SCALABLE PEOPLE RE-IDENTIFICATION BASED ON A ONE-AGAINST-SOME CLASSIFICATION SCHEME

William Robson Schwartz

Department of Computer Science, Federal University of Minas Gerais, Belo Horizonte-MG, Brazil

ABSTRACT

People re-identification is a problem of increasing interest in computer vision, mainly in applications such as video surveillance and dynamic environment monitoring. However, the large amount of data captured from multiple cameras, the large number of agents involved and poor acquisition conditions make it a difficult problem to solve. Recent works have shown that the use of multiple feature extraction methods combined by a weighting technique considering a one-against-all classification scheme provide accurate results for applications such as face recognition and appearance-based modeling. However, to enroll new subjects, all models need to be rebuild, which results in an increasingly computational time. To reduce this problem, this work proposes a classification scheme, called one-against-some, to allow scalable enrollment of new individuals without reducing the accuracy when compared to the one-against-all classification scheme.

***Index Terms*—** People re-identification, robust feature descriptors, one-against-all classification scheme, partial least squares

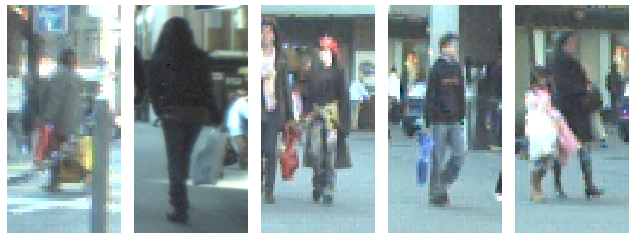
1. INTRODUCTION

In scenarios where the environment is observed by multiple cameras, not necessarily with intersection in the field of view, people re-identification aims at the identification of individuals that have been previously detected and uniquely identified. People re-identification has been considered on several application domains, such as video surveillance and monitoring [1] and sport events [2].

Recent people re-identification methods have considered approach based on feature descriptors to model the appearance [3], creation of panoramic appearance maps [4], detection of interest points to identify previously known subjects considering a *KD-tree* [5] and auto-similarity images [6]. However, such methods still suffer with the change of appearance of individuals captured by different cameras and low quality of the samples due to variations in illumination and shadows, as illustrated in Figure 1.

Due to the low quality of the samples, facial information, which presents the most discriminative characteristics for humans, is hard to be captured. Therefore, several approaches have been applied to model the appearance based mainly on the clothing using color-based feature descriptors such as color histograms [7], non-parametric kernel density estimation [8] or spatial information by representing appearances in joint color spatial spaces [9]. However, to capture subtle difference between appearance, it is important to extract features in a fine way so that discriminative clues to distinguish individuals can be detected.

Methods proposed by Schwartz et al. [10, 11, 12], have addressed the problem of extracting descriptors from several feature channels in a very fine way so that subtle differences between low quality appearances can be captured. After extracting the features, a one-against-all classification scheme based on Partial Least Squares (PLS) models is applied to find the regions of an individual appearance are more important to distinguish that particular individual from the remaining ones.



(a) (b) (c) (d) (e)

Fig. 1: Samples of the ETHZ people re-identification dataset.

The combination of using large feature sets to capture subtle differences and the use of one-against-all classification scheme leads to issues regarding the scalability of the identification system. As pointed out in [11], there are two main scalability issues. First, the number of subjects considered can be large, so that common search techniques, such as brute force nearest neighbor, to match samples do not scale well, this issue was addressed in [11], in the domain of face identification. Second, in applications such as people re-identification, in which new subjects are added incrementally, the need of rebuilding models every time a new subject is first seen compromises the computational performance of the system.

To address the scalability issue regarding the need for rebuilding PLS models every time a new subject is considered, this work proposes the use of a one-against-some classification scheme, an incremental approach in which only a subset of subjects need to be considered when a new subject is added and previously constructed models are maintained.

Experimental results show that the use of a one-against-some approach has achieved similar results as the one-against-all classification scheme. Furthermore, the addition of new subjects without the need for rebuilding PLS models has shown promising results.

The text is organized as follows. The one-against-some classification scheme is described in Section 2 together with some concepts used in this work. Experimental results obtained with the proposed method presented in Section 3. Finally, Section 4 concludes the paper with final remarks.

2. METHODOLOGY

This section describes the proposed approach and reviews some important concepts necessary to its development. First we describe

Thanks to FAPESP, FAPEMIG, CNPq and CAPES for the financial support.

the feature extraction process responsible to obtain a description of the samples. Then we review the regression and dimension reduction method called Partial Least Squares, which is used to create discriminative models to distinguish different subject's appearance. Finally, the proposed one-against-some classification scheme will be described.

2.1. Feature Extraction

Due to the dataset used in this work, only one exemplar is provided for each appearance i in the form of an image window. To extract fine details from the sole subject sample, the window around the person is decomposed into overlapping blocks and a set of features is extracted for each block to construct a feature vector. Therefore, for a given subject i, we obtain one sample described by a high dimensional feature vector $\boldsymbol{v}_i$.

In this work we considered texture information using the gray level co-occurrence matrices (GLCM) [13], a method widely used for texture analysis. GLCM represents second order texture information by capturing the joint probability distribution of gray-level pairs of neighboring pixels. We use a set with 12 descriptors: angular second-moment, contrast, correlation, variance, inverse difference moment, sum average, sum variance, sum entropy, entropy, difference variance, difference entropy, and directionality.

The blocks used have sizes of 16×16 and 32×32 with strides of 8 and 16 pixels, respectively, resulting in 70 blocks per window for each color band. We work in the HSV color space. For each color band, we create four GLCM, one for each of the ($0°$, $45°$, $90°$, and $135°$) directions. The displacement considered is 1 pixel and each color band is quantized into 16 bins.

Once the feature extraction process is performed for all blocks inside an image window, feature descriptors are concatenated creating a high dimensional feature vector $\boldsymbol{v}_i$, with $10,080$ feature descriptors.

2.2. Partial Least Squares

Partial Least Squares is a statistical method employed to model relations between sets of observed variables by the estimation of a low dimensional latent space [14, 15]. It aims at estimating a low dimensional space to exploit the separation between samples from different classes, causing samples with similar characteristics to be clustered in the latent space. Its uses, formulation and how to apply PLS for data classification are discussed as follows.

PLS was created by Herman Wold in the 1970s and has been exploited in several areas, such as Bioinformatics, Chemometrics and Neurosciences. More recently, PLS has been successfully employed to Computer Vision problems to perform data classification, dimension reduction and regression [11, 16, 10, 17].

Specifically, PLS estimates latent variables as linear combinations of the original variables in a matrix $\boldsymbol{X}$, composed of variables used to describe samples, and a matrix $\boldsymbol{Y}$ containing a set of response variables (when a single response variable is considered, a vector $\boldsymbol{y}$ is used). It presents the advantage of being allowed to work in problems containing high dimensional data and very few samples. The following paragraphs describe the PLS decomposition and latent space estimation.

Let n be the number of samples described by d variables each, stored in a mean-centered matrix $\boldsymbol{X}_{n \times d}$, and let k be the number of response variables associated, stored in a mean-centered matrix $\boldsymbol{Y}_{n \times k}$. PLS estimates a p-dimensional space ($p \ll d$) by decomposing $\boldsymbol{X}$ and $\boldsymbol{Y}$ into

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^T + \boldsymbol{E}$$
$$\boldsymbol{Y} = \boldsymbol{U}\boldsymbol{Q}^T + \boldsymbol{F}$$

where $\boldsymbol{T}_{n \times p}$ and $\boldsymbol{U}_{n \times p}$ are matrices containing the latent variables, matrices $\boldsymbol{P}_{d \times p}$ and $\boldsymbol{Q}_{k \times p}$ represent the loadings, and matrices $\boldsymbol{E}_{n \times d}$ and $\boldsymbol{F}_{n \times k}$ are the residuals. The nonlinear iterative partial least squares (NIPALS) algorithm [14] is employed to perform such decomposition by estimating a set of projection vectors $\boldsymbol{w}_i$, ($i = 1, 2, \ldots, p$) which will be stored in a matrix $\boldsymbol{W} = (\boldsymbol{w}_1, \boldsymbol{w}_2, \ldots, \boldsymbol{w}_p)$, such that

$$[cov(\boldsymbol{t}_i, \boldsymbol{u}_i)]^2 = \max_{|\boldsymbol{w}_i| = |\boldsymbol{u}_i| = 1} [cov(\boldsymbol{X}\boldsymbol{w}_i, \boldsymbol{Y}\boldsymbol{u}_i)]^2 \tag{1}$$

where $|\boldsymbol{w}_i|$ and $|\boldsymbol{u}_i|$ denote the 2-norm of vectors $\boldsymbol{w}_i$ and $\boldsymbol{u}_i$, respectively. $\boldsymbol{t}_i$ represents the i-th columns of matrices $\boldsymbol{T}$, and $cov(\boldsymbol{t}_i, \boldsymbol{u}_i)$ is the sample covariance between latent vectors $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$. This algorithm extracts the latent variables $\boldsymbol{t}_i$ and $\boldsymbol{u}_i$ iteratively. Matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ are deflated by subtracting their rank-one approximations as

$$\boldsymbol{X}_{i+1} = \boldsymbol{X}_i - \boldsymbol{t}_i\boldsymbol{p}_i^T$$
$$\boldsymbol{Y}_{i+1} = \boldsymbol{Y}_i - t_i\boldsymbol{q}_i^T$$

where $\boldsymbol{X}_i$ and $\boldsymbol{Y}_i$ represent matrices $\boldsymbol{X}$ and $\boldsymbol{Y}$ at the i-th iteration ($\boldsymbol{X}_1 = \boldsymbol{X}$ and $\boldsymbol{Y}_1 = \boldsymbol{Y}$), and $\boldsymbol{p}_i$ and $\boldsymbol{q}_i$ denote the i-th columns of the matrices $\boldsymbol{P}$ and $\boldsymbol{Q}$, respectively.

After the extraction of the projection vectors, the p-dimensional representation of $\boldsymbol{X}_{n \times d}$ is given by $\boldsymbol{T}_{n \times p}$, used to extract the regression coefficients $\boldsymbol{\beta}_{d \times k}$ by

$$\boldsymbol{\beta} = \boldsymbol{W}(\boldsymbol{P}^T\boldsymbol{W})^{-1}\boldsymbol{T}^T\boldsymbol{Y}. \tag{2}$$

Then, the regression responses, Y_v, for a feature vector $\boldsymbol{v}_{d \times 1}$ are computed by

$$Y_v = \overline{Y} + \boldsymbol{\beta}^T\boldsymbol{v}S \tag{3}$$

where $\overline{Y}_{1 \times k}$ is the sample mean of each variable of $\boldsymbol{Y}$ and $S_{1 \times k}$ is the standard deviation of the variables in $\boldsymbol{Y}$.

2.3. One-Against-Some Approach

Aiming at maximizing the discrimination between C different classes, the one-against-all classification scheme has been applied to estimates C PLS models considering single response variables [10, 11]. This way, the response variable $\boldsymbol{Y}$, represented by a matrix in Section 2.2, becomes a vector, $\boldsymbol{y}$, and its entries have class indicators. When a test sample is presented, it is projected to each model, resulting in a set with C responses (one per class) and the best matching class is associated to the model presenting the highest response.

The proposed method follows the same steps of our previous work for appearance-based modeling [10], as illustrated in Figure 2(a). The main difference is that for the one-against-some, we consider only a subset of individuals as negative instead of all remaining subjects as in the one-against-all classification scheme. The diagram in Figure 2(b) shows the steps performed on the proposed method, discussed as follows.

It can be seen in the diagram (b) in Figure 2 that, on the one hand, the one-against-some approach considers a subset of subjects as the negative class instead of all remaining subjects. This way not all subjects need to be known by the time of the enrollment of a

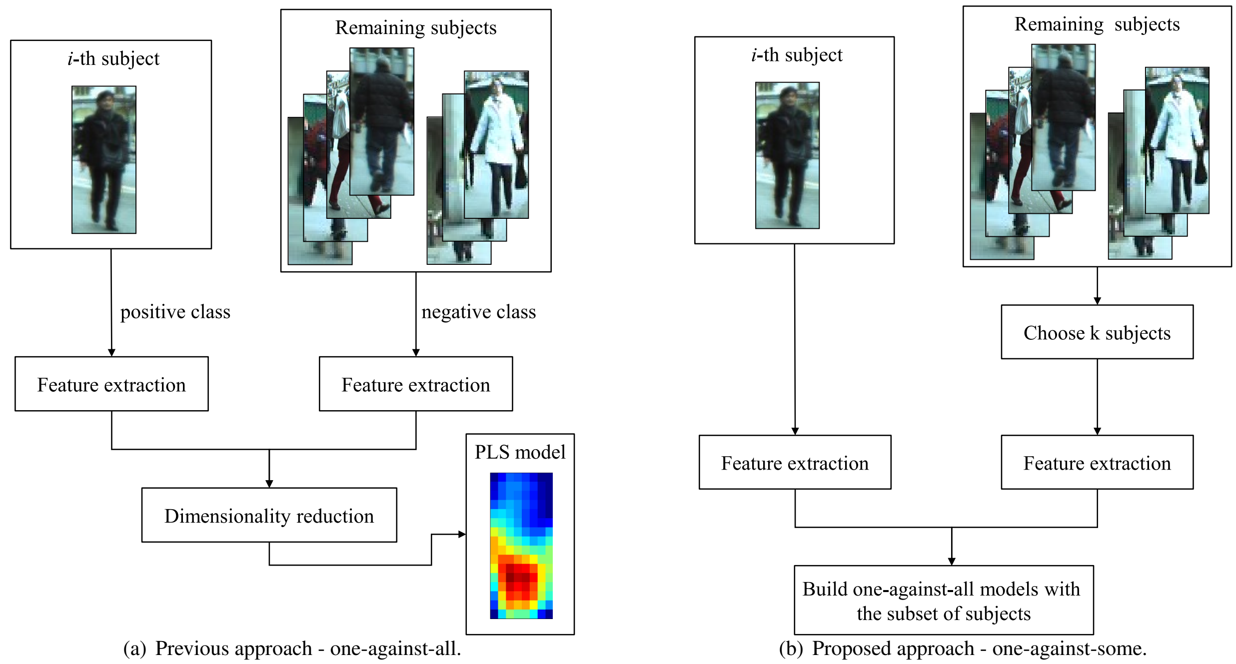


Fig. 2: Diagrams of the one-against-all and one-against-some classification schemes. (a) one-against-all approach for each subject; (b) one-against-some approach for each subject – once a subset of the remaining subjects is chosen, the one-against-all approach is applied.

new individual. However, on the other hand, more than one model is built considering a given subject as positive sample (this number is estimated experimentally). Therefore, there will be more than one regression response for each subject. As in the one-against-all approach, the subject associated with the highest response is considered the best match for the sample being tested.

The main advantage of the one-against-some classification scheme is that when a new subject j is presented, the existing models do not need to be rebuild to consider this new individual, few PLS models considering j as the positive class are built choosing random subsets of already enrolled subject to represent the negative class. The next section shows the results achieved with this simple change on the one-against-all classification scheme.

3. EXPERIMENTAL RESULTS

In this section, we present the experiments we performed to validate the proposed method for people re-identification based on a one-against-some classification scheme. For the experiments, we used the ETHZ Dataset for Appearance-Based Modeling [10, 18][1]. All experiments were conducted on an Intel Xeon 5160, 3 GHz dual core processor with 8GB of RAM running Windows operating system.

In the first experiment, we compare the one-against-all to the one-against-some classification schemes. The construction of the PLS models is performed considering exclusively samples from the training, which are disjunct from the samples in the test. Experimental results showed that models considering subsets of the remaining subject with size 7 achieved the best results. Therefore, since the individuals on the subsets are chosen randomly, it is expected that each subject appear in seven PLS models.

Table 1 shows the rank-1 recognition rate for the three video sequences of the ETHZ dataset. We can see that the results for both approaches are very similar, which validates experimentally the use of the one-against-some classification scheme.

[1] http://www.dcc.ufmg.br/∼william/datasets.html

Table 1: Rank-1 recognition rates achieved for the ETHZ dataset.

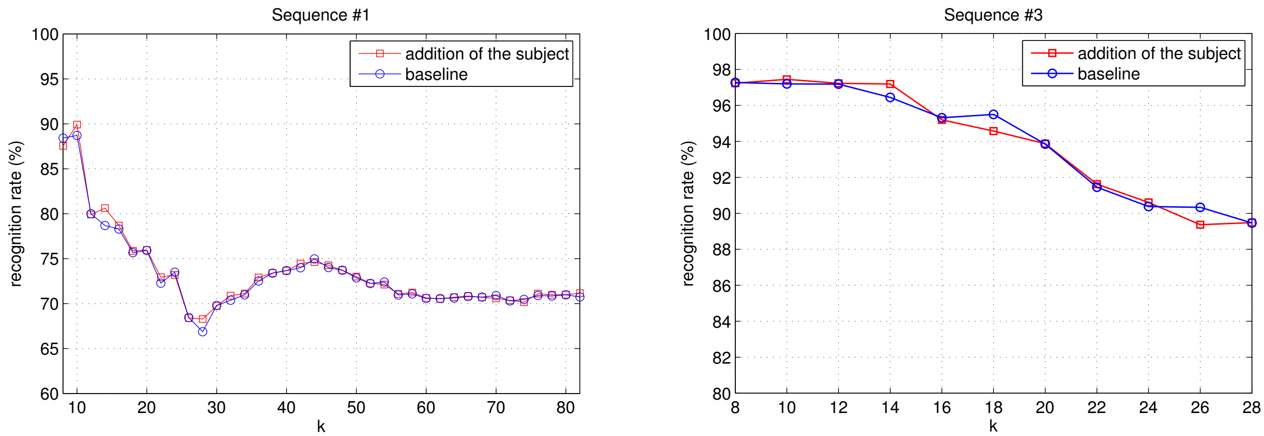
Sequence	One-against-all (%)	One-against-some (%)
#1	71.90	70.74
#2	73.50	72.85
#3	90.50	89.47

In the second experiment, we evaluate the recognition rates when new subjects are added incrementally. For a subset of k subjects, we build one-against-some models considering a set of k subjects to have a baseline recognition rate; then, we remove one subject and build models considering only $k-1$ subject; finally, we build models adding the removed subject.

Figure 3 shows that the addition of new subjects without rebuilding the previously constructed models achieved similar results to those obtained when all subjects are known before-hand. Such results make scalable the incremental subject enrollment and the development of online people re-identification systems (when not all subjects are known previously), since no PLS model needs to be reconstructed.

4. CONCLUSIONS AND FUTURE WORK

A problem of great interest in computer vision, people re-identification presents several challenges, such as the large amount of data and subjects, that need to be tackled. This work proposed a one-against-some classification scheme, an incremental approach in which only a subset of subjects need to be considered when a new subject is added and previously constructed models are maintained. Therefore, the scalable re-identification systems can be designed.



(a) Sequence #1

(b) Sequence #3

Fig. 3: Recognition rates as function of the number of subjects considered.

Experimental results shown that the proposed approach is promising and it might be used to build scalable people re-identification systems that focus on the discriminability of the appearances based on feature descriptors. As future work, more investigation need to be devoted to the one-against-some classification scheme and on the reduction of projections necessary to evaluate a sample on all models available.

5. REFERENCES

[1] Peter Henry Tu, Gianfranco Doretto, Nils O. Krahnstoever, a.G. Amitha Perera, Frederick W. Wheeler, Xiaoming Liu, Jens Rittscher, Thomas B. Sebastian, Ting Yu, and Kevin G. Harding, "An Intelligent Video Framework for Homeland Protection," in *Defence and Security Symposium*, 2007.

[2] Horesh Ben Shitrit, Jerome Berclaz, Francois Fleuret, and Pascal Fua, "Tracking Multiple People Under Global Appearance Constraints," in *IEEE Intl. Conference on Computer Vision*, 2011.

[3] Niloofar Gheissari, Thomas B. Sebastian, Peter Henry Tu, Jens Rittscher, and Richard Hartley, "Person Reidentification Using Spatiotemporal Appearance," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2006.

[4] Tarak Gandhi and Mohan Manubhai Trivedi, "Person Tracking and Reidentification : Introducing Panoramic Appearance Map (PAM) for Feature Representation," *Machine Vision and Applications*, vol. 18, no. 3, pp. 207–220, 2007.

[5] Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu, and Bruno Steux, "Interest Points Harvesting in Video Sequences for Efficient Person Identification," in *European Conference on Computer Vision*, 2008.

[6] Yinghao Cai and Matti Pietikainen, "Person Re-identification Based on Global Color Context," in *Intl. Workshop on Visual Surveillance*, 2010.

[7] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based Object Tracking," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[8] Zhe Lin and Larry S. Davis, "Learning Pairwise Dissimilarity Profiles for Appearance Recognition in Visual Surveillance," in *International Symposium on Advances in Visual Computing*, 2008, pp. 23–34.

[9] A. Elgammal, R. Duraiswami, and L.S. Davis, "Probabilistic Tracking in Joint Feature-Spatial Spaces," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2003, pp. 781–788.

[10] W. R. Schwartz and L. S. Davis, "Learning Discriminative Appearance-Based Models Using Partial Least Squares," in *Brazilian Symposium on Computer Graphics and Image Processing*, 2009.

[11] W. R. Schwartz, H. Guo, and L. S. Davis, "A Robust and Scalable Approach to Face Identification," in *European Conference on Computer Vision*, 2010, vol. 6316 of *Lecture Notes in Computer Science*, pp. 476–489.

[12] W. R. Schwartz, H. Guo, J. Choi, and L. S. Davis, "Face Identification Using Large Feature Sets," *IEEE Transactions on Image Processing*, vol. 21, no. 4, pp. 2245–2255, 2012.

[13] Robert M. Haralick, K. Shanmugam, and Dinstein, "Textural Features for Image Classification," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 3, no. 6, pp. 610–621, 1973.

[14] H. Wold, "Partial Least Squares," in *Encyclopedia of Statistical Sciences*, vol. 6, pp. 581–591. Wiley, New York, NY, USA, 1985.

[15] R. Rosipal and N. Kramer, "Overview and Recent Advances in Partial Least Squares," *Lecture Notes in Computer Science*, vol. 3940, pp. 34–51, 2006.

[16] W.R. Schwartz, A. Kembhavi, D. Harwood, and L.S. Davis, "Human Detection Using Partial Least Squares Analysis," in *IEEE Intl. Conference on Computer Vision*, 2009, pp. 24–31.

[17] A. Sharma and D. Jacobs, "Bypassing Synthesis: PLS for Face Recognition with Pose, Low-Resolution and Sketch," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2011.

[18] A. Ess, B. Leibe, K. Schindler, , and L. van Gool, "A mobile vision system for robust multi-person tracking," in *IEEE Intl. Conference on Computer Vision and Pattern Recognition*, 2008.