RE-IDENTIFYING PEOPLE IN THE WILD

*Javier Oliver, Alberto Albiol, Antonio Albiol, José Manuel Mossi**

Universitat Politècnica de València
iTEAM Research Institute
Camino de Vera s/n, 46022 Valencia, Spain

ABSTRACT

People re-identification in uncontrolled scenarios is a difficult task since people appearance may significantly vary along time due to changes in illumination, changes in the person pose or the presence of undesired objects in the scene. In order to cope with this temporal variability in the person appearance, we introduce the concept of Bags of Appearances (BoA) to describe each person. A BoA is a container of color features that fully represents a person by collecting all their different appearances along time. Matching of bags is performed in a probabilistic framework by accumulating the probability of matching for all of the elements of each bag. Experiments have been conducted in a real shop where clients were re-identified at the entrance and exit. Results improve state-of-the art methods and confirm that our proposal successfully copes with rough changes in the people appearance.

Index Terms— Re-Identification, Appearance Descriptor, Knn, Kinect

1. INTRODUCTION

In this work we address the problem of people re-identification in shopping areas, using two non-overlapping cameras covering the entrance and the exit. In this kind of scenarios, people appearance may significantly vary along time and among cameras. Variations in the person appearance may be caused by changes in illumination, occlusions, the use of shopping baskets, the appearance of carried objects, or simply caused by the changes in the person pose due to the articulation of the human body.

People re-identification in camera networks has traditionally been addressed using appearance information, which includes color, texture and shape features. Color histograms have been proposed in many works [1] [2] [3] [4] [5] to describe the person appearance. Dikmen et al. [6] combine color histograms with spatial information by dividing the image into a fixed grid where color histograms are extracted. Albiol et al. [7] finely segment each person and divide their silhouette into horizontal bands in which they calculate the temporal and spatial average color. Hu et al. [8] model the color appearance over the silhouettes principal axis, which requires a robust background subtraction. Bazzani et al. [9] combine local patterns together with color histograms to provide more complete descriptors.

Algorithms for describing people can be also grouped into holistic [10] [11] and part-based [12] [13] [14]. An example of the holistic representation is the Bag-of-features approach [15] [16] [17], which represents an image or a portion of an image by extracting local descriptors according to a predefined diccionary of appearances. The holistic approach has shown the best results in challenging scenarios. However, the part-based strategy is becoming popular and provide promising results.

Re-identification methods can also be grouped into single or multiple-shot. In the former, the person is described using a single image [18]. In the latter, the person is described by different images acquired by the same camera using tracking or by several cameras [9]. Multiple-shot methods can exploit other contextual cues such as spatio-temporal reasoning [19].

The use of the temporal information can be exploited to learn the person appearance and their variation over time. Bedagkar-Gala et al [20] extracts the meaningful color appearance of the person in a part-based approach by exploiting spatio-temporal information. The model learns the principal appearance of the person, discards noise and outliers, and retains the relevant variations. Hamdoun et al [10] describe the main person appearance and their variability along time by accumulating interest points at each frame. Gheissari et al. [21] extract invariant spatio-temporal signatures combining normalized color and salient edgels histograms.

Recently, authors are using 3D information to tackle the re-identification problem. Baltieri et al [22] use a calibrated system to generate 3D representations of the people appearance. Albiol et al [7] propose a camera network composed by rgb-depth camera nodes, where segmentation and tracking are performed using depth information. Description is performed by dividing the person into horizontal stripes and calculating the mean colors at each one. Their method provides a temporal signature of the person appearance that is summarized into a mean descriptor, named Bodyprint. Oliver et al [23] propose

*The work presented in this paper has been funded by the Spanish Ministry of Science and Technology under the CICYT contract TEVISMART, TEC2009-09146.

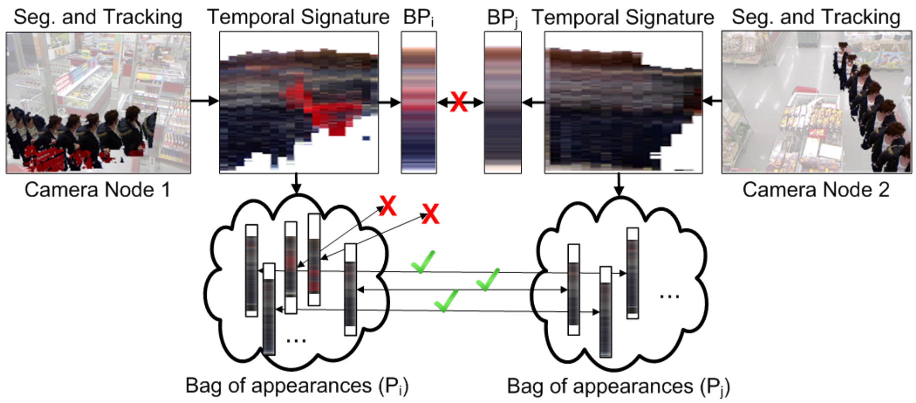

Fig. 1. Example of the variability of the person appearance along time seen from different cameras.

the generation of 3D Bodyprints, which map the appearance features into an 3D cylindrical grid.

The rest of the paper is organized as follows: Section 2 discusses the contribution of this paper to prior work. The proposed algorithm description is provided in Sec. 3 and the assessment is performed in Sec. 4. Finally, conclusions are drawn in Sec. 5.

2. RELATION TO PRIOR WORK

In our previous work [7], we developed a re-identification system that used kinect devices at each camera node to segment and track people. The great advantage of using a depth sensor is that the problems of image segmentation and 3D alignment can be solved precisely. In this work, we use the person segmentation and tracking algorithms of our previous work (see [7] for details). For each frame of a tracked person, we described his appearance by dividing the person height into a set of horizontal stripes and calculating the mean colors at each stripe. We called Temporal Signature to the image obtained by stacking all the frame appearances along time. Figure 1 illustrates this concept, where the same person is seen from two different cameras, named node 1 and node 2. For both cameras, the person segmentation along time has been overimposed on the background image. In our experiments, height is quantized using 110 vertical bins of 2cm.

Finally, for each person we extracted a Bodyprint by taking a weighted average of the mean colors along time at each stripe. Re-identification using Bodyprint descriptors demonstrated to yield good results in real scenarios. However, the algorithm failed in cases where the person's appearance drastically changed along time. An example of this situation is shown in Figure 1 where the mean colors of the regions belonging to the pelvis and legs are clearly affected by the influence of the shopping basket, which is taken from the pile and is connected to the body due to a bad segmentation. In this work we present a different person representation and matching to face this problem.

3. ALGORITHM DESCRIPTION

3.1. Bag of appearances (BoA)

In this work, we propose a different person representation which we call Bag of Appearances (BoA) to deal with the problems introduced in the previous section.

BoA characterizes a person by collecting all person's frame appearances into a bag (which is different for each person). The BoA is useful because it allows to model the person's appearance from different points of view, poses and deal with outliers. Figure 1 illustrates this concept, where a cloud of frame appearances collects all the appearances for a particular person. The concept of BoA is similar to the Bag of Features [15] [16] [17]. However, fundamental differences arise due to the nature of the color features and the huge variability in the items representing each class. The main difference is that we use separate bags for each person and do not perform global clustering. Although clustering of the appearances in each bag is a promising idea to reduce the computational cost of the algorithm, we decided not to use it in this work and to focus only on how multi-instance representation affects to the re-identification rate.

Since re-identification is performed across different cameras it is very important to compensate for illumination variations. To do so, we follow the same approach as in [7], where the brightness of the i-th person, $\mathcal{B}_i$, is obtained as the mean of all the R, G, B pixel values of a person track $\mathcal{P}_i$:

$$\mathcal{B}_i = \sum_{p_i \in \mathcal{P}_i} \frac{R_{p_i} + G_{p_i} + B_{p_i}}{3} \tag{1}$$

then $\mathcal{B}_i$ is subtracted from the color components of the appearance vectors.

3.2. Probabilistic Framework for BoA matching

Suppose that we want to find the person X_i from camera A among the set of N_B persons from camera B denoted as:

$\{Y_j\}_{j=1..N_B}$. If person X_i is represented by his bag of appearances, $\{\mathbf{x}_{im}\}_{m=1..N_i}$, for each $\mathbf{x}_{im}$ we find its k-nearest neighbours among all the items from all the bags of camera B $\{\mathbf{y}_{jn}\}_{j=1..N_B,n=1..N_j}$. The comparison between appearance vectors has been carried out using the weighted correlation described in [7] which allows to deal with missing data. The information provided by the k nearest neighbours of $\mathbf{x}_{im}$ is accumulated in a histogram H_i. Each bin of H_i corresponds to a different person of camera B. However, in order to incorporate the information about the confidence of each element, we convert the distances into posterior probabilities. The posterior probabilities are estimated from the distance likelihoods using the Bayes rule. Figure 2 shows the likelihood distributions of the distance between $\mathbf{x}_{im}$ and $\mathbf{y}_{jn}$ for the match and mismatch cases. In this work, we approximate these distributions using the Gaussian function.

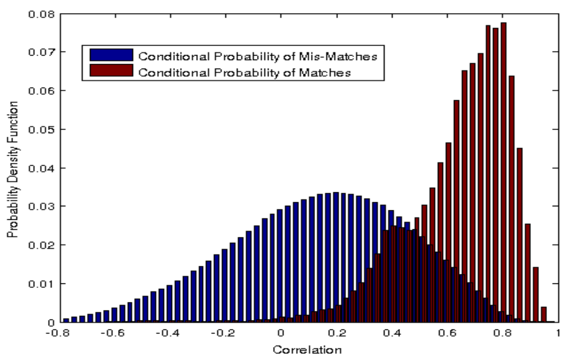

Fig. 2. Likelihood distribution of appearance distance values for the match and mismatch cases.

As mentioned above, our weighted correlation, used to compare frame appearances, relies only on the common visible parts of both appearance vectors to deal with missing data (oclusions). Unfortunately, this method discards the person's height information which is also very discriminant. In order to enhance the probability of matching, we obtain the difference of the average heights between pairs of persons' tracks Δh . This difference is also transformed into posterior probability P_h that measures the probability that two tracks are from the same person given Δh. Since both, appearance and height information, are converted into probabilities and they are independent, we multiply their values to fuse their information.

The overall process of matching can be seen in the self-explanatory pseudocode at Algorithm 1.

Definition X := People in node A;
Y := People in node B;
h_X := Heights of people in node A;
h_Y := Heights of people in node B;
C_i := Number different people in the datasets;
Dataset normalisation Remove brightness variations of each person independently in nodes A and B;
foreach *person* X_i *in Node* A **do**
    Define H_i as the class histogram that represents the probability of matching person X_i with dataset Y;
    foreach *appearance* $\mathbf{x}_{im}$ **do**
        % Find k nearest neighb. of $\mathbf{x}_{im}$ in cam. B;
        foreach $\mathbf{y}_{jn} \in \{knn\_of(\mathbf{x}_{im})\}$ **do**
            % Convert distance to probability
            $P_a \leftarrow Dist(\mathbf{x}_{im}, \mathbf{y}_{jn})$
            % Calculate height difference between persons X_i and Y_j;
            $\Delta h_{ij} = abs(h_{X_i} - h_{Y_j})$;
            % Transform Δh_{ij} to probability;
            $P_h \leftarrow \Delta h_{ij}$;
            % Add observation to probability histogram;
            $H_i(j) = H_i(j) + P_a.P_h$;
        end
    end
    % Find person match;
    $match = argmax(H_i)$;
end

Algorithm 1: People matching using bag of appearances

4. EVALUATION

In this section we will analyse the system performance by calculating the probability of re-identification of the system. To that end, we identify two stages. In the former, the parameters of the Normal distributions that model the correlation and height difference need to be retrieved. Also, the parameters of the knn algorithm need to be set up. In the latter, we use the parameters found during training to test the system performance.

4.1. Database

In order to be able to test our algorithm, we used our public database [24], which gathers information of 144 different people in a store behaving without any restrictions. Two different camera nodes with no overlapping area have been considered in this experiment. Half of the dataset has been used for tuning up the model parameters. The other half has been used for testing.

4.2. Training

In this stage we first focus on the extraction of the Gaussian parameters that model the correlation distribution of the same-class matches and different-class matches. Figure 3.2 shows the distribution of both groups, where it can be easily seen that both approximately follow a Gaussian shape. We model the *match* distribution by the Normal $N(0.63, 0.19)$, and the *mismatch* distribution by $N(0.13, 0.32)$.

Second, we focus on the extraction of the Gaussian distribution parameters for the case of difference of heights, where we model the *match* and *mismatch* distributions by $N(0.60, 1.89)$ and $N(0.55, 27.71)$ respectively.

In the third place, we seek the optimal number of neighbours, k_{opt}, for the knn algorithm by using the same training dataset. The maximum probability of re-identification P_R varying the number of neighbours from 1 to 10 is found for $k_{opt} = 2$.

4.3. Test

After tunning the parameters of the knn algorithm and modeling the correlation and heights into a probabilistic framework, we test our algorithm using an independent test set. We obtain a probability of re-identification of 58.11%, which clearly outperforms the re-identification rates obtained using Bodyprints for the same dataset, which was of 53.05%. Figure 3 shows some challenging people that motivate the use of the BoA strategy since their temporal appearance presents significant variance. Note that for all these cases, the mean colors retrieved by Bodyprints would not be fully representative of each person. For the man with label 23 pushing a trolley, there is a period of time where his hand is segmented correctly and is connected to the body, but there is also a period where the hand does not appear. The women with labels 50, 43 and 28 are taking a shopping basket from the pile. The inaccurate segmentation yields a red color during several frames of the temporal signature. On other hand, both the women with labels 28 and 14 hold a white paper at the end of the signature. Finally, the man with label 3 shows a problem of illumination variability together with the contribution of the shopping basket.

The reason of the improvement achieved by BoA relies on the independent treatment of the temporal information of the appearance. Figure 4 shows the accumulated probability of matching for two different persons. The sparsity of the histogram of person with ID 28 shows that several people presented similar appearance to the person of interest during small instants, mainly at the central and final frames of the track. However, the contribution of the first frames, where the appearance is not corrupted yet, are essential for providing the correct match of this person. The histogram of person 43 shows an easier example of matching, although most of the first frames are badly classified.

Fig. 3. Examples of challenging tracks

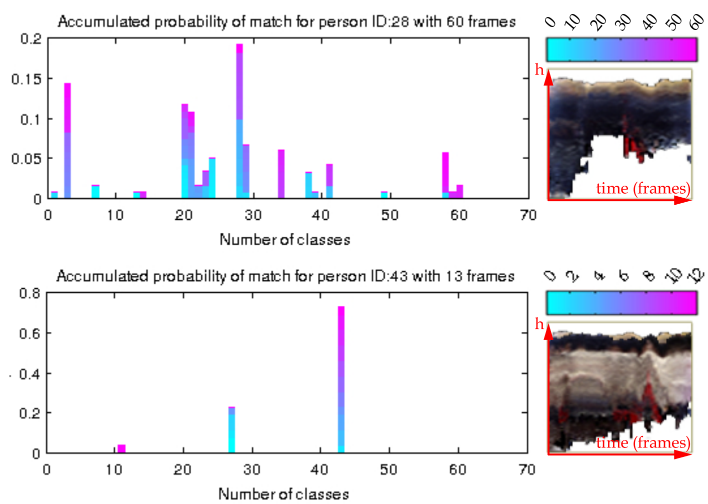

Fig. 4. Accumulated probability-of-matching of a person at the entrance against 73 different people gathered at exit. Top and bottom figures represent the probabilities for people with IDs of 28 and 43, respectively.

5. CONCLUSIONS

In this paper we have introduced the concept of Bag of Appearances (BoA) for long-term appearance-based people re-identification using depth cameras. This new methodology consists of describing a person by their multiple representations along time. The BoA has been demonstrated to work in real and complex scenarios where outliers and other disrupting effects are likely to occur during tracking. Re-identification rates obtained by our approach clearly outperforms state-of-the-art algorithms for re-identification using depth cameras.

6. REFERENCES

[1] Oliveira I.O. and Souza Pio J.L., "People reidentification in a camera network," in *Eighth IEEE International Conference on Dependable, Autonomic and Secure Computing*, dec. 2009, pp. 461–466.

[2] Farenzena M., Bazzani L., Perina a., Murino V., and Cristani M., "Person re-identification by symmetry-driven accumulation of local features," in *Proceedings of CVPR*, San Francisco, CA, USA, 2010.

[3] Prosser B., Zheng W.S., Gong S., and Xiang T., "Person re-identification by support vector ranking," in *Proceedings of the BMVC*. 2010, pp. 21.1–21.11.

[4] Kuo C.H., Huang C, and Nevatia R., "Inter-camera association of multi-target tracks by on-line learned appearance affinity models," in *Proceedings of the ECCV*, Berlin, Heidelberg, 2010, pp. 383–396, Springer-Verlag.

[5] Zheng W.S., Gong S., and Xiang T., "Person re-identification by probabilistic relative distance comparison," in *Proceedings of CVPR*, june 2011, pp. 649–656.

[6] Dikmen M., Akbas E., Huang T.S., and Ahuja N., "Pedestrian recognition with a learned metric," in *Asian Conference in Computer Vision*, 2010.

[7] Albiol A., Albiol Al., Oliver J., and Mossi J.M., "Who is who at different cameras. matching people using depth cameras," *IET Computer Vision*, 2012.

[8] Hu W., Hu M., Zhou X., Tan T., Lou J., and Maybank S., "Principal axis-based correspondence between multiple cameras for people tracking," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 28, no. 4, pp. 663 –671, april 2006.

[9] Bazzani L., Cristani M., Perina A., Farenzena M., and Murino V., "Multiple-shot person re-identification by hpe signature," in *Proceedings of the ICPR*, Washington, DC, USA, 2010, pp. 1413–1416.

[10] Hamdoun, O., Moutarde, F., Stanciulescu, B. and Steux, B., "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences," in *Proceedings of Distributed Smart Cameras*, 2008.

[11] Doretto G., Sebastian T., Tu P., and Rittscher J., "Appearance-based person reidentification in camera networks: Problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, pp. 1–25, 2011.

[12] Weber M., Bauml M., and Stiefelhagen R., "Part-based Clothing Segmentation for Person Retrieval," in *IEEE International Conference on Advanced Video and Signal-Based Surveillance*, Aug. 2011, p. 6.

[13] Martinel N. and Micheloni C., "Re-identify people in wide area camera network," in *Computer Vision and Pattern Recognition Workshops (CVPRW)*, IEEE, Ed., Providence RI, June 2012, pp. 31–36.

[14] Aziz K.E., Merad D., and Fertil B., "People re-identification across multiple non-overlapping cameras system by appearance classification and silhouette part segmentation," in *Proceedings of Advanced Video and Signal-Based Surveillance*, Aug. 2011, p. 6.

[15] Varma M. and Zisserman A., "A statistical approach to texture classification from single images," *International Journal of Computer Vision: Special Issue on Texture Analysis and Synthesis*, vol. 62, no. 1–2, pp. 61–81, April 2005.

[16] Winn J., Criminisi A., and Minka T., "Object categorization by learned universal visual dictionary," in *Proceedings of ICCV*, oct. 2005, vol. 2, pp. 1800 –1807 Vol. 2.

[17] Fei-Fei L. and Perona P., "A bayesian hierarchical model for learning natural scene categories," in *Proceedings of CVPR*, june 2005, vol. 2, pp. 524 – 531 vol. 2.

[18] Wang X., Doretto G., Sebastian T. B., Rittscher J., and P. H. Tu, "Shape and appearance context modeling," 2007, pp. 1–8.

[19] Simonnet D., Lewandowski M., Velastin S., Orwell J., and Turkbeyler E., "Re-identification of pedestrians in crowds using dynamic time warping.," in *ECCV Workshops*, pp. 423–432.

[20] Bedagkar-Gala A. and Shishir K. Shah, "Multiple person re-identification using part based spatio-temporal color appearance model," in *Computational Methods for the Innovative Design of Electrical Devices'11*, 2011, pp. 1721–1728.

[21] Gheissari N., Sebastian T.B., and Hartley R., "Person reidentification using spatiotemporal appearance," in *Proceedings of CVPR*, 2006, pp. 1528–1535.

[22] Baltieri D., Vezzani R., Cucchiara R., Utasi A., BenedeK C., and Szirányi T., "Multi-view people surveillance using 3d information," in *ICCV Workshops*, 2011, pp. 1817–1824.

[23] Oliver J., Albiol A., and Albiol Al., "3D descriptor for people re-identification," in *Proceedings of ICPR*, Tsukuba, Japan, November 2012.

[24] http://www.gpiv.upv.es/videoresearch/personindexing.html.