A GRAPH-BASED SIGNATURE GENERATION FOR PEOPLE RE-IDENTIFICATION IN A MULTI-CAMERA SURVEILLANCE SYSTEM

T. D'Orazio[1] and C. Guaragnella[2]
[1]*ISSIA CNR, via Amnedola 122/D-I 70126 Bari, Italy*
[2]*DEE Politecnico di Bari, Via Orabona 70126 Bari, Italy*

Keywords: Multi-view Video Analysis, People Re-identification, Color Features, Graph-based Signature.

Abstract: In this paper we investigate the problem of people re-identification in case of non overlapping and non calibrated cameras. We propose a novel method for signature generation that considers both color and spatial features along a video sequence and a distance measure to estimate the similarity between silhouettes. A graph based representation has been introduced to model different people, in which uniform regions represent nodes and contiguities among regions represent edges. Comparisons with a standard approach based on histogram similarity have been provided to evaluate the proposed methodology.

1 INTRODUCTION

In this paper, the problem of people tracking in the context of a non overlapping views of a multiple camera video surveillance system, commonly referred to as the *people re-identification* problem, is addressed.

In recent literature, different re-identification methods have been developed, some on them focusing on the trajectories matching of people moving in the scene, others focusing solely on the appearance of the body. The latter are referred to as *appearance-based methods*, and can be grouped in two sets. The first group is composed by the single-shot methods (Chai et al., 2010; Alahi et al., 2010; de Oliveira and de Souza Pio, 2009; Gheissari et al., 2006; Mazzeo et al., 2009; D'Orazio et al., 2009), that model a person analyzing a single image and are applied when tracking information is absent. The second group of algorithms encloses the multiple-shot approaches; they employ multiple images of a person (usually obtained via tracking) to build a signature ((Bazzani et al., 2010; Cong et al., 2010; Hamdoun et al., 2008)). Some approaches try to learn the camera network topology in order to simplify the people association problem by predicting the relationship between events like a person exiting the camera view in a given location and the time lag between its reappearance in another view of a camera placed in another location. Alternative approaches to people tracking across multiple un-calibrated cameras use *gait analysis* as a new biometrics technique.

In this paper we investigate the problem of people re-identification and in particular we consider the more general case of non overlapping and non calibrated cameras. Neither constraints on the knowledge of camera view topologies, nor possible predictions of the expected time/location of people appearance among neighbor cameras are imposed. We propose a novel method for signature generation that considers both color features and spatial distribution of uniform regions along a video sequence and a distance measure to estimate the similarity of different sequences. Comparisons with other methodologies on a standard data sets have been provided to evaluate the proposed approach.

2 THE PROPOSED APPROACH

In order to generate a robust signature characterizing people moving in a camera field of view, it is necessary to extract moving people and track them along the video sequence. In this paper a background subtraction algorithm is proposed, resulting as an evolution of the works proposed in (Spagnolo et al., 2005). At the end of this step, people silhouettes have been extracted in each frame of the sequence; assume also that the image resolution is enough to appreciate colors and textures of clothes. The proposed method

D'Orazio T. and Guaragnella C..
A GRAPH-BASED SIGNATURE GENERATION FOR PEOPLE RE-IDENTIFICATION IN A MULTI-CAMERA SURVEILLANCE SYSTEM.
DOI: 10.5220/0003842104140417
In *Proceedings of the International Conference on Computer Vision Theory and Applications* (VISAPP-2012), pages 414-417
ISBN: 978-989-8565-03-7
Copyright © 2012 SCITEPRESS (Science and Technology Publications, Lda.)

is based on the evaluation of color similarity among uniform regions and the extraction of robust relative geometric information that are persistent also when people move in the scene. The signature can be estimated along a number of frames, and a distance measure is introduced to compare signatures extracted by different video sequences. The method consists of the following steps: 1) first of all for each frame a segmentation of uniform regions is carried out; 2) for each region some color and area information are evaluated; 3) the extracted data are evaluated on a number of consecutive frames; 4) a connected graph is generated (nodes contain the information of uniform regions such as color histograms and area occupancy, while connections among nodes contain information on the contiguity of regions); 5) a similarity measure is introduced to compare graphs generated by different video sequences that considers some relaxation rules to handle the different appearances of the same person when observed by different point of views.

In the remaining part of this section we describe the methodologies for the region information extraction, the edge information extraction, the graph generation, and the similarity measure to compare graphs.

Moving Region Segmentation. The detection of moving objects in a complex background scene is an important initial step for any computer vision application. In this paper we propose a modification of a statistical background subtraction algorithm (Chiu et al., 2010) producing good performance in the context of people detection both in indoor and outdoor contexts. In order to construct a reference background model, a video sequence during which moving objects can pass through the scene is observed. For each pixel the values of the three RGB components are recorded and the clusters of the most probable are maintained. The background values are those belonging to the most probable clusters if the corresponding probability is over a dynamic threshold. The background extraction phase continues until all the pixels have been extracted. Then moving objects in the foreground can be segmented by comparing each pixel of the current frame with the corresponding one in the background model. At this point, the result of this segmentation step is highly dependent on the lighting variations, the presence of multiple shadows (especially in indoor contexts) and the similarity between foreground and background values (camouflage effects). For this reason we have introduced new procedures for background model updating and shadow removal both based on a cross correlation approach.

Extraction of Uniform Region. After the extraction of moving blobs by the background subtraction and shadow removing algorithms, the resulting segmented areas are analyzed for the evaluation of color and shape information. A color-based segmentation has been applied to detect, inside each blob, uniform connected regions. The segmentation algorithm proposed in (Nock and Nielsen, 2004) has been used. In order to control the coarseness of the segmentation, the method uses a parameter Q: the smaller it is, the less numerous are the region in the final segmentation. The resulting segmentation is processed to extract inter-region and intra-region information. For each region a feature vector is considered containing different information such as: the color histogram, the mean color, the center of mass of the region, the area, the perimeter, the region position with respect to the whole body figure (head, central and lower). The body silhouette has been divided in three areas of different dimensions: the *head* part corresponds to $1/6$ of upper region, the *central* part falls between $1/6$ and $1/3$ of the silhouette, while the *lower* part corresponds to the remaining inferior area. All the other features have been normalized with respect to the ratio between the area dimension and the total body area. The relations between each couple of regions are examined and an adjacency matrix is generated. This step has been repeated for all the images of the same person of the sequence. Corresponding regions in the sequence have been extracted and the feature vectors have been updated to weight the measure coming from different frames.

Connected Graph Signature. The proposed people signature uses a relational graph. Such a graph consists of a finite numbers of nodes and a finite number of edges. Each node has a number of associated attributes that correspond to the elements of the feature vector described above. In this way, detecting the similarity among people turns into determining the similarity among graphs, which is also referred as graph matching. In this section we describe the processing steps to extract the information for graph generation, while in the following section we will describe the proposed graph matching procedure. First of all, in order to reduce the number of considered regions when people wear dresses with varying texture, small regions that fall inside larger regions are joined with the external ones and the corresponding feature vectors are updated. Small regions that are adjacent other larger ones are neglected (for example those corresponding to hands, shoes, hair or small shadow areas). In figure 1 an example of the region extraction process is reported. The considered regions are representative of the nodes of the graph, while the edges will represent the adjacency relations among the nodes. For example in figure 1 the central node corresponding to

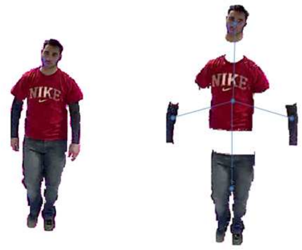

Figure 1: On the left the considered person, on the right the resulting regions after the segmentation and the combination of internal regions.

the shirt is connected to all the other nodes since this region is adjacent to all the other regions (head, arms, and trousers).

Similarity Measure Among Graphs. Connected graphs have been largely used for object recognition: given a database of known objects and a query, the task is to retrieve one objects from the database that is similar to the query. However, in the context of people re-identification, the same person observed by different points of view produces different signatures both in terms of resulting nodes in the graph, both in terms of attributes for each node. Therefore it is necessary to consider some degree of error tolerance in the graph matching process, considering both subgraph matching and not precise matches between corresponding node attributes. Since we observed that the central nodes are those that contain the most useful data for an initial people discrimination we started the graph matching procedure from these nodes. In particular among all the nodes that belong to the cental area, we selected the one corresponding to the torso as the node whose center of mass is the closest to center of mass of the whole silhouette and whose area is larger than a percentage of the total area. Then, for each couple of graphs, starting from these central torso nodes the connected nodes belonging to the upper and the lower regions have been considered to evaluate geometric relationships based on distances and color/texture characteristics. The total similarity is carried out summing up the weighted differences among all the couples of nodes as follows:

$$D_{tot}(n,m) = w_T \cdot D(n_T,m_T) + w_{CN} \cdot D(n_{CN},m_{CN}) + \\ + w_{UP} \cdot D(n_{UP},m_{UP}) + w_{LO} \cdot D(n_{LO},m_{LO})$$

where n,m are the two considered graphs, n_T,m_T are torso nodes, n_{CN},m_{CN} are central nodes, n_{UP},m_{UP} are upper nodes, n_{LO},m_{LO} are lower nodes, while $D()$ is the measure of similarity evaluated on the distance between the nodes' attributes. In particular we decided to weight in a different way the differences between central, upper, and lower nodes. The central nodes corresponding to the torso are considered the most significant for the re-identification problem, then the weight w_T has been fixed to 0.4. The remaining weights have been set to a fixed values of 0.2. If the two considered graphs are not isomorph (for example they have a different number of central, lower, and upper nodes) comparisons are carried out considering all the other possible combinations. The match providing the lowest distance among the different possible matches is considered as the correct one.

3 EXPERIMENTAL RESULTS

Tests have been carried out on a real data set acquired by two different cameras placed in two different corridors of an office. People observed in one camera have to be recognized when they pass in the second camera field of view. In this paper results are presented referred to 35 target persons compared against 47 persons observed in the second camera. In order to evaluate the performances of the proposed method we have compared the results with a color position histogram approach extracted by the work presented in (Cong et al., 2010). The silhouette signature is generated dividing the silhouette in n equal parts and characterizing each part with a color histogram. Preliminary tests have been carried out to decide the number of regions that provide the best performances.

On the same test sets, the results obtained with the color position histogram signature and the proposed graph-based signature have been compared. For each view a set of key frames in which people are entirely visible has been considered in order to characterize the signature with the two methods. The re-identification has been assigned to the couple of images that obtains the minimum value of the similarity distance. The considered metric is the standard histogram distance both for the color position histogram signature, and the proposed graph based approach. Several experiments were carried out in order to detect the best color space and the best quantization values that guarantee the invariance to lighting variations. In table 1 the best results are reported, obtained in the HSI space and with a color map with 1050 colors.

Table 1: The results obtained with the color position histogram signature and the proposed graph based signature.

	True Positive	False Negative
Col. Pos. Hist Sig.	82%	18%
Graph based Sig.	88%	12%

The proposed approach improves the detection performances because the evaluation of similarity be-

Figure 2: In the first line the result of the color position histogram signature, while in the second line the result of the proposed graph based signature. The first one on the left is the target object and the remaining ones are the objects identified with a growing similarity distance.

tween uniform regions instead of body part regions with fixed dimensions allows a more precise characterization of the color differences. Color histograms, even if applied in different stripes of the body, lose the spatial information about the color distribution. Instead the initial selection of uniform regions, that characterize the nodes of the graph and the comparisons between couple of uniform regions maintain the spatial information about colors.

In the first line of figure 2 one case of wrong detection of the color position histogram signature method is reported, while in the second line the corresponding result obtained with the proposed method is shown. The evaluation of histograms in constant stripes of the body silhouette makes similar two different persons whose color components are the same but their spatial distribution is different. This is particular evident in the two considered cases, in which people wear similar dresses unless for an inscription on the chest.

The proposed graph based signatures, anyway, suffers in few cases of people re-identification. In particular when people wear similar dresses with the same colors, but they differ only for small details, such as the shoes color, the hair color, and so on, the method is not able to disambiguate since the simplification of small internal region inclusion or small extreme region elimination causes the lost of precious information for the discrimination ability. These are, of course, extreme situations in which many feature based signature approaches fail. Future work will be addressed to the solution of this problem, introducing an adaptive similarity measures that will consider all the possible uniform regions in the graph matching procedure only when the similarity among main regions does not allow the disambiguation.

REFERENCES

Alahi, A., Vandergheynst, P., Bierlaire, M., and Kunt, M. (2010). Cascade of descriptors to detect and track objects across any network of cameras. *Computer Vision and Image Understanding*, 114:624–640.

Bazzani, L., Cristani, M., Perina, A., Farenzena, M., and Murino, V. (2010). Multiple-shot person reidentification by hpe signature. In *20th International Conference on Pattern Recognition*, pages 1413–1416, Istanbul Turkey.

Chai, Y., Takala, V., and Pietikainen, M. (2010). Matching groups of people by covariance descriptor. In *20th International Conference on Pattern Recognition*, pages 2744–2747, Istanbul Turkey August 23-August 26.

Chiu, C. C., Ku, M., and Liang, L. (2010). A robust object segmentation system using a probability based background extraction algorithm. *IEEE Transaction on Cirtuits and Systems for Video Technology*, 20(4):518 – 528.

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

de Oliveira, I. and de Souza Pio, J. (2009). People reidentification in a camera network. In *Eighth IEEE International Conference on Dependable*, Autonomic and Secure Computing, pages 461–466.

D'Orazio, T., Mazzeo, P., and Spagnolo, P. (2009). Color brightness transfer function evaluation for non overlapping multi camera tracking. In *Third ACM/IEEE International Conference on Distributed Smart Cameras*, Como Italy.

Gheissari, N., Sebastian, T. B., Tu, P. H., Rittscher, J., and Hartley, R. (2006). Person reidentification using spatiotemporal appearance. In *Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, CVPR.

Hamdoun, O., Moutarde, F., Stanciulescu, B., and Steux, B. (2008). Person re-identifcation in multi-camera system by signature based on interest point descriptors collected on short video sequences. In *Proceedings of the IEEE Conference on Distributed Smart Cameras*, pages 1–6.

Mazzeo, P., Spagnolo, P., and D'Orazio, T. (2009). Object tracking by non-overlapping distributed camera networks. In *Advanced Concepts for Intelligent Vision Systems September*, ACIVS, Mercure Chateau Chartrons, Bordeaux, France.

Nock, R. and Nielsen, F. (2004). Statistical region merging. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 26(11).

Spagnolo, P., D'Orazio, T., M.Leo, and Distante, A. (2005). Advances in background updating and shadow removing for motion detection algorithm. In *Lecture Notes in Computer Science*, pages 398–406.

SCIENCE AND TECHNOLOGY PUBLICATIONS