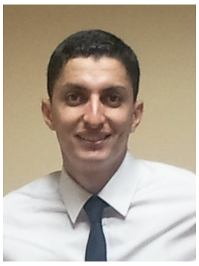
Portions of the research in this paper use the CASIA Gait Database collected by Institute of Automation, Chinese Academy of Sciences.

The authors gratefully acknowledge financial support from China Scholarship Council (CSC).

References

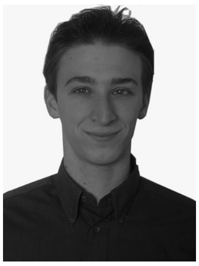
[1] D. Truong Cong, L. Khoudour, C. Achard, C. Meurie, O. Lezoray, People re-identification by spectral classification of silhouettes, Signal Process. 90 (8) (2010) 2362–2374.
[2] D. Roark, A. O'Toole, H. Abdi, Human recognition of familiar and unfamiliar people in naturalistic video, in: IEEE International Workshop on Analysis and Modeling of Faces and Gestures, AMFG 2003, Nice, France, 2003, pp. 36–43.
[3] D. Kaziska, A. Srivastava, Gait-based human recognition by classification of cyclostationary processes on nonlinear shape manifolds, J. Am. Stat. Assoc. 102 (480) (2007) 1114–1124.
[4] X. Zhou, B. Bhanu, Feature fusion of side face and gait for video-based human identification, Pattern Recognition 41 (3) (2008) 778–795.
[5] X. Zhou, B. Bhanu, Integrating face and gait for human recognition at a distance in video, IEEE Trans. Syst. Man Cybern. Part B Cybern. 37 (5) (2007) 1119–1137.
[6] D. Makrisa, N. Doulamisc, S. Middletond, Vision-based production of personalized video, Signal Process. Image Comput. 24 (5) (2009) 158–176.
[7] K. Yoon, D. Harwood, L. Davis, Appearance-based person recognition using color/path-length profile, J. Visual Commun. Image Representation 17 (3) (2006) 605–622.
[8] E. Hörster, J. Lux, R. Lienhart, Recognizing persons in images by learning from videos, in: Proceedings of SPIE, vol. 6506, 2007, pp. 65060D.1–65060D.9.
[9] D. Truong Cong, L. Khoudour, C. Achard, L. Douadi, People detection and re-identification in complex environments, IEICE Trans. Inf. Syst. 93 (7) (2010) 1761–1772.
[10] E. Bredensteiner, K. Bennett, Multicategory classification by support vector machines, Comput. Optim. Appl. 12 (1) (1999) 53–79.
[11] S. Agarwal, V. Vijaya Saradhi, H. Karnick, Kernel-based online machine learning and support vector reduction, Neurocomputing 71 (7) (2008) 1230–1237.
[12] Z. Liang, Y. Li, Incremental support vector machine learning in the primal and applications, Neurocomputing 72 (10–12) (2009) 2249–2258.
[13] N.A. Syed, S. Huan, L. Kah, K. Sung, Incremental learning with support vector machines, in: Workshop on Support Vector Machines at the International Joint Conference on Artificial Intelligence, 1999.
[14] N. Syed, H. Liu, K. Sung, Handling concept drifts in incremental learning with support vector machines, in: Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, 1999, p. 321.
[15] S. Rüping, Incremental learning with support vector machines, in: Proceedings of the IEEE International Data Mining ICDM 2001 Conference, 2001, pp. 641–642.
[16] J. Kivinen, A. Smola, R. Williamson, Online learning with kernels, IEEE Trans. Signal Process. 52 (8) (2004) 2165–2176.
[17] G. Cauwenberghs, T. Poggio, Incremental and decremental support vector machine learning, Adv. Neural Inf. Process. Syst. 13 (2001) 409–415.
[18] C. Diehl, G. Cauwenberghs, SVM incremental learning, adaptation and optimization, in: Proceedings of the International Joint Conference on Neural Networks, 2003, vol. 4, IEEE, 2003, pp. 2685–2690.
[19] K. Boukharouba, L. Bako, S. Lecoeuche, Incremental and decremental multi-category classification by support vector machines, in: 2009 International Conference on Machine Learning and Applications, IEEE, 2009, pp. 294–300.
[20] CASIA Gait Database, URL: ⟨http://www.sinobiometrics.com⟩ (2001).
[21] G. Gasser, N. Bird, O. Masoud, N. Papanikolopoulos, Human activities monitoring at bus stops, in: Proceedings of the IEEE International Conference on Robotics and Automation, vol. 1, New Orleans, LA, 2004, pp. 90–95.
[22] G. Finlayson, B. Schiele, J. Crowley, Comprehensive colour image normalization, in: ECCV'98 Fifth European Conference on Computer Vision, 1998, pp. 475–490.
[23] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, IEEE Trans. Syst. Man Cybern. 3 (6) (1973) 610–621.
[24] G. John, R. Kohavi, K. Pfleger, Irrelevant features and the subset selection problem, in: Proceedings of the Eleventh International Conference on Machine Learning, vol. 129, 1994, pp. 121–129.
[25] Y. Saeys, I. Inza, P. Larrañaga, A review of feature selection techniques in bioinformatics, Bioinformatics 23 (19) (2007) 2507–2517.
[26] K. Pearson, On lines and planes of closest fit to systems of points in space, Philos. Mag. 2 (6) (1901) 559–572.
[27] M. Hall, Correlation-Based Feature Selection for Machine Learning, Ph.D. Thesis, University of Waikato, New-Zealand, 1999.
[28] A. Bouchachia, Incremental Learning, Encyclopedia of Data Warehousing and Mining, 2nd ed., IGI Global, 2009.
[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, I.H. Witten, The weka data mining software: an update, SIGKDD Explor. 11 (1) (2009) 10–18.

Yanyun LU is a Ph.D. student at the Department of Computer Science and Control, Ecole des Mines de Douai, France. She received her B.S. degree in Electronic and Information Engineering in 2007, and M.S. degree in Communication and Information System in 2009, at the Department of Computer and Information Engineering, Hohai University, Jiangsu, China. Her research interests include Image Processing, Machine Learning, Pattern Recognition and Intelligent System.



Khaled Boukharouba received his B.S. degree in Electrical Engineering (with high honors) from the Université des Sciences et Technologies de Lille in 2005 and his M.S. in Automatic, Computer Engineering and Vision (with honors) from the same university in 2007. He obtained his Ph.D. in Automatic Control, Computer Engineering and Signal and Image Processing from the Université des Sciences et Technologies de Lille and Ecole des Mines de Douai in 2011. He is now a post-doc fellow attached to CNRS in SIGMA Laboratory at ESPCI. His research interests include machine learning (SVM, online clustering, Incremental multi-category classification), hybrid system modeling and computer vision.

Jacques Boonært received the Engineer degree in Robotics and Mechanics from the Ecole des Mines de Douai in 1993 and the M.S. degree in Systems Control from University of Compiégne, France, in 1993. He received the Ph.D. degree in Systems Control from University of Compiégne, France, in 1998. He is currently Assistant Professor at the Ecole des Mines de Douai. His work is mainly focused on behaviors and gestures classification from video signal. The aim of this research is to achieve a high level description of the observed scenes so that inferences on the related information are possible.



Anthony Fleury received an Engineer (Computer Science) and a M.Sc. (Signal Processing) degree in 2005 in Grenoble and a Ph.D. degree in Signal Processing from the University Joseph Fourier of Grenoble in 2008 for his work on Health Smart Homes and activity recognition. He joined then the LMAM team at Swiss Federal Institute of Technology and is now, since September 2009, Assistant Professor at Ecole des Mines de Douai. His research interests include the modeling of human behaviors and activities, machine learning and pattern recognition with applications to biomedical engineering.

Stéphane Lecœuche was born in Auchel, France, in 1970. He received the Ph.D. degree from the University of Lille, Lille, France, in 1998. Since 2005, he has been a Professor with the Computer Science and Automatic Control Labs, Mines Douai, Douai, France. His research interests include system identification and dynamical learning applied to the modeling and monitoring of complex evolving systems.

Appendix C

Identification/Réidentification de personnes sur des vidéos grâce des algorithmes de classification incrémentaux. Gretsi, 2013

Identification/Réidentification de personnes sur des vidéos grâce à des algorithmes de classification incrémentaux.

Yanyun LU, Anthony FLEURY, Jacques BOONAERT, Stéphane LECOEUCHE

Unité de Recherche en Informatique et Automatique, École des Mines de Douai
941 Rue Charles Bourseul, CS 10838, 59508 France
yanyun.lu@mines-douai.fr, anthony.fleury@mines-douai.fr
jacques.boonaert@mines-douai.fr, stephane.lecoeuche@mines-douai.fr

Résumé – La reconnaissance de personnes dans le cadre de la vidéo-surveillance est un thème très important de nos jours. Cette communication s'intéresse à la reconnaissance automatique, avec des données qui ne sont pas complètes lors de la phase initiale d'apprentissage, et donc qui nécessite des algorithmes de classification en ligne. Sur une base de 20 personnes présentées avec des orientations différentes et en marchant, nous donnons les résultats de la classification de ces personnes avec tout d'abord un algorithme de SVM incrémental, qui a une connaissance de toutes les classes mais avec un très faible nombre d'éléments pour chaque classe à l'apprentissage puis s'adapte, et dans un second temps avec un algorithme dénommé SAKM, qui va permettre l'apprentissage de nouveautés et donc la découverte de nouvelles personnes arrivant. Nous obtenons des résultats à plus de 95% de bonne classification dans les deux cas, le premier la classification des 20 personnes avec une très faible base d'apprentissage (une première partie de séquence) et dans le second cas avec l'apprentissage uniquement sur trois classes et la découverte de sept classes au total, dont 4 nouvelles.

Abstract – People identification in video-surveillance systems is a problem of interest. This paper deals with automatic recognition of person with incomplete data during training phase requiring on-line training algorithms. In a database including 20 persons walking with different orientation comparing to the camera, we obtained results, first, on classification using Incremental SVM, using at the begining small amount of data belonging to the whole set of classes but adapting the models as the data comes (training set contains a very few number of instances). In a second experiment, we used SAKM algorithm, which can deal with novelty detection. Both experiments gave very good results (over 95% of correct classification) in the first case with the 20 persons and in the second case by training on few images of 3 persons and testing on 7 persons.

1 Introduction

La reconnaissance de personnes est une thématique primordiale dans les systèmes de vidéo-surveillance, de plus en plus utilisés dans les transports et les lieux publics, pour laquelle de nombreuses techniques ont été développées. Ces techniques permettent de détecter une personne et de l'isoler dans l'image, mais aussi de l'identifier [1]. De nombreuses recherches ont émergé ces dernières années au sein de cette thématique. Ces recherches sont, pour un bon nombre, basées sur des caractéristiques biométriques (marche, visage, iris, etc.) [2], mais certaines se basent maintenant sur des caractéristiques d'apparence [3], spécifiquement pour des applications liées à la détection des piétons et plus généralement tous les systèmes d'analyse multi-caméra [4]. L'apparence se base sur les vêtements visibles de la personne qui sont facilement détectables et obtenus par élimination du fond. Pendant une période de temps relativement courte, nous pouvons espérer que l'apparence de la personne soit constante (pas de changement de vêtements). Cependant, les conditions d'illumination et l'orientation de la personne par rapport à la caméra peuvent, elles, changer.

L'apprentissage de toutes les orientations avec toutes les illuminations ne pouvant être réalisé, nous avons précédemment utilisé des techniques d'apprentissage incrémentales basé sur les SVM pour réaliser cette identification de personne [5]. Cela permet d'apprendre sur une courte séquence et de reconnaître la personne avec des orientations ou des illuminations différentes dans d'autres séquences. Cependant, un autre problème émerge alors, il s'agit de pouvoir, dans des endroits par exemple publics, identifier une personne qui est encore inconnue du système (l'apprentissage de nouveautés). Avec les techniques basées sur les SVM incrémentaux, cela est rendu très compliqué car il faudrait trouver une condition sur les données qui soit assez spécifique pour permettre de détecter la différence entre une nouvelle classe et une évolution d'une classe connue et ainsi adapter l'ensemble du SVM avec une nouvelle classe à partir de données inconnues. Pour surmonter cet obstacle, nous avons choisi d'appliquer l'algorithme SAKM (Self-Adaptive Kernel Machine) [9] sur les données images et de comparer avec les résultats des exécutions précédentes utilisant les SVM incrémentaux.

La section 2 s'attardera sur la sélection des attributs que nous présenterons au classifieur. La section 3 présentera les deux algorithmes de classification qui sont comparés, avant que la section 4 ne présente les résultats de l'exécution de ces différents algorithmes. Ces résultats seront discutés dans la section 5 qui introduira également les limitations et futurs travaux.

2 Extraction et selection d'attributs

La base de données CASIA est une base de données vidéo contenant 20 personnes différentes, dans laquelle les personnes sont filmées en marchant selon différentes orientations. Chaque personne marche avec 6 orientations différentes par rapport à la caméra vidéo. Ces personnes n'ont pas changé de vêtements entre les différents essais. Des exemples d'images obtenues sont montrés sur la Fig. 1. Il est à noter que bien que cette base de données soit à l'origine destinée à l'analyse de la marche, elle peut très bien être utilisée dans ce cadre d'identification de personnes. Le fait qu'elle offre différentes illuminations et orientations de la personne sans changement de vêtement nous permet d'évaluer la pertinence et l'efficacité de nos choix algorithmiques.

La base est fournie avec un fond déjà soustrait. En plus des images initiales sont aussi fournies les images représentant les silhouettes des personnes (sous forme de masques). À partir de ces silhouettes, nous avons segmenté le corps en trois différentes parties : la tête, le buste puis les jambes. Afin d'être insensible aux conditions d'illumination, une normalisation a été effectuée sur les couleurs.

Pour la couleur, nous calculons 54 attributs à partir des valeurs moyennes, des écarts types, des histogrammes, etc. sur les trois couleurs de base et les différentes parties du corps prises indépendamment. À ceci s'ajoutent 13 attributs évaluant la texture présente dans les parties de l'image, évaluation qui se fait en niveau de gris et suivant les attributs définis par Haralick [6, 5]. Avec tout ceci, un total de 93 attributs (auxquels nous ajoutons la classe) est alors calculé pour caractériser une personne dans chacune des images.

Afin de sélectionner les informations discriminantes (en enlevant des attributs trop corrélés, trop peu corrélés à la classe ou des attributs ne variant pas assez) dans l'ensemble d'apprentissage, nous avons utilisé trois méthodes de réduction de la dimension, nous permettant d'obtenir trois ensembles d'attributs de taille différente. La première est l'analyse en composantes principales (PCA). Dans cette analyse, nous avons retenu un ensemble de 26 attributs (avec comme critère de réglage la conservation d'au moins 95% de la variance totale). La seconde est une méthode de sélection d'attributs basée sur les corrélations (CFS). Cette méthode se base sur le calcul d'une heuristique [10] :

$$M_s = \frac{k \cdot \overline{r_{cf}}}{\sqrt{k + k \cdot (k-1) \cdot \overline{r_{ff}}}}$$

considérant la corrélation classe-attribut (r_{cf}) ainsi que les corrélations attribut-attribut (r_{ff}) et le nombre d'attributs (k). La finalité était de maximiser cette expression. Afin de rechercher le sous-ensemble optimal (sans parcourir l'ensemble des possibilités qui est bien trop important) d'attributs dans les 93 initiaux, une recherche basée sur du Greedy Hill Climbing bidirectionnel [7] est appliquée avec un critère d'arrêt définissant le nombre de pas sans amélioration possible. L'exécution de cet algorithme nous donne un ensemble de 40 attributs au total.

Enfin, la dernière méthode (Wrapper) se base sur le même algorithme de recherche mais, en lieu et place du calcul d'une heuristique, les performances d'un classifieur donné (ici SVM classique) sont évaluées pour chaque sous-ensemble testé. Le critère est donc de trouver le sous-ensemble pour lequel la classification avec l'algorithme considéré donne les résultats optimaux en termes de bonne classification. Pour cette dernière méthode, 16 attributs sont retenus.

La PCA effectue une projection de l'espace d'attributs de départ dans un nouvel espace dont la base est une combinaison linéaire des précédents vecteurs. L'interprétabilité des résultats n'est donc pas vraiment possible, surtout en dimension 26. Pour ce qui est des deux autres méthodes, les attributs sélectionnés sont des attributs de l'ensemble d'origine. Dans ces deux ensembles, la plupart des attributs sont bien sûr basés sur la couleur. Il est à noter également qu'une grande partie des attributs donnés par la troisième méthode se retrouvent également dans l'ensemble de la seconde méthode (11 attributs sur les 16)

3 Apprentissage par SVM Incrémentaux et SAKM

Le cadre applicatif de ce travail est la reconnaissance de personnes par méthodes de classification. Notre but est d'effectuer l'apprentissage du classifieur sur un faible ensemble de données et d'adapter ce classifieur par la suite pour qu'il puisse reconnaître tout de même la personne même orienté de manière inconnue ou apprendre une nouvelle classe d'individu. Deux algorithmes sont implémentés à cet effet. Le premier, basé sur les SVM incrémentaux, connait l'ensemble des classes au début et apprend au fur et à mesure alors que l'algorithme SAKM adapte l'espace de décision et le nombre de clusters au fil de l'eau. Ces deux algorithmes sont présentés par la suite.

3.1 Apprentissage par SVM Incrémental

La base de cet algorithme est d'entraîner un SVM avec une portion réduite de l'ensemble de données, de conserver uniquement les vecteurs supports et les vecteurs erreurs à chaque étapes du processus d'apprentissage, qui créera un ensemble d'apprentissage pour l'étape suivante.

La classification par SVM est définie par :

$$x_i \in C_k; k = \underset{j=1,\ldots,K}{\arg} \max f_j(x_i)$$

où les fonctions de décision sont $f_i(x) = w_i^T \Phi(x) + b_i$ (avec w_i et b_i définissant les coefficients de la séparatrice et Φ est la fonction noyau). Afin d'obtenir la meilleure marge entre les classes, la méthode des multiplieurs de Lagrange est appliquée et les fonctions de décisions sont réécrites avec les coefficients de Lagrange α_i. Ensuite, en se basant sur les conditions dites

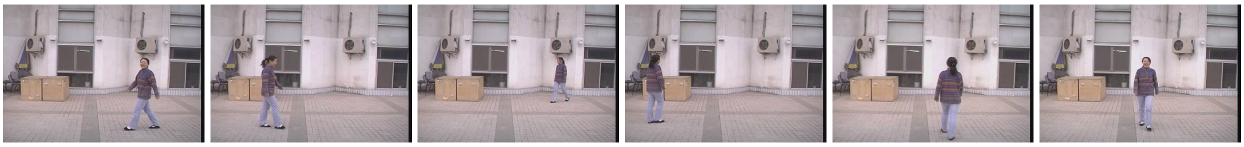


FIGURE 1 – Une personne effectuant les 6 actions de marche différentes dans la base CASIA.

KKT en un point x_m, l'ensemble des données T est divisé en trois catégories en fonction des valeurs de g_m :

$$g_m = \begin{cases} > 0; & if \quad \alpha_m = 0; & \text{Ensemble des vecteurs de données} \quad DV \\ = 0; & if \quad 0 < \alpha_m < C; & \text{Ensemble des vecteurs supports} \quad SV \\ < 0; & if \quad \alpha_m = C; & \text{Ensemble des vecteurs erreurs} \quad EV \end{cases}$$

où $C \geq 0$ permet de paramétrer le nombre d'erreurs admissibles.

L'apprentissage incrémental par SVM repose sur le fait de préserver les conditions KKT sur tous les vecteurs de la base d'entrainement lorsqu'un nouveau vecteur est ajouté de manière adiabatique. La procédure incrémentale est succinctement donnée par l'algorithme 1.

Algorithm 1: Classification en ligne par SVM incrémental [8]

```
Nouvelle donnée $x_c$ ; Initialise $\alpha_c = 0$ ;
if $g_c > 0$ then terminate;
else if $g_c \leq 0$ then
    Applique le plus grand incrément $\alpha_c$ afin que l'une des conditions suivantes se produise;
    (1)$g_c = 0$ : Ajoute $x_c$ aux vecteurs supports $SV$, met à jour la fonction de décision et termine;
    (2)$\alpha_c = C$ : Ajoute $x_c$ aux vecteurs erreurs $EV$ et termine;
    (3)$\alpha_c$ proche de 0, les vecteurs supports correspondant bougent dans la marge et la fonction de décision est mise à jour;
    Puis répéter tant que nécessaire.
end
```

Une description plus précise et complète de l'ensemble de cet algorithme est donnée par [8].

3.2 Self-Adaptive Kernel Machine (SAKM)

SAKM [9] est un algorithme de clustering en ligne permettant d'agir sur des données non-stationnaires dans un contexte multiclasse. L'idée principale est de calculer le degré de proximité d'un nouvel échantillon avec les plus proches vecteurs supports de chacun des clusters dans l'espace de données, ceci avec une métrique induite par le noyau. Une nouvelle fonction de similarité est introduite en considérant $\mu\phi_{t,m}$. Dans l'espace de Hilbert défini par un noyau Gaussien, la distance d'un nouvel échantillon X_t par rapport à chaque cluster C_m^t au temps t s'exprime par :

$$\begin{aligned} \mu\phi_{t,m} &= \frac{\delta}{\sqrt{2}} \left\| \phi(X_t) - \phi(SV_{win,m}) \right\|_\Gamma \\ &= \delta \times \sqrt{1 - \kappa(X_t, SV_{win,m})} \\ &= \delta \times \sqrt{1 - \exp(-\lambda \left\| X_t - SV_{win,m} \right\|)} \end{aligned}$$

où $SV_{win,n}$ est le vecteur support gagnant du cluster C_m^t et δ est une fonction signe égal à 1 si l'évaluation de la fonction définissant la marge en X_t est négative.

Lorsqu'une nouvelle donnée X_t est présentée, la procédure à suivre (création, adaptation, fusion) est sélectionnée, comme montré dans l'algorithme 2, suivant le critère :

$$\Omega^{win} = \left\{ C_m^t \in \Omega^t / \mu\phi(X_t, C_m^t) \leq \varepsilon_{th} \right\}$$

où ε_{th} est un paramètre d'acceptabilité d'un cluster.

Algorithm 2: Classification en ligne avec SAKM [9]

```
Récupération des données en ligne $X$ ; Configuration des paramètres et seuils;
Initialisation des fonctions de décision et des paramètres des noyaux avec les données d'apprentissage.;
while Nouvelles données disponibles do
    Évaluation de la fonction de similarité basée sur les noyaux : $\mu\phi_{t,m}$ ; Critère de similarité des clusters : $\Omega^{win}$;
    if $card(\Omega^{win} = 0)$ then Procédure de création;
    else if $card(\Omega^{win} = 1)$ then Procédure d'adaptation;
    else $card(\Omega^{win} \geq 2)$Procédure de fusion;
    if Nombre de données dans les clusters inférieur au seuil then Procédure d'élimination
end
```

4 Résultats

Pour tester la méthode basée sur les SVM incrémentaux, 10% des images de la base de données sont réservés à l'apprentissage, et le reste est utilisé pour tester en mettant à jour la base de connaissance. En effet, à chaque test d'un élément, ce même élément est ensuite inséré comme élément d'apprentissage avec comme classe celle qui a été décidée. Les résultats sont montrés dans le tableau 1. Les résultats sont très bons et très proches des 99.5% donnés par les SVM avec un apprentissage plus classique (60% de la base de départ) et sans incrémentation. Cependant, cette méthode, aussi bons soient ses résultats, ne permet pas de découvrir de nouvelles classes.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	C11	C12	C13	C14	C15	C16	C17	C18	C19	C20	Global
Wholeset	92.16	99.44	100	95.77	99.10	99.37	100	100	100	100	99.78	99.31	98.98	97.34	100	96.87	98.23	91.89	97.42	99.22	98.21
PCA	91.07	99.72	98.74	95.37	99	92.15	93.65	99.76	100	100	99.23	96.11	94	94.42	91.69	97.74	98.11	85.70	85.77	100	95.6
CFS	92.59	99.44	100	97.99	99.10	99.79	100	100	100	100	99.78	98.28	99.39	97.72	100	98.62	98.23	92	96.56	100	98.46
Wrapper	91.83	99.44	99.58	91.15	99.10	92.25	99.08	99.88	100	100	96.92	95.65	86.98	93.03	94.16	95.11	97.99	83.88	93.37	99.67	95.39

TABLE 1 – Résultats de classification avec les SVM incrémentaux

Le second tableau présente les résultats avec l'algorithme SAKM. Pour ces résultats, trois classes sont au départ considérées comme connues. Pour ces trois classes, 10% des données sont prises (en début de la première séquence) pour former une base d'apprentissage et apprendre trois clusters. Une fois cet apprentissage réalisé, le reste des éléments est utilisé pour l'apprentissage en ligne. Afin de tester la possibilité de découverte de nouveautés, nous avons ainsi ajouté à cet ensemble de test quatre autres classes qui n'ont jamais été apprises. Le but étant que l'algorithme puisse détecter ces nouveautés afin de créer automatiquement une nouvelle classe. La dernière colonne du tableau nous indique le nombre total de cluster trouvé dans les données. Une étape de fusion est ensuite nécessaire pour rassembler les clusters appartenant à une même personne. À noter aussi que pour cette expérimentation nous avons choisi aléatoirement 7 individus dans le set de données. Ainsi, P1 ne correspond pas forcément à C1 du premier tableau. Les hyper-paramètres des deux classifieurs ont été optimisés.

	P1	P2	P3	P4	P5	P6	P7	Nb Cluster
Wholeset	100	100	97.48	97.53	95.34	90	100	23
PCA	100	100	95.27	98.02	88.34	84.33	100	16
CFS	100	100	99.26	100	96.64	90.83	100	16
Wrapper	100	100	97.69	100	95.96	89.17	100	9

TABLE 2 – Résultats de classification avec SAKM et la découverte de nouveautés (3 classes apprises et 7 testées).

Les deux résultats sont très bons et se rapprochent de ceux de la classification non incrémentale bien qu'un nombre plus important de données d'apprentissage ait été utilisé. Nous avons ensuite conduit une seconde expérimentation avec l'algorithme non-incrémental dans laquelle les données présentées pour l'apprentissage sont limitées à celles présentées initialement dans l'algorithme incrémental. Nous observons une diminution significative du pourcentage de bonne classification sous les 85%.

5 Discussion et Conclusion

Les deux techniques d'apprentissage en ligne présentées nous montrent qu'en identification de personnes, nous arrivons à atteindre des résultats proches des algorithmes de classification habituels avec un apprentissage initial plus léger. La seconde méthode nous permet même de ne pas avoir la connaissance originelle du nombre de personnes et des personnes à classer.

Ces résultats sont cependant à approfondir pour plusieurs raisons. La première est que, pour les algorithmes en ligne, le résultat final dépend logiquement de l'ordre de présentation des images. Il faut ainsi déterminer l'impact sur les résultats d'un ordre d'arrivée des images en adéquation avec un système de vidéo-surveillance multi-caméra dans un endroit public donné. De plus, le réglage des algorithmes est relativement sensible et la généralisation à un système de vidéo-surveillance en conditions réelles n'est pas triviale, notamment il sera nécessaire de faire un meilleur usage des procédures d'oubli.

Références

[1] D. Truong Cong, L. Khoudour, C. Achard, C. Meurie, O. Lezoray, People re-identification by spectral classification of silhouettes, Signal Processing 90 (8) (2010) 2362–2374.

[2] X. Zhou, B. Bhanu, Feature fusion of side face and gait for video-based human identification, Pattern Recognition 41 (3) (2008) 778 – 795.

[3] D. Makrisa, N. Doulamisc, S. Middletond, Vision-Based Production of Personalized Video, Signal Processing : Image Computation 24 (5) (2009) 158–176.

[4] D. Truong Cong, L. Khoudour, C. Achard, L. Douadi, People Detection and Re-Identification in Complex Environments, IEICE Transactions on Information and Systems 93 (7) (2010) 1761–1772.

[5] Y. Lu, K. Boukharouba, J. Boonaert, A. Fleury, S. Lecœuche, Application of an Incremental SVM algorithm for On-line human recognition from video surveillance using texture and color features, Neurocomputing (2013) In Press. DOI :10.1016/j.neucom.2012.08.071.

[6] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, Systems, Man and Cybernetics, IEEE Transactions on 3 (6) (1973) 610–621.

[7] J. Pearl, Heuristics : Intelligent Search Strategies for Computer Problem Solving, Addison-Wesley, 1984.

[8] K. Boukharouba, L. Bako, S. Lecoeuche, Incremental and Decremental Multi-category Classification by Support Vector Machines, in : 2009 International Conference on Machine Learning and Applications, IEEE, 2009, pp. 294–300.

[9] H.A. Boubacar, S. Lecoeuche, S. Maouche, SAKM : Self-adaptive kernel machine A kernel-based algorithm for online clustering, Neural Networks 21 (9) (2008) 1287–1301.

[10] M. Hall, Correlation-based feature selection for machine learning, Ph.D. thesis, University of Waikato, New-Zealand (1999).

Bibliography

[AABC04] A. Arauzo-Azofra, J.M. Benitez, and J.L. Castro. A feature set measure based on relief. In *Proceedings of the fifth international conference on Recent Advances in Soft Computing*, pages 104–109, 2004.

[ANRS07] A.F. Abate, M. Nappi, D. Riccio, and G. Sabatino. 2d and 3d face recognition: A survey. *Pattern Recognition Letters*, 28(14):1885–1906, October 2007.

[ASK08] S. Agarwal, V.V. Saradhi, and H. Karnick. Kernel-based online machine learning and support vector reduction. *Neurocomputing*, 71(7):1230–1237, 2008.

[Bąk12] S. Bąk. *Human re-identification through a video camera network.* PhD thesis, Université de Nice, 2012.

[BB99] E.J. Bredensteiner and K.P. Bennett. Multicategory classification by support vector machines. *Computational Optimization and Applications*, 12(1):53–79, 1999.

[BBL09] K. Boukharouba, L. Bako, and S. Lecoeuche. Incremental and decremental multi-category classification by Support Vector Machines. In *2009 International Conference on Machine Learning and Applications*, pages 294–300, 2009.

[BCBT10] S. Bąk, E. Corvee, F. Brémond, and M. Thonnat. Person re-identification using spatial covariance regions of human body parts. In *Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on*, pages 435–440, 2010.

[BDSP07] R. Bodor, A. Drenner, P. Schrater, and N. Papanikolopoulos. Optimal camera placement for automated surveillance tasks. *Journal of Intelligent and Robotic Systems*, 50(3):257–295, 2007.

[BGV92] B.E. Boser, I.M. Guyon, and V.N. Vapnik. A training algorithm for optimal margin classifiers. In *Proceedings of the fifth annual workshop on Computational learning theory*, pages 144–152, 1992.

[BHK97] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):711–720, 1997.

[BL02] H. Byun and S.W. Lee. Applications of support vector machines for pattern recognition: A survey. *Pattern Recognition with Support Vector Machines*, pages 571–591, 2002.

[BL08] K. Boukharouba and S. Lecoeuche. Online clustering of non-stationary data using incremental and decremental svm. In *Artificial Neural Networks-ICANN 2008*, pages 336–345, 2008.

[BLM05] H.A. Boubacar, S. Lecoeuche, and S. Maouche. Self-adaptive kernel machine: online clustering in rkhs. In *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*, volume 3, pages 1977–1982, 2005.

[BM10] N. Bhoi and M.N. Mohanty. Template matching based eye detection in facial image. *International Journal of Computer Applications (0975–8887) Volume*, 2010.

[BMP04] R. Bodor, R. Morlok, and N. Papanikolopoulos. Dual-camera system for multi-level activity recognition. In *Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, volume 1, pages 643–648, 2004.

[BMS02] M.S. Bartlett, J.R. Movellan, and T.J. Sejnowski. Face recognition by independent component analysis. *Neural Networks, IEEE Transactions on*, 13(6):1450–1464, 2002.

[Bor03] S. Borer. *New support vector algorithms for multicategorical data applied to real-time object recognition.* PhD thesis, École Polytechnique Fédérale de Lausanne, 2003.

[Bou09] R. Bouckaert. A hierarchical face recognition algorithm. *Advances in Machine Learning*, pages 38–50, 2009.

[Bou11] K. Boukharouba. *Modélisation et classification de comportements dynamiques des systemes hybrides*. PhD thesis, Phd Thesis, Université de Lille, France2011, 2011.

[BP93] R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(10):1042–1052, October 1993.

[BRL$^+$11] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool. Online multiperson tracking-by-detection from a single, uncalibrated camera. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(9):1820–1833, 2011.

[BS87] B Brumback and M Srinath. A chi-square test for fault-detection in kalman filters. *Automatic Control, IEEE Transactions on*, 32(6):552–554, 1987.

[Bur98] C.J.C. Burges. A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2(2):121–167, 1998.

[CAS01] CASIA Gait Database. Downloadable on the internet at : http://www.sinobiometrics.com, 2001.

[CGS12] J.F. Connolly, E. Granger, and R. Sabourin. An adaptive classification system for video-based face recognition. *Information Sciences*, 192:50–70, 2012.

[CGY96] I.J. Cox, J. Ghosn, and P.N. Yianilos. Feature-based face recognition using mixture-distance. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 209–216, 1996.

[CJ11] W. Cheng and C. Juang. An incremental support vector machine-trained ts-type fuzzy system for online classification problems. *Fuzzy Sets and Systems*, 163(1):24–44, 2011.

[Cla02] D.A. Clausi. An analysis of co-occurrence texture statistics as a function of grey level quantization. *Canadian Journal of remote sensing*, 28(1):45–62, 2002.

[CMBT03] N. Chikahito, P. Massimiliano, H. Bernd, and P. Tomaso. Full-body person recognition system. *Pattern Recognition*, 36(9):1997–2006, 2003.

[CP01] G. Cauwenberghs and T. Poggio. Incremental and decremental support vector machine learning. page 409, 2001.

[CST00] N. Cristianini and J. Shawe-Taylor. *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press, 2000.

[CV95] C. Cortes and V. Vapnik. Support-vector networks. *Machine learning*, 20(3):273–297, 1995.

[DC03] C.P. Diehl and G. Cauwenberghs. Svm incremental learning, adaptation and optimization. In *Neural Networks, 2003. Proceedings of the International Joint Conference on*, volume 4, pages 2685–2690, 2003.

[DL97] M. Dash and H. Liu. Feature selection for classification. *Intelligent data analysis*, 1(1-4):131–156, 1997.

[DSTR11] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher. Appearance-based person reidentification in camera networks: problem overview and current approaches. *Journal of Ambient Intelligence and Humanized Computing*, 2(2):127–151, 2011.

[DT05] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 886–893, 2005.

[FBP$^+$10] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2010)*, San Francisco, CA, USA, 2010. IEEE Computer Society.

[FGH10] Y. Fu, G. Guo, and T.S. Huang. Age synthesis and estimation via faces: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2010.

[FGMR13] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Visual object detection with deformable part models. *Communications of the ACM*, 56(9):97–105, 2013.

[FHST05] G. Finlayson, S. Hordley, G. Schaefer, and G.Y. Tian. Illuminant and device invariant colour using histogram equalisation. *Pattern Recognition*, 38:2005, 2005.

[Fri96] J. Friedman. Another approach to polychotomous classifcation. Technical report, Technical report, Stanford University, Department of Statistics, 1996.

[FRP07] Li F., F. Rob, and P. Pietro. Learning generative visual models from few training examples: An incremental bayesian approach tested on 101 object categories. *Computer Vision and Image Understanding*, 106(1):59–70, 2007.

[FSC98] G.D. Finlayson, B. Schiele, and J.L. Crowley. Comprehensive colour image normalization. *Computer Vision—ECCV'98*, pages 475–490, 1998.

[FYN$^+$12] G. Fanelli, A. Yao, P.L. Noel, J. Gall, and L. Van Gool. Hough forest-based facial expression recognition from video sequences. In *Trends and Topics in Computer Vision*, pages 195–206. Springer, 2012.

[GK86] P. Geladi and B.R. Kowalski. Partial least-squares regression: a tutorial. *Analytica chimica acta*, 185:1–17, 1986.

[GKVL06] A.B. Gardner, A.M. Krieger, G. Vachtsevanos, and B. Litt. One-class novelty detection for seizure analysis from intracranial eeg. *The Journal of Machine Learning Research*, 7:1025–1044, 2006.

[GPK02] S.E. Grigorescu, N. Petkov, and P. Kruizinga. Comparison of texture features based on gabor filters. *Image Processing, IEEE Transactions on*, 11(10):1160–1167, 2002.

[GS01] R. Gross and J. Shi. The cmu motion of body (mobo) database. 2001.

[GSH06] N. Gheissari, T.B. Sebastian, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 1528–1535, 2006.

[GT08] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. *Computer Vision–ECCV 2008*, pages 262–275, 2008.

[HAA07] M.T. Harandi, M.N. Ahmadabadi, and B.N. Araabi. A Hierarchical Face Identification System Based on Facial Components. pages 669–675, 2007.

[Hal99] M.A. Hall. *Correlation-based feature selection for machine learning*. PhD thesis, Citeseer, 1999.

[Ham11] L. Hamoudi. *Application de techniques d'apprentissage pour la détection et la reconnaissance d'individus*. PhD thesis, Lille 1, 2011.

[HAN02] J.B. Hayfron-Acquah and J.N. Nixon, M.S.and Carter. Human identification by spatio-temporal symmetry. In *Pattern Recognition, 2002. Proceedings. 16th International Conference on*, volume 1, pages 632–635, 2002.

[HBL10] L. Hamoudi, J. Boonaert, and S. Lecoeuche. Appearance-based person recognition using clothes classification. In *IPCV'10*, pages 415–421, 2010.

[HGV13] M. Hanmandlu, D. Gupta, and S. Vasikarla. Face recognition using elastic bunch graph matching. In *Applied Imagery Pattern Recognition Workshop: Sensing for Control and Augmentation, 2013 IEEE (AIPR)*, pages 1–7, 2013.

[HHN99] P.S. Huang, C.J. Harris, and M.S. Nixon. Human gait recognition in canonical space using temporal templates. In *Vision, Image and Signal Processing, IEE Proceedings-*, volume 146, pages 93–100, 1999.

[HJLQ11] H. Hu, S. Jiang, Z. Li, and Z. Qu. Abnormal pedestrian crossing detection based on video processing. In *11th International Conference of Chinese Transportation Professionals (ICCTP)*, 2011.

[HKK04] M. Hahnel, D. Klunder, and K.F. Kraiss. Color and texture features for person recognition. In *Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on*, volume 1, 2004.

[HL02] C.W. Hsu and C.J. Lin. A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, 13(2):415–425, 2002.

[HLL07] E. Hörster, J. Lux, and R. Lienhart. Recognizing persons in images by learning from videos. In *Proceedings of SPIE*, volume 6506, page 65060D, 2007.

[HMD13] T. Huynh, R. Min, and J.L. Dugelay. An efficient lbp-based descriptor for facial depth images applied to gender recognition using rgb-d face data. In *Computer Vision-ACCV 2012 Workshops*, pages 133–145, 2013.

[HP06] M. Heikkila and M. Pietikainen. A texture-based method for modeling the background and detecting moving objects. *IEEE transactions on pattern analysis and machine intelligence*, 28(4):657–662, 2006.

[HSD73] R.M. Haralick, K. Shanmugam, and I. Dinstein. Textural features for image classification. *Systems, Man and Cybernetics, IEEE Transactions on*, 3(6):610–621, 1973.

[IML$^+$01] I. Inza, M. Merino, P. Larranaga, J. Quiroga, B. Sierra, and M. Girala. Feature subset selection by genetic algorithms and estimation of distribution algorithms: A case study in the survival of cirrhotic patients treated with tips. *Artificial Intelligence in Medicine*, 23(2):187–205, 2001.

[JA09] R. Jafri and H.R. Arabnia. A survey of face recognition techniques. *journal of Information Processing Systems*, 5, 2009.

[JKP94] G.H. John, R. Kohavi, and K. Pfleger. Irrelevant features and the subset selection problem. In *Proceedings of the eleventh international conference on machine learning*, volume 129, pages 121–129, 1994.

[Jol05] I. Jolliffe. *Principal component analysis*. Wiley Online Library, 2005.

[Kan74] T. Kanade. Picture processing system by computer complex and recognition of human faces. 1974.

[KDM$^+$09] D.I. Kosmopoulos, A. Doulamis, A. Makris, N. Doulamis, S. Chatzis, and S.E. Middleton. Vision-based production of personalized video. *Image Commun.*, 24:158–176, March 2009.

[KHM97] T. Kurita, K. Hotta, and T. Mishima. Scale and rotation invariant recognition method using higher-order local autocorrelation features of log-polar image. In *Computer Vision—ACCV'98*, pages 89–96. Springer, 1997.

[KHST12] M. Karasuyama, N. Harada, M. Sugiyama, and I. Takeuchi. Multi-parametric solution-path algorithm for instance-weighted support vector machines. *Machine learning*, 88(3):297–330, 2012.

[KJ97] R. Kohavi and G.H. John. Wrappers for feature subset selection. *Artificial intelligence*, 97(1):273–324, 1997.

[KLJ11] B. Klare, Z. Li, and A.K. Jain. Matching forensic sketches to mug shot photos. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):639–646, 2011.

[KP07] K.C. Kwak and W. Pedrycz. Face recognition using an enhanced independent component analysis approach. *Neural Networks, IEEE Transactions on*, 18(2):530–541, 2007.

[KPD$^+$90] S. Knerr, L. Personnaz, G. Dreyfus, J. Fogelman, A. Agresti, M. Ajiz, A. Jennings, F. Alizadeh, F. Alizadeh, J. Haeberly, et al. Single-layer learning revisited: A stepwise procedure for building and training a neural network. *Optimization Methods and Software*, 1:23–34, 1990.

[KS96] D. Koller and M. Sahami. Toward optimal feature selection. 1996.

[KSW04] J. Kivinen, A.J. Smola, and R.C. Williamson. Online learning with kernels. *Signal Processing, IEEE Transactions on*, 52(8):2165–2176, 2004.

[KT10] M. Karasuyama and I. Takeuchi. Multiple incremental decremental learning of support vector machines. *Neural Networks, IEEE Transactions on*, 21(7):1048–1059, 2010.

[LD08] Z. Lin and L. Davis. Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance. *Advances in Visual Computing*, pages 23–34, 2008.

[LHLM02] Q. Liu, R. Huang, H. Lu, and S. Ma. Face recognition using kernel-based fisher discriminant analysis. In *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*, pages 197–201, 2002.

[LHTX03] K. Li, H. Huang, S. Tian, and W. Xu. Improving one-class svm-based for anomaly detection. In *Machine Learning and Cybernetics, 2003 International Conference on*, volume 5, pages 3077–3081, 2003.

[LL09] Z. Liang and Y.F. Li. Incremental support vector machine learning in the primal and applications. *Neurocomputing*, 72(10-12):2249–2258, 2009.

[Low04] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60(2):91–110, 2004.

[LS04] Z. Liu and S. Sarkar. Simplest representation yet for gait recognition: Averaged silhouette. In *Pattern Recognition, ICPR 2004. Proceedings of the 17th International Conference on*, volume 4, pages 211–214, 2004.

[LW02] C. Liu and H. Wechsler. Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition. *Image processing, IEEE Transactions on*, 11(4):467–476, 2002.

[Mey97] D. Meyer. Human gait classification based on hidden markov models. In *3D Image Analysis and Synthesis*, volume 97, pages 139–146, 1997.

[MLS12] B. Meden, F. Lerasle, and P. Sayd. Mcmc supervision for people re-identification in nonoverlapping cameras. In *BMVC*, pages 1–11, 2012.

[MM96] B.S. Manjunath and W. Ma. Texture features for browsing and retrieval of image data. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 18(8):837–842, 1996.

[MR03] G.L. Marcialis and F. Roli. Fusion of face recognition algorithms for video-based surveillance systems. pages 235–249, 2003.

[MTP03] J. Ma, J. Theiler, and S. Perkins. Accurate on-line support vector regression. *Neural Computation*, 15(11):2683–2703, 2003.

[MY02a] L.M. Manevitz and M. Yousef. One-class svms for document classification. *The Journal of Machine Learning Research*, 2:139–154, 2002.

[MY02b] B. Moghaddam and M.H. Yang. Learning gender with support faces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):707–711, 2002.

[MZBH09] D. Ming, C. Zhang, B. Bai, Y.and Wan, and K.D.K. Hu, Y.and Luk. Gait recognition based on multiple views fusion of wavelet descriptor and human skeleton model. In *Virtual Environments, Human-Computer Interfaces and Measurements Systems, 2009. VECIMS'09. IEEE International Conference on*, pages 246–249, 2009.

[NCC+02] M.S. Nixon, J.N. Carter, D Cunado, P.S. Huang, and S.V. Stevenage. Automatic gait recognition. *Biometrics*, pages 231–249, 2002.

[PD06] S. Paris and F. Durand. A fast approximation of the bilateral filter using a signal processing approach. In *Computer Vision–ECCV 2006*, pages 568–580. Springer, 2006.

[PHS11] U. Prabhu, J. Heo, and M. Savvides. Unconstrained pose-invariant face recognition using 3d generic elastic models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(10):1952–1961, 2011.

[PJK+06] U. Park, A.K. Jain, I. Kitahara, K. Kogure, and N. Hagita. Vise: Visual search engine using multiple networked cameras. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, volume 3, pages 1204–1207, 2006.

[PJW00] D.K. Park, Y.S. Jeon, and C.S. Won. Efficient use of local edge histogram descriptor. In *Proceedings of the 2000 ACM workshops on Multimedia*, pages 51–54, 2000.

[PNK94] P. Pudil, J. Novovičová, and J. Kittler. Floating search methods in feature selection. *Pattern recognition letters*, 15(11):1119–1125, 1994.

[PRJ13] J.M. Pandya, D. Rathod, and J.J. Jadav. A survey of face recognition approach. *International Journal of Engineering Research and Applications (IJERA)*, 3(1):632–635, 2013.

[PSR+02] P.J. Phillips, S. Sarkar, I. Robledo, P. Grother, and K. Bowyer. Baseline results for the challenge problem of humanid using gait analysis. In *Automatic Face and Gesture Recognition, Proceedings. Fifth IEEE International Conference on*, pages 130–135, 2002.

[Ros03] R. Rosipal. Kernel partial least squares for nonlinear regression and discrimination. *Neural Network World*, 13(3):291–300, 2003.

[Rüp01] S. Rüping. Incremental learning with support vector machines. pages 641–642, 2001.

[RWZD07] Y. Ran, I. Weiss, Q. Zheng, and L.S. Davis. Pedestrian detection via periodic motion analysis. *International Journal of Computer Vision*, 71(2):143–160, 2007.

[SC00] B. Schiele and J.L. Crowley. Recognition without correspondence using multidimensional receptive field histograms. *International Journal of Computer Vision*, 36(1):31–50, 2000.

[SD09] W.R. Schwartz and L.S. Davis. Learning discriminative appearance-based models using partial least squares. In *Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on*, pages 322–329, 2009.

[SHH09] R. Senaratne, S. Halgamuge, and A. Hsu. Face recognition by extending elastic bunch graph matching with particle swarm optimization. *Journal of Multimedia*, 4(4), 2009.

[SHKS99] N.A. Syed, S. Huan, L. Kah, and K. Sung. Incremental learning with support vector machines, 1999.

[SIL07] Y. Saeys, I. Inza, and P. Larrañaga. A review of feature selection techniques in bioinformatics. *Bioinformatics*, 23(19):2507–2517, 2007.

[SL13] R. Sakurai and J.H. Lee. Classification based person identification in group living environment. In *System Integration (SII), 2013 IEEE/SICE International Symposium on*, pages 665–670, 2013.

[Sou13] M. Souded. *People detection, tracking and re-identification through a video camera network.* PhD thesis, Université de Nice, 2013.

[SPL$^+$05] S. Sarkar, P.J. Phillips, Z. Liu, I.R. Vega, P. Grother, and K.W. Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(2):162–177, 2005.

[SPST+01] B. Schölkopf, J.C. Platt, J. Shawe-Taylor, A.J.S. Smola, and R.C. Williamson. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7):1443–1471, 2001.

[SR03] A. Sundaresan and R. RoyChowdhury, A.and Chellappa. A hidden markov model based framework for recognition of humans from gait sequences. In *Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on*, volume 2, pages II–93, 2003.

[SSM07] A. Samal, V. Subramani, and D. Marx. Analysis of sexual dimorphism in human face. *Journal of Visual Communication and Image Representation*, 18(6):453–463, 2007.

[SWZ+09] J. Suo, T. Wu, S. Zhu, S. Shan, X. Chen, and W. Gao. Design sparse features for age estimation using hierarchical face model. pages 1–6, 2009.

[TB01] R. Tanawongsuwan and A. Bobick. Gait recognition from time-normalized joint-angle trajectories in the walking plane. In *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, volume 2, pages II–726, 2001.

[TCA10] L. Truong Cong, D.and Khoudour and C. Achard. People reacquisition across multiple cameras with disjoint views. *Image and Signal Processing*, pages 488–495, 2010.

[TCAKD09] D.N. Truong Cong, C. Achard, L. Khoudour, and L. Douadi. Video sequences association for people re-identification across multiple non-overlapping cameras. *Image Analysis and Processing–ICIAP 2009*, pages 179–189, 2009.

[TCKA+10] D.N. Truong Cong, L. Khoudour, C. Achard, C. Meurie, and O. Lezoray. People re-identification by spectral classification of silhouettes. *Signal Processing*, 90(8):2362–2374, 2010.

[TCKAD10] D.N. Truong Cong, L. Khoudour, C. Achard, and L. Douadi. People detection and re-identification in complex environments. *IEICE Transactions on Information and Systems*, 93(7):1761–1772, 2010.

[TL03] D.M.J. Tax and P. Laskov. Online svm learning: from classification to data description and back. In *Neural Networks for Signal*

Processing, 2003. NNSP'03. 2003 IEEE 13th Workshop on, pages 499–508, 2003.

[TP91] M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of cognitive neuroscience*, 3(1):71–86, 1991.

[Vap00] V.N. Vapnik. *The nature of statistical learning theory*. Springer Verlag, 2000.

[VLBB08] U. Von Luxburg, M. Belkin, and O. Bousquet. Consistency of spectral clustering. *The Annals of Statistics*, pages 555–586, 2008.

[WFKvdM97] L. Wiskott, J.M. Fellous, N. Kuiger, and C. von der Malsburg. Face recognition by elastic bunch graph matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):775–779, 1997.

[WL11] Z. Wang and S. Li. Face recognition using skin color segmentation and template matching algorithms. *Information Technology Journal*, 10(12), 2011.

[WTNH03] L. Wang, T. Tan, H. Ning, and W. Hu. Silhouette analysis-based gait recognition for human identification. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 25(12):1505–1518, 2003.

[XCL$^+$10] J. Xu, W. Cong, J. Li, L. Wang, L. Li, and H. Liang. Gait recognition based on key frame and elliptical model. In *Information and Automation (ICIA), 2010 IEEE International Conference on*, pages 2483–2487, 2010.

[YH98] J. Yang and V. Honavar. Feature subset selection using a genetic algorithm. *Intelligent Systems and Their Applications, IEEE*, 13(2):44–49, 1998.

[YHD06] K. Yoon, D. Harwood, and L. Davis. Appearance-based person recognition using color/path-length profile. *Journal of Visual Communication and Image Representation*, 17(3):605–622, 2006.

[YKR08] M. Yang, K. Kpalma, and J. Ronsin. A survey of shape feature extraction techniques. *Pattern Recognition*, pages 43–90, 2008.

[YL03] L. Yu and H. Liu. Feature selection for high-dimensional data: A fast correlation-based filter solution. In *International Conference on Machine Learning*, volume 20, page 856, 2003.

[YNC02] C.Y. Yam, M.S. Nixon, and J.N. Carter. Gait recognition by walking and running: a model-based approach. 2002.

[YO07] J. Yao and J.M. Odobez. Multi-layer background subtraction based on color and texture. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8, 2007.

[Yui91] A.L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991.

[ZB08] X. Zhou and B. Bhanu. Feature fusion of side face and gait for video-based human identification. *Pattern Recognition*, 41(3):778 – 795, 2008. Part Special issue: Feature Generation and Machine Learning for Robust Multimodal Biometrics.

[ZCPR03] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *Acm Computing Surveys (CSUR)*, 35(4):399–458, 2003.

Abstract

Video surveillance is nowadays an important topic to address, as it is broadly used for security and it brings problems related to big data processing. A part of it is identification and re-identification of persons in multicamera environments. The objective of this thesis work is to design a complete automatic appearance-based human recognition system working in real-life environment, with the goal to achieve two main tasks: person re-identification and new person discovery. The proposed system consists of four modules: video data acquisition; background extraction and silhouette extraction; feature extraction and selection; and person recognition. For evaluation purposes, in addition to the public available CASIA Database, a more challenging new database has been created under low constraints. Grey-world normalized color features and Haralick texture features are extracted as initial feature subset, then features selection approaches are tested and compared. These optimized subsets of features are then used firstly for person re-identification using Multi-category Incremental and Decremental SVM (MID-SVM) algorithm with the advantage of training only with few initial images and secondly for person discovery and classification using Self-Adaptive Kernel Machine (SAKM) algorithm able to differentiate existing persons who can be classified from new persons who have to be learned and added. The proposed system succeed in person re-identification with classification rate of over 95% and achieved satisfying performances on person discovery with accuracy rate of over 90%.