Application of an Incremental SVM Algorithm for On-line Human Recognition from Video Surveillance Using Texture and Color Features

Yanyun Lu, Khaled Boukharouba, Jacques Boonaert, Anthony Fleury,
Stéphane Lecoeuche

Application of an Incremental SVM Algorithm for On-line Human Recognition from Video Surveillance Using Texture and Color Features

Yanyun Lu, Khaled Boukharouba, Jacques Boonært, Anthony Fleury*, Stéphane Lecœuche

Univ Lille Nord de France, F-59000 Lille, France
EMDouai, IA, F-59500 Douai, France

Abstract

The goal of the 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 periodrely on only 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 has been 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.

Keywords: Video surveillance, Human recognition, Incremental Support Vector Machine, On-line multiclass classification

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, 3, 4, 5], but also take into account non-biometric features (appearance) [6, 7, 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 camera for which the illumination is different or if the person is moving and, as a consequence, changes of orientation. Even a very huge static database of people images can not 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 (environmental, 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. In addition, in order to get better performances in classification, the off-line approaches require a large amount of training images, are time-consuming and cannot incorporate prior information into the learning process. However, in our case, the human recognition problem, the classes properties are time-varying and the system should have the ability to deal with large scale dynamic data because during long observations, the complete description of the class (the person) is not known. On-line methods are particularly useful in the situations that involve 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 contact information: Département Informatique et Automatique, École des Mines de Douai, 941, Rue Charles Bourseul, C.S. 10838 - 59508 Douai, France, Tel.: +33(0)3 27 71 2381; Fax: +33(0)3 27 71 2917/2980.

Email addresses: yanyun.lu@mines-douai.fr (Yanyun Lu), khaled.boukharouba@mines-douai.fr (Khaled Boukharouba), jacques.boonaert@mines-douai.fr (Jacques Boonært), anthony.fleury@mines-douai.fr (Anthony Fleury), stephane.lecoeuche@mines-douai.fr (Stéphane Lecœuche)

The work of Syed et al. in 1999 [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 apply it to solve human recognition problems in a network of camera.

The main contribution of this paper is the application of incremental SVM techniques to the problem of identification/re-identification of person 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 these parameters and then selection of the features and it ends with the incremental classification of these images. Tests have been run on a collection of almost 20000 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 details 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 train 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 for our time-varying problem. As the algorithm is incremental, each decision taken for a new frame will be used to update the SVM classifier (even if this decision is wrong). 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, 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 performances were compared to the one of the classical SVM (without incremental learning). The structure of our system is described by Fig. 1.
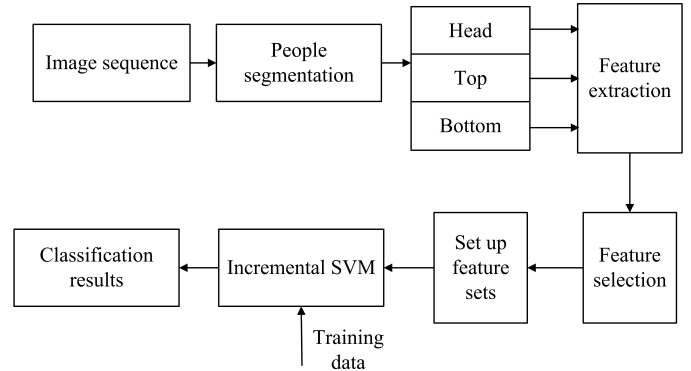

Figure 1: The structure of proposed human recognition system

Figure 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 design to recognize gait parameters in video images. However, for our application, and even if we are not interested by 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

2

for every people, depending on the orientation, the statistics on the clothes can be different. Testing on this dataset allow to test in the same condition than 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 adapt to the different view of a person in movement. That is the reason why, before collecting and creating our own and important database, we tested with this one, that is freely and publicly available, and with which all the results of the article can be verified and reproduced.

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

Table 1: The distribution of 20 classes within the 19135 images in CASIA gait database

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.
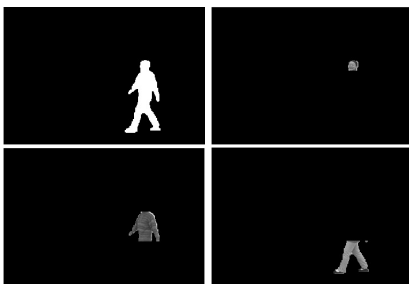

Figure 3: The silhouette picture and the three parts of the body that have been considered

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 and describe. 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 conditions, 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 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. We consider the segmentation of the body in three parts and convert them in grey levels. Then a one-dimension matrix is obtained to represent regions, with values between 0 (black) and 255 (white). 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

3

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

Table 2: The initial features based on color and texture

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 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 method) and embedded methods (closed-loop method - that can also 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 attribute. 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 19135 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 M. A. 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 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 will 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.

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

Table 3: The features selected by each data set

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 looses 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 article applies this algorithm to a new problem (identification of person) and does not change the way the algorithm is defined.

4.1. Multi-category SVM and the KKT conditions

Let's consider a training 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, $K \geq 2$. The SVM classifier used for data classification is defined by:

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

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 class i and j is $2/||w_i - w_j||$. In order to get the largest margin between class i and j, minimization of the sum of $||w_i - w_j||^2$ for all $i, j = 1, \ldots, K$ 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 hyper-plan ij of the lth training vector. So, the proposed quadratic function is as the follows:

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

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:

$$L = \frac{1}{2}\sum_{i=1}^{K}\sum_{j=i+1}^{K}||w_i - w_j||^2 + \frac{1}{2}\sum_{i=1}^{K}||w_i||^2$$
$$+ C\sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l\in C_{ij}}\xi_l^{ij} - \sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l\in C_{ij}}\mu_l^{ij}\xi_l^{ij}$$
$$- \sum_{i=1}^{K}\sum_{j=i+1}^{K}\sum_{x_l\in C_{ij}}\alpha_l^{ij}\left(y_l^{ij}[(w_i - w_j)^T\Phi(x_l)\right.$$
$$\left. + (b_i - b_j)] - 1 + \xi_l^{ij}\right) \quad (4)$$

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

The optimization problem is now written as:

$$\min_{w_i,b_i,\xi_l^{ij}}\max_{\alpha_l^{ij},\mu_l^{ij}} L \quad (5)$$
$$s.t. \quad \alpha_l^{ij},\mu_l^{ij} \geq 0, i \neq j$$

In order to obtain the expression of w_i, $i = 1,\dots,K$, we compute the following gradient: $\frac{\partial L}{\partial w_i} = 0$, $\frac{\partial L}{\partial b_i} = 0$ and $\frac{\partial L}{\partial \xi_l^{ij}} = 0$. In consequence, we get:

$$w_i = \frac{1}{K+1}\sum_{\substack{j=1\\j\neq i}}^{K}\left(\sum_{x_l\in C_i}\alpha_l^{ij}\Phi(x_l) - \sum_{x_l\in C_j}\alpha_l^{ij}\Phi(x_l)\right) \quad (6)$$

$$\sum_{\substack{j=1\\j\neq i}}^{K}\left(\sum_{x_l\in C_i}\alpha_l^{ij} - \sum_{x_l\in C_j}\alpha_l^{ij}\right) = 0 \quad (7)$$

$$\alpha_l^{ij} + \mu_l^{ij} = C \quad (8)$$

Then by replacing w_i by its expression (Equation (6)), 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 an 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 $i = 1,\dots,K$, $j = i+1,\dots,K$:

$$g_m^{ij} = \frac{\partial W}{\partial \alpha_m^{ij}} \begin{cases} > 0; & if \quad \alpha_m^{ij} = 0; & D(dv_m^{ij}) \\ = 0; & if \quad 0 < \alpha_m^{ij} < C; & S(sv_m^{ij}) \\ < 0; & if \quad \alpha_m^{ij} = C; & E(ev_m^{ij}) \end{cases} \quad (9)$$
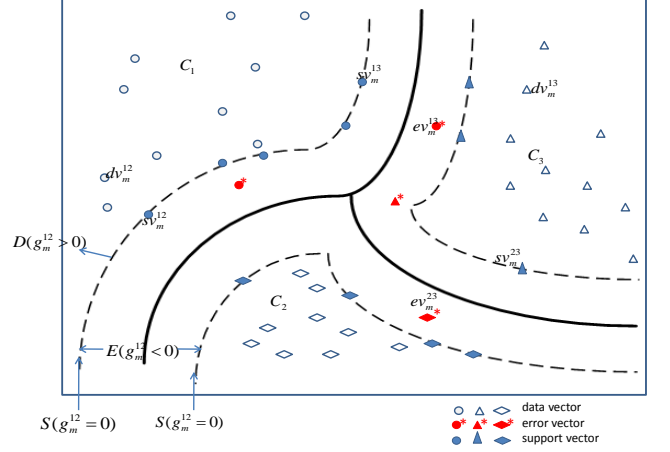

Figure 4: Three sets obtained from training samples, considering three classes

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 the coefficient $\alpha_c^{pq} = 0$, $p = 1,\dots,K, q = p+1,\dots,K$. The coefficients of the existing samples in the training dataset (only the support vectors) change value during each incremental step to meet the KKT conditions. In particular, the adaptation of g_m^{ij} when new data x_c is added can be expressed differentially as:

$$\Delta g_m^{ij} = y_m^{ij}\left(\beta^{ij,pq}\Delta\alpha_c^{pq}K_{cm} + \sum_{x_l\in C_i}\left(2\Delta\alpha_l^{ij}\right.\right.$$
$$+ \sum_{\substack{n=1\\n\neq i,j}}^{K}\Delta\alpha_l^{in}\right)K_{lm} - \sum_{x_l\in C_j}\left(2\Delta\alpha_l^{ij} + \sum_{\substack{n=1\\n\neq i,j}}^{K}\Delta\alpha_l^{nj}\right)K_{lm}$$
$$\left.- \sum_{\substack{n=1\\n\neq i,j}}^{K}\sum_{x_l\in C_n}\left(\Delta\alpha_l^{in} - \Delta\alpha_l^{nj}\right)K_{lm} + \left(\Delta b_i - \Delta b_j\right)\right) \quad (10)$$

$$\gamma^{j,pq}\Delta\alpha_c^{pq} + \sum_{\substack{n=1\\n\neq i}}^{K}\left(\sum_{x_l\in C_i}\Delta\alpha_l^{in} - \sum_{x_l\in C_j}\Delta\alpha_l^{in}\right) = 0 \quad (11)$$

where $i = 1,\dots,K, j = i+1,\dots,K$, 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 $\beta^{ij,pq}, \gamma^{j,pq}$ are defined in [19].

From Equation (9), 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 Equations (10) and (11) can be written as the following matrix equation: (as described in detail in [19])

$$\begin{bmatrix} \Delta b \\ \Delta\alpha \end{bmatrix} = -RH^{pq}\Delta\alpha_c^{pq} \quad (12)$$

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

6

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

4.3. Adding a new vector

When a new sample x_c is added, we set the training set as $T^* = S_p \cup x_c$, where S_p is the support vector set of the previous step. 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. 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 Equation (12) 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 bookkeeping 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 neighbour 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.

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 are used for testing (and updating). 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.
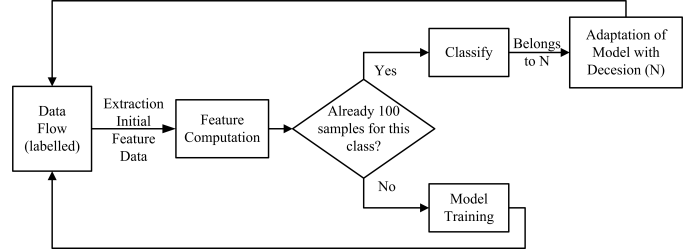

Figure 5: Incremental learning work flow

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 least, 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

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

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.

	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

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

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 that the one used for incremental learning experiment has been used to train the Support Vector Machines. 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 for 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 vector) 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.

References

[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] 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, Journal of the American Statistical Association 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, Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on 37 (5) (2007) 1119–1137.

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

[7] K. Yoon, D. Harwood, L. Davis, Appearance-based person recognition using color/path-length profile, Journal of Visual Communication and 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 Transactions on Information and Systems 93 (7) (2010) 1761–1772.

[10] E. Bredensteiner, K. Bennett, Multicategory classification by support vector machines, Computational Optimization and Applications 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: Proc. IEEE Int Data Mining ICDM 2001 Conf, 2001, pp. 641–642.

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

[17] G. Cauwenberghs, T. Poggio, Incremental and decremental support vector machine learning, Advances in Neural Information Processing Systems 13 (2001) 409–415.

[18] C. Diehl, G. Cauwenberghs, Svm incremental learning, adaptation and optimization, in: Neural Networks, 2003. Proceedings of the International Joint Conference on, 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 Inernational 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, ECCV'98 Fifth European Conference on Computer Vision (1998) 475–490.

[23] R. Haralick, K. Shanmugam, I. Dinstein, Textural features for image classification, Systems, Man and Cybernetics, IEEE Transactions on 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, Philosophical Magazine 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, Second Edition, 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 Explorations 11 (1) (2009) 10–18.