WYA^2: Optimal Individual Features Extraction for VideoSurveillance Re-identification

Isaac Martín de Diego, Ignacio San Román, Cristina Conde and Enrique Cabello
FRAV, Face Recognition and Artificial Vision Group, Rey Juan Carlos University, C/Tulipan s/n, Móstoles, Spain

Keywords: VideoSurveillance, Re-identification, Kernels Combination, Bag of Features.

Abstract: A novel method for re-identification based on optimal features extraction in VideoSurveillance environments is presented in this paper. A high number of features are extracted for each detected person in a dataset obtained from a camera in a scenario. An evaluation of the relative discriminate power of each bag of features for each person is performed. We propose a forward method in a Support Vector framework to obtained the optimal individual bags of features. These bags of features are used in a new scenario in order to detect suspicious persons using the images from a non-overlapping camera. The results obtained demonstrate the promising potential of the presented approach. The proposed method can be enriched with future enhancements that can further improve its effectiveness and complexity in more complex VideoSurveillance situations.

1 INTRODUCTION

Recognizing people moving through different non-overlapping cameras, is a challenging problem usually called re-identification. For example, in an airport, if a person is tagged as a suspect, we want to learn his/her appearace for going after him/her through all cameras he/she passes. This two step re-identification problem is called ”tag-and-track”.
Unfortunately, current tag-and-track algorithms are likely to fail in real-world scenarios for several reasons. On the one hand, different lighting conditions, points of view between cameras, zoom, camera quality, etc, can induce serious errors in the matching of target candidates (An et al., 2013). On the other hand, most re-identification algorithms approach use the same descriptors for all people, regardless of intrinsic features of each person (see, for instance (Moon and Pan, 2010),(Martinson et al., 2013),(Li et al., 2014),(Tome-Gonzalez et al., 2014)).
In (Moctezuma et al., 2015) a method for human identification in mono and multi-camera surveillance environments is presented. Several approaches have been proposed to measure the relevance of each extracted feature in order to use a weighted combination of them. However, the same relevant features and weights are extracted for all the individuals based on gobal information from the complete dataset.
A VideoSurveillance system is a combination of hardware and software components that are used to capture and analyze video. The primary aim of these kind of systems is to monitor the behaviour of objects (usually people) in order to check for suspicious or abnormal behaviours, using the extracted information of those objects (physical features, trajectories, speed...) from a variety of sensors (usually cameras). INVISUM (Intelligent VideoSurveillance System) is a project funded by the Spanish Ministry of Economy and Competitivity focused on the development of an advanced and complete security system (INV, 2014). The goal of the project is the development of an intelligent VideoSurveillance system that addresses the limitations of scalability and flexibility of current VideoSurveillance systems incorporating new compression techniques, pattern detection, decision support, and advanced architectures to maximize the efficiency of the system. As part of the INVISUM system a suspect detector process needs to be developed.
In this paper, we propose a novel features selection process to re-identify target humans in non-overlapping cameras. We call this system WYA^2. WYA^2 stands for “Why You Are Who You Are”. The main idea behind our method is to select, for each target person in the dataset his/her personal most discriminative bag of features. The approach is designed to be directly applicable to typical real-world surveillance camera networks. In a nutshell the method performs as follows. First we extract a high number of features from every person detected in a camera.

Diego, I., Román, I., Conde, C. and Cabello, E.
WYA^2: Optimal Individual Features Extraction for VideoSurveillance Re-identification.
DOI: 10.5220/0006435400350041
In *Proceedings of the 14th International Joint Conference on e-Business and Telecommunications (ICETE 2017) - Volume 5: SIGMAP*, pages 35-41
ISBN: 978-989-758-260-8

These features are related to color and texture. Then, a selection of the most discriminative features is performed for each target person in a Support Vector Machine framework. Then, in other non-overlapping camera, only these discriminate features are used to detect the target person. If the set of individuals in both cameras presents similar global features, the most discriminative features obtained in camera 1 should be able to detect the target suspicious person in the second camera. The method has been tested with several cases of study extracted from the INVISUM dataset acquired at the Universidad Rey Juan Carlos and first described in (Roman et al., 2017).

The rest of the paper is organized as follows. Section 2 presents the proposed methodology. The database collected to test our method is presented in Section 3. The experiments to evaluate the performance of the method are in Section 4. Conclusions and a brief discussion about the method and results are presented in Section 5.

2 WYA^2 METHOD

A key issue for human re-identification in real-world camera network is that each person should be described using different discriminative features. In this paper we propose a method for choosing the most discriminative characteristics for each person. A general sequence diagram of the proposed method is presented in Figure 1.

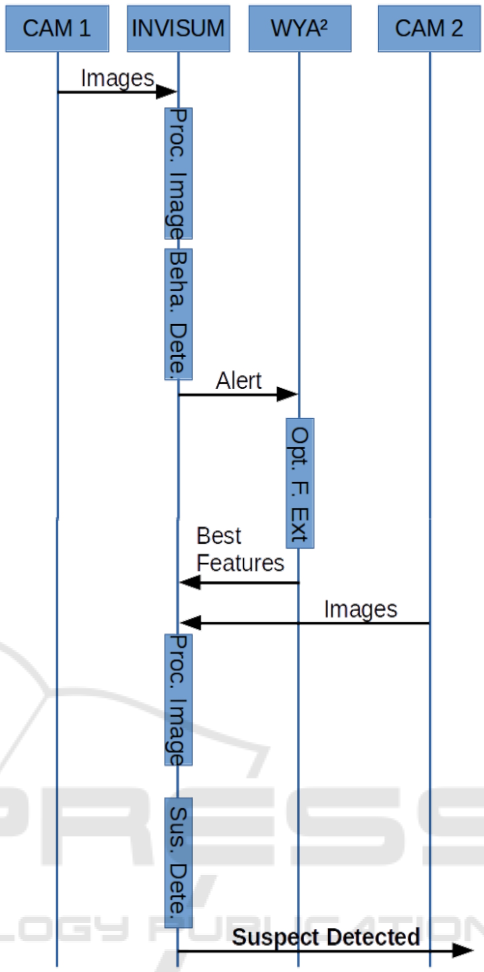

Figure 1: Sequence diagram of the WYA^2 Method.

For simplicity, consider a VideoSurveillance system with two non-overlapping cameras in two different scenarios. The method starts detecting persons in the first scenario. Features of every person are extracted and stored in a data base from each frame in which the person is detected. The INVISUM system has a Behaviour Detector. A person could be tagged as a suspect, based on the person anomalous behaviour (anomalous trajectory, presence in a restriced area,...) or based on security staff information. When a person is detected as a suspect, an alert is generated and the information is sent to the WYA^2 process. A set of Support Vector Machines (SVMs) are trained to detect his/hers discriminative features. To train the SVM, each frame of the suspicious person is labeled as $+1$ and the rest of the people detected is labeled as -1. To extract the most significative bag of features, a forward process similar to the well-known forward variable selection used in stepwise regression, is proposed. This approach involves starting with no features in the model, testing the addition of each bag of features using a selection criterion based on the error of the model, and repeating this process until no improves are achieved. In the first step, a SVM is trained for each bag of features. The global error (our selection criterion) of each SVM is calculated as the sum of the false acceptance rate (FAR) and false rejection rate (FRR). The best bag of features, those corresponding to the lowest error, are selected. The optimal error is set as the best bag of features error. Next, the second best bag of features are included in the model given the presence of the best bag of features in the model. That is, a new SVM using both bag of features is calculated. Each bag of features generates a kernel, and these kernels are combined using a combination of kernels similar to the presented in (de Diego et al., 2010). The fused kernel is used to train a new SVM. If the new error is lower than the optimal error, then the new best model is the model with the two best bag of features, and the new optimal error is updated. This forward process continues until the optimal error does not decrease. Thus, for each suspected person a set of optimal individual discriminative features is selected from camera 1. That is, a set of features that best discriminate the suspected person related to the other people in the database. Notice that most of the methods presented in the literature (see for instace (Moctezuma et al., 2015)) calculate the best discriminate features over the complete datasets. However, WYA^2 calculates the best discriminative features over each suspicious person. Thus, it is expected that different sets of bags of features will be extracted for different people, regarding their discriminative power.

The aim of any re-identification system is to detect the suspicious person in the second camera. For each detected person in camera 2, a SVM is trained using the optimal individual features for the suspicious person obtained in camera 1. This is, the same kind of features are used, but with new kernels calculated from camera 2 data. Feature space in camera 2 could be very different than camera 1 due to different lighting conditions, points of view between cameras, zoom, etc. However, same kind of features seems to be discriminative. In order to train the SVM, each frame of the detected person is labeled as $+1$ and the rest of the frames are labeled as -1. In this case, the relevant error measure is the false negative rate (FNR): the proportion of positives that are nor correctly identified as such. It is expected that this error will be low for the suspected person (if he/his is present in scenario 2), and it will be high for other detected people. If this error is lower than a threshold, then an alert is generated: the target person in scenario 1 has been detected as suspicious person in scenario 2. Notice that when the WYA^2 method returns as optimal features a number of characteristics with low discriminative power in camera 1, it is expected that the FNR error in scenario 2 will be high. That is, if the target person in scenario 1 has no significative different features regarding the other persons in such scenario, it is unlikely that those features will be able to detect as suspicious the target person in scenario 2.

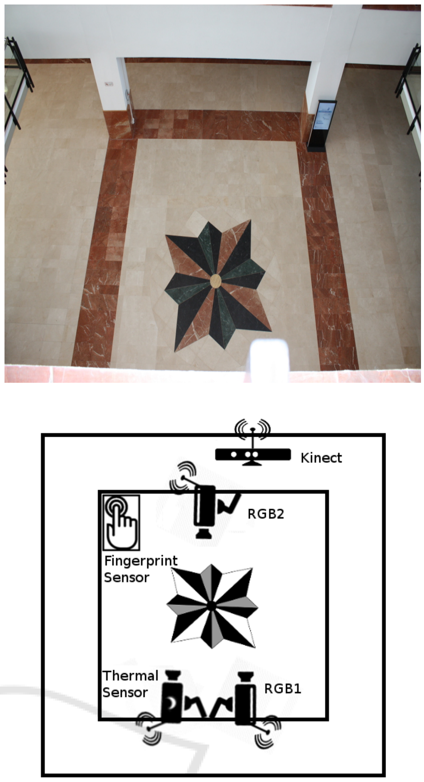

Figure 2: Indoor Scenario and sensors positions.

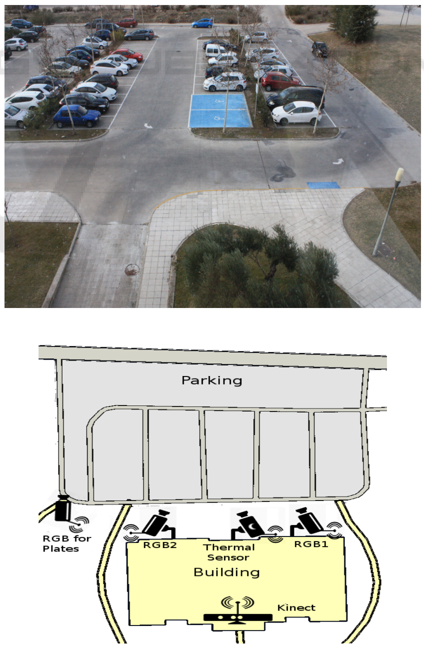

Figure 3: Outdoor Scenario and sensors positions.

3 DATA BASE AND FEATURES

In order to evaluate the performance of the WYA^2 method, we have performed real experiments using two academic scenarios at the campus of the University Rey Juan Carlos in Mostoles, Spain.

The first scenario is an indoor scene shown in Figure 2. The image covers most of the main hall of a classroom building. The hall is mainly used to move from one classroom to other and to enter the building. The second area of interest is an outdoor scene shown in Figure 3. The image covers a parking area close to the classroom building considered in the previous scene.

In Figures 2 and 3 all the INVISUM sensors are presented. However, for the purposes of the present work, images from RGB1 camera in scenario 1 and images from RGB1 camera in scenario 2 are used.

Figure 4 shows sample images from these cameras.
The dataset used for our experiments has got more than 5000 images, from which 40 people were extracted (18 in the first scenario, and 22 in the second). For more details on the use of this database, please contact the authors. For testing purposes, a suspicious target person (see Figure 5) was considered and used in both scenarios. Thus, when the WYA^2 method is tested using the suspicious target as target person, it is expected that proper bag of features is going to be selected in scenario 1. Using such a bag of features, it is expected that this person is going to be detected as suspicious in scenario 2. No other people were presented in both scenarios. Thus, when the WYA^2 is tested using as target person a non-suspicious person in scenario 1, it is expected that no suspicious person is going to be detected in scenario 2.

3.1 Features Extraction

The first step for Features Extraction is to detect and extract the humans presented in each image using the background substraction method based on Gaussian Mixture of Models (GMM) for object detection. Then, a set of well-known soft-biometric features is extracted from each person. In most of the literature works, few features are considered. However, we consider a complete set of different feature categories. These categories are the RGB color space, the HSV color space, the co-ocurrence matrix (Haralick, 1979), and the Local Binary Pattern (Ojala et al., 1994). Thus, we have taken into account a total of 222 bags of features. A complete list of the Bag of Features considered is presented in Table 1. For instance, the first six bags of features correspond to the Mean and Typical Deviation of the channels R,G,B, H, S and V using the image as source of information. The next bags of features are related the Mean, Typical Deviation, Skewness and Kurtosis of each channel of the Histogram. The next bags of features are related the Dispersion, Energy and Entropy of each channel of the Histogram, an so on.

4 EXPERIMENTS

In order to test the performance of our approach, the WYA^2 method was tested on the 18 detected persons in scenario 1. Thus, the most discriminate features for each target person were extracted, and used to re-identification in scenario 2. The main results are presented in Table 2. For each target person the ID of the bags of features selected as optimal for re-identification tasks using the WYA^2 method are shown. In addition, the suspicious person detected (if any) in scenario 2 when the optimal set of features was used is presented.
It is clear that the set of bag of features selected depends on each person. In fact, different number of bags of features were selected in any case. For instance, for person number 1 (the suspicious target presented in both scenarios), the optimal bag of features are: Sum Mean, Sum Std, Sum Entropy obtained from channel R, source SGLD Matrix 90°, and [Mean, Typical Deviation], channel S, source Image. In this case, the suspicious target (ID 1) is correctly detected as suspicious person in scenario 2 with the information extracted by the WYA^2 method in scenario 1.
For person number 2 the optimal bag of features are: [Mean, Typical Deviation], channel G, source Image, and [Dispersion, Energy, Entropy], channel G, source Histogram. This information causes that the WYA^2 detects as suspicious person the ID 1 (the suspicious target). Thus, an error occurs. To analyze this error the discriminative power of the bags of features obtained for person number 2 were tested over person number 1 in scenario 1. As expected these bags of features were able to discriminate properly the suspicious target. That is, when bags of features number 2 and number 14 are considered individually, they have a high discriminative power regarding person number 1. However these bags of features were not selected during the forward selection method proposed in WYA^2. Their discriminative power is very low when bags of features numbers 5 and 79 are in the model.
Images of people ID 1 and ID 2 in scenario 1, and person ID 1 in scenario 2 are presented in Figure 6.
When the WYA^2 was trained for the rest of the people in scenario 1, no suspicious people were detected in scenario 2. That is, the information that best discriminate these persons was not able to detect suspicious people in a new environment. This is expected since the bags of features that best discriminate the suspicious target in scenario 1 (labeled as 5 and 79), were not obtained as relevant bag of features for any other person.

5 CONCLUSIONS

In this paper, a novel methodology for human reidentification in multi-camera VideoSurveillance environment scenarios has been presented. The method, has been called WYA^2: "Why You Are Who You Are". WYA^2 is designed to select the best individual bags of features for each individual in the dataset

Figure 4: Sample Images from scenario 1 and scenario 2.

Table 1: Bag of features for R,G,B,H,S, and V channels. Six bags of features in each line in the table.

Id	Source	Bag of Feature
1-6	Image	[Mean, Typical Deviation]
7-12	Histogram	[Mean, Typical Deviation, Skewness, Kurtosis]
13-18	Histogram	[Dispersion, Energy, Entropy]
19-24	SGLD Matrix 0°	[Energy, Entropy, Inertia, Inverse Inertia]
25-30	SGLD Matrix 0°	[Correlation 0, Correlation 1, Correlation 2]
31-36	SGLD Matrix 0°	[Sum Mean, Sum Std, Sum Entropy]
37-42	SGLD Matrix 0°	[Dif Mean, Dif Std, Dif Entropy]
43-48	SGLD Matrix 45°	[Energy, Entropy, Inertia, Inverse Inertia]
47-54	SGLD Matrix 45°	[Correlation 0, Correlation 1, Correlation 2]
55-60	SGLD Matrix 45°	[Sum Mean, Sum Std, Sum Entropy]
61-66	SGLD Matrix 45°	[Dif Mean, Dif Std, Dif Entropy]
67-72	SGLD Matrix 90°	[Energy, Entropy, Inertia, Inverse Inertia]
73-78	SGLD Matrix 90°	[Correlation 0, Correlation 1, Correlation 2]
79-84	SGLD Matrix 90°	[Sum Mean, Sum Std, Sum Entropy]
85-90	SGLD Matrix 90°	[Dif Mean, Dif Std, Dif Entropy]
91-96	SGLD Matrix 135°	[Energy, Entropy, Inertia, Inverse Inertia]
97-102	SGLD Matrix 135°	[Correlation 0, Correlation 1, Correlation 2]
103-108	SGLD Matrix 135°	[Sum Mean, Sum Std, Sum Entropy]
109-114	SGLD Matrix 135°	[Dif Mean, Dif Std, Dif Entropy]
115-120	Histogram of LBP	[Mean, Typical Deviation, Skewness, Kurtosis]
121-126	Histogram of LBP	[Dispersion, Energy, Entropy]
127-132	SGLD Matrix 0° of LBP	[Energy, Entropy, Inertia, Inverse Inertia]
133-138	SGLD Matrix 0° of LBP	[Correlation 0, Correlation 1, Correlation 2]
139-144	SGLD Matrix 0° of LBP	[Sum Mean, Sum Std, Sum Entropy]
145-150	SGLD Matrix 0° of LBP	[Dif Mean, Dif Std, Dif Entropy]
151-156	SGLD Matrix 45° of LBP	[Energy, Entropy, Inertia, Inverse Inertia]
157-162	SGLD Matrix 45° of LBP	[Correlation 0, Correlation 1, Correlation 2]
163-168	SGLD Matrix 45° of LBP	[Sum Mean, Sum Std, Sum Entropy]
169-174	SGLD Matrix 45° of LBP	[Dif Mean, Dif Std, Dif Entropy]
175-180	SGLD Matrix 90° of LBP	[Energy, Entropy, Inertia, Inverse Inertia]
181-186	SGLD Matrix 90° of LBP	[Correlation 0, Correlation 1, Correlation 2]
187-192	SGLD Matrix 90° of LBP	[Sum Mean, Sum Std, Sum Entropy]
191-198	SGLD Matrix 90° of LBP	[Dif Mean, Dif Std, Dif Entropy]
199-204	SGLD Matrix 135° of LBP	[Energy, Entropy, Inertia, Inverse Inertia]
205-210	SGLD Matrix 135° of LBP	[Correlation 0, Correlation 1, Correlation 2]
209-216	SGLD Matrix 135° of LBP	[Sum Mean, Sum Std, Sum Entropy]
217-222	SGLD Matrix 135° of LBP	[Dif Mean, Dif Std, Dif Entropy]

Table 2: Most relevant bags of features for each person in scenario 1 and suspicious detected using that information in scenario 2.

Person ID	Bag of Features ID	Suspicious Detected
1	5,79	1
2	2,14	1
3	6,167,16	Not Detected.
4	23,168,196	Not Detected.
5	27,122,138	Not Detected.
6	14,85,162	Not Detected.
7	14,175	Not Detected.
8	99,107,196	Not Detected.
9	19,99,211	Not Detected.
10	6,132,36,9	Not Detected.
11	138,63,1,162	Not Detected.
12	102,70,150	Not Detected.
13	146,177,114,112,51	Not Detected.
14	146	Not Detected.
15	112,197	Not Detected.
16	114,5	Not Detected.
17	144,117,93,9,97	Not Detected.
18	112,135	Not Detected.

using a forward selection method in a Support Vector framework. The proposed method has been tested over a VideoSurveillance dataset. Overall, the results have been promising and the proposed methodology can serve as the foundation for further research.
Future research directions include to apply several methods for the fusion of kernels information during the SVM train. In addition, it will be necessary to test our methodology in additional datasets to show their relative performance when compared with other re-identification methods. Besides, It could be added other features not based on colour and texture, as for example gate or Gabor features, which are less sensitive to light variations. In (Moctezuma et al., 2015), it was shown that it is very important to weight the features according to each scenario. In this paper, we have shown that it is very important to consider the features according to each person. These features are what makes you who you are.

Figure 5: Suspicious target to be detected in scenario 2.

ACKNOWLEDGMENTS

This work is supported by the Ministerio de Economa y Competitividad from Spain INVISUM (RTC-2014-2346-8). This work has been part of the ABC4EU project and has received funding from the European Unions Seventh Framework Programme for research, technological development and demonstration under grant agreement No 312797.

REFERENCES

(2014). INVISUM. http://www.invisum.es.

An, L., Chen, X., Kafai, M., Yang, S., and Bhanu, B. (2013). Improving person re-identification by soft biometrics based reranking. In *Seventh International Conference on Distributed Smart Cameras, ICDSC 2013, October 29 2013-November 1, 2013, Palm Springs, CA, USA*, pages 1–6.

de Diego, I. M., Muñoz, A., and Moguerza, J. M. (2010). Methods for the combination of kernel matrices within a support vector framework. *Machine Learning*, 78(1-2):137–174.

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, 67(5):786–804.

Li, A., Liu, L., and Yan, S. (2014). Person re-identification by attribute-assisted clothes appearance. In *Person Re-Identification*, pages 119–138.

Martinson, E., Lawson, W. E., and Trafton, J. G. (2013). Identifying people with soft-biometrics at fleet week.

Figure 6: Target persons in scenario 1 that generate a detected alert in scenario 2, and suspicious target.

In *ACM/IEEE International Conference on Human-Robot Interaction, HRI 2013, Tokyo, Japan, March 3-6, 2013*, pages 49–56.

Moctezuma, D., Conde, C., de Diego, I. M., and Cabello, E. (2015). Soft-biometrics evaluation for people re-identification in uncontrolled multi-camera environments. *EURASIP J. Image and Video Processing*, 2015:28.

Moon, H. and Pan, S. B. (2010). A new human identification method for intelligent video surveillance system. In *Proceedings of the 19th International Conference on Computer Communications and Networks, IEEE ICCCN 2010, Zürich, Switzerland, August 2-5, 2010*, pages 1–6.

Ojala, T., Pietikainen, M., and Harwood, D. (1994). Performance evaluation of texture measures with classification based on kullback discrimination of distributions. In *Pattern Recognition, 1994. Vol. 1-Conference A: Computer Vision & Image Processing., Proceedings of the 12th IAPR International Conference on*, volume 1, pages 582–585. IEEE.

Roman, I. S., de Diego, I. M., Conde, C., and Cabello, E. (2017). Context-aware distance for anomalous humantrajectories detection. In *to appear in Pattern Recognition and Image Analysis - 8th Iberian Conference, IbPRIA 2017, Faro, Portugal, June 20-23, 2017. Proceedings*.

Tome-Gonzalez, P., Fiérrez, J., Vera-Rodríguez, R., and Nixon, M. S. (2014). Soft biometrics and their application in person recognition at a distance. *IEEE Trans. Information Forensics and Security*, 9(3):464–475.