Appearance-based Re-Identification of Humans in Low-Resolution Videos using Means of Covariance Descriptors

Jürgen Metzler
Fraunhofer Institute of Optronics, System Technologies and Image Exploitation IOSB
76131 Karlsruhe, Germany
juergen.metzler@iosb.fraunhofer.de

***Abstract*—The objective of human re-identification is to recognize a specific individual on different locations and to determine whether an individual has already appeared. This is especially in multi-camera networks with non-overlapping fields of view of interest. However, this is still an unsolved computer vision task due to several challenges, e.g. significant changes of appearance of humans as well as different illumination, camera parameters etc. In addition, for instance, in surveillance scenarios only low-resolution videos are usually available, so that biometric approaches may not be applied. This paper presents a whole-body appearance-based human re-identification approach for low-resolution videos. We propose a novel appearance model computed from several images of an individual. The model is based on means of covariance descriptors determined by spectral clustering techniques. The proposed approach is tested on a multi-camera data set of a typical surveillance scenario and compared to a color histogram based method.**

I. Introduction

There are several approaches for re-identifying humans in multi-camera networks. Depending on the image resolution of persons, biometric approaches can be applied such as re-identification by face [1], gait [2] or their combination [3], to name but a few. Other approaches use soft biometric features such as skin color, hair color, tattoos or other body decorations, height or width as well as behavioral traits of the individuals [4], [5]. However, if only low-resolution videos are available, these approaches may probably not applicable, for the simple reason that such features cannot be extracted due to the low resolution.

In this paper, we propose a human re-identification method based on the whole-body appearance. Appearance-based techniques perform the re-identification on color or texture information of person's clothing. Recently, there are several appearance-based approaches, an overview can be found in [6]. The use of local features for human re-identification in low-resolution videos are generally not adequate as persons don't often show up significant textures. One common suitable approach may be using color histograms which represent the whole-body appearances. Another approach is the use of covariance matrices as image region descriptors which have been proposed by Tuzel et al. [7]. The authors got convincing results for different computer vision tasks especially according to their discriminability (see e.g. [7], [8] and [9]). Two examples for human re-identification methods using covariance descriptors are [10] and [11]. One drawback here is the computation time caused mainly through the high number of necessary comparisons of the covariance descriptors. Furthermore, the grid-based approach is not adequate for low-resolution videos. In our proposed approach, we first build a representative set of means of covariance descriptors representing the whole-body appearance and compare only means with each other instead of comparing all descriptors with each other, which decreases the number of necessary comparisons by approximately factor 1000. Thereby, it is important to get representative and discriminative means that as well as their number are determined by spectral clustering techniques.

Our approach is presented in Section IV, after we treated the basics of the covariance descriptors and their corresponding Riemannian manifold in Section II as well as spectral clustering fundamentals in Section III. Experimental results are shown in Section V before we conclude our contribution.

II. Covariance Descriptors

Covariance descriptors were introduced by Porikli et al. [7]. A covariance descriptor represents an image region by a covariance matrix of image features. It proposes a natural way of fusing multiple features which might be correlated with each other, where diagonal entries of the covariance matrix represent the variance of each feature and the off-diagonal entries represent the correlations between the features. In other words, a covariance descriptor contains information about spatial and statistical properties of the image region as well as linear correlations between these properties.

As pointed out in [7] there are several advantages of using covariance descriptors:

- Support of scale invariant features/properties
- Invariant to mean changes (e.g. invariant to identical shifting of color values)
- Insensitive to noise
- Efficient fusion of multiple features

- Feature set may be easily extended or modified

Using covariance matrices as descriptors, there is need for non-Euclidean metrics since covariance matrices do not lie in a vector space. For that, the set of positive definite symmetric matrices can be formulated as Riemannian manifold as described in the following sections.

A. Covariance Descriptor Computation

Let R_1 be an image region. First, for each pixel inside R_1 features such as color, gradients, filter responses, etc. are computed. Then, d-dimensional feature vectors are constructed - one for each pixel inside R_1. For human re-identification we use the y-coordinates of the image pixels as well as the color values R, G and B.

Let$\{f_i\}_{(i=1\dots n)}$ be a set of feature vectors of the W-width and H-height rectangular R_1 and

$$f_i = (y, R(x,y), G(x,y), B(x,y))^{\mathrm{T}} \tag{1}$$

a feature vector at the pixel with the coordinates (x, y) and $R(x,y), G(x,y), B(x,y)$ the corresponding color values. The covariance matrix representing R_1 is then given by

$$Cov_{R_1} = \frac{1}{WH} \sum_{i=1}^{WH} (f_i - \mu_{R_1})(f_i - \mu_{R_1})^{\mathrm{T}} \tag{2}$$

where μ_{R_1} denotes the mean-vector of $\{f_i\}_{(i=1\dots n)}$.

More details about the covariance descriptor computation including an efficient method by using integral images can be found in [12].

B. Space of Positive Definite Covariance Matrices

A covariance matrix contains information about statistical dispersions and linear relationships of random variables. Let

$$Cov(X_i, X_j) =$$

$$E\left[(X_i - E(X_i))(X_j - E(X_j))\right], \tag{3}$$

$$i = 1 \dots n, j = 1 \dots n,$$

the pairwise covariances, the covariance matrix Σ is then given by

$$\sum = \begin{pmatrix} Cov(X_1, X_1) & \cdots & Cov(X_1, X_n) \\ \vdots & \ddots & \vdots \\ Cov(X_n, X_1) & \cdots & Cov(X_n, X_n) \end{pmatrix}. \tag{4}$$

The set of positive definite covariance matrices (nonsingular covariance matrices) describes a Riemannian manifold and is denoted by Sym_n^+. A Riemannian manifold or Riemannian space is a topological space that is only locally Euclidean: there is a tangent space at each element of the manifold (in our case at each covariance matrix). Hence, Euclidean geometry is not appropriate to compare covariance matrices. In past years functions like the trace or determinant of a covariance matrix have been used to measure the similarity. However, these measures are not suitable [13] and hence Foerstner et al. [13] and Pennec et al. [14] deduced invariant Riemannian metrics. These metrics are equivalent, thus it is sufficient to concentrate on Pennec's:

$$< y, z >_{\sum_1} = trace\left(\Sigma_1^{-\frac{1}{2}} y \Sigma_1^{-1} z \Sigma_1^{-\frac{1}{2}}\right), \tag{5}$$

Σ_1 is a covariance matrix and y, z are elements of the tangent space at Σ_1. y and z are computed by a diffeomorphism which maps elements of the tangent spaces into the manifold of covariance matrices. Associated to the Riemannian metric (5) it is defined by the exponential map

$$exp_{\Sigma_1}(y) = \Sigma_1^{\frac{1}{2}} \exp\left(\Sigma_1^{-\frac{1}{2}} y \Sigma_1^{-\frac{1}{2}}\right) \Sigma_1^{\frac{1}{2}}. \tag{6}$$

The exponential map is global in Sym_n^+ and thus there is an inverse mapping (logarithmic map) which is uniquely defined everywhere:

$$log_{\Sigma_1}(\Sigma_2) = \Sigma_1^{\frac{1}{2}} \log\left(\Sigma_1^{-\frac{1}{2}} \Sigma_2 \Sigma_1^{-\frac{1}{2}}\right) \Sigma_1^{\frac{1}{2}}. \tag{7}$$

It maps points of the manifold into tangent spaces. Now, by substituting Equation (7) into (5) we get the following equation for Pennec's metric:

$$\begin{aligned} &< y, z >_{\sum_1} \\ &= \quad trace\left(\log^2\left(\sum_1^{-\frac{1}{2}} \sum_2 \sum_1^{-\frac{1}{2}}\right)\right) \quad . \\ &= \quad < log_{\sum_1}\left(\sum_2\right), log_{\sum_1}\left(\sum_2\right) >_{\sum_1} \end{aligned} \tag{8}$$

The exp and log are the ordinary matrix exponential and logarithm operators.

C. Empirical Mean Value

There are several definitions of the (empirical) mean value for a set of measures of the same positive definite symmetric matrix [15]. One applicable mean is the so-called Karcher or Fréchet mean which minimizes the sum of the squared distances between the matrices. According to [15], [16] and [17] this mean exists and is even unique in Sym_n^+, as this manifold has a non-positive curvature [18]. It can be computed by a gradient descent algorithm [15]: matrices are mapped into the tangent space first, where then the Euclidean mean is calculated. Eventually, the mean value of the covariance matrices is given by mapping back the Euclidean mean.

Let $\Sigma_1 \dots \Sigma_n$ be a set of n measures of the positive definite symmetric matrix $\bar{\Sigma}_t$, then the new mean $\bar{\Sigma}_{t+1}$ of this set is given by

$$\bar{\Sigma}_{t+1} = exp_{\bar{\Sigma}_t}\left(\frac{1}{n} \sum_{i=1}^{n} log_{\bar{\Sigma}_t}(\Sigma_i)\right). \tag{9}$$

An important point using this computation is to determine a good starting point. If there is no $\bar{\sum}_t$, in general, an

element of $\sum_1 \dots \sum_n$ can be selected randomly as starting point. If required, the matrices may be weighted differently, for instance, by their distances to the mean value [8].

III. Spectral Clustering

This section gives a brief overview of the proposed spectral clustering approach used in this contribution (details about well-known spectral clustering algorithms can be found e.g. in [19], [20]). Let $\Sigma_1 \dots \Sigma_n$ be a set of n covariance descriptors. First, a symmetric adjacency matrix $A = (a_{ij})$ is computed:

$$a_{ij} = a_{ji} = \begin{cases} 1 & , \quad if\ \Sigma_j \in N_{\Sigma_i} \\ 0 & , \quad otherwise \end{cases}, \tag{10}$$

using the Riemannian Metric $d(\Sigma_i, \Sigma_j)$ as stated in Equation (5) respectively Equation (8) and $\Sigma_j \in N_{\Sigma_i}$ means that Σ_j is a neighbor of Σ_i. There are several methods to determine the neighborhood relationships of data points respectively descriptors. For instance, all descriptors can be connected with each other or the k-nearest neighbors of a descriptor can be determined. In addition, the connections can be weighted by the corresponding pairwise similarities or distances of the descriptors regardless of the chosen method. In our approach, we consider a ϵ-neighborhood, where all descriptors whose pairwise distances are smaller than ϵ are connected without weighting the connections. In our experiments we got the best results with this neighborhood, however, please note that in general - as mentioned in [19] - theoretical results on the question how the choice of the similarity graph influences the spectral clustering result are not known.

Then the normalized graph Laplacian matrix L_{sym} is computed. It is defined as

$$L_{sym} = I - D^{-\frac{1}{2}} A D^{-\frac{1}{2}} \tag{11}$$

with the identity matrix I. $D = (d_{ij})$ denotes a diagonal matrix whose diagonal entry d_{ii} is given by the sum of the i^{th} row of A.

Properties of the $n \times n$ normalized graph Laplacian matrix L_{sym}, which are relevant within this paper, are listed hereafter (the proof can be found in [19]):

- 0 is an eigenvalue of L_{sym} with eigenvector $D^{\frac{1}{2}} \overrightarrow{\mathbf{1}}$.
- L_{sym} is positive semi-definite and have n non-negative real valued eigenvalues $0 = \lambda_1 \leq \dots \leq \lambda_n$.
- Due to the graph W is undirected and only have non-negative weights, the multiplicity k of the eigenvalue 0 of L_{sym} equals the number of connected components in the graph.
- W.l.o.g., assuming that the elements of each connected component are ordered according to the component they belong to, then both the adjacency matrix A as well as the Laplacian matrix L_{sym} have a block diagonal form:

$$L_{sym} = \begin{pmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{pmatrix}. \tag{12}$$

- The spectrum of L_{sym} is given by the union of the spectra of L_i, and the corresponding eigenvectors of L_{sym} are the eigenvectors of L_i, filled with 0 at the positions of the other blocks.

After computation of the normalized Laplacian matrix, the clustering can be performed. Let k the number of the wanted clusters. In that case, the first k eigenvectors $u_1, \dots, u_k$ of L_{sym} are computed and the matrix $T \in \mathbb{R}^{n \times k}$ is formed by the k eigenvectors as columns of T. Additionally, the rows of T are normalized to 1. As a final step, the rows of T, which represent the input covariance descriptors, are grouped into k clusters (details about the determination of k in our proposed approach is described in the next section). For this, the k-means algorithm is typically applied on the rows of T. In the context of our proposed approach we run the k-means algorithm only on blocks of the Laplacian matrix L_{sym} with block size equal or greater than x (x corresponds to the minimum number of covariance descriptors wanted for the computation of means). Blocks with a size smaller than x are not considered, so that we got $j \leq k$ clusters - at least one mean of covariance descriptors for every block with the size equal or greater than x.

IV. Re-Identification

In our context, the objective of human re-identification is to recognize a specific person captured by a tracker in a multi-camera network respectively to merge partial trajectories, for instance, caused through gaps in the cameras' field of views without having any information concerning sensor topology or the like. In other words, we have an image sequence of a connected specific individual's track and compare this with sequences of other connected tracks in order to determine whether the (partial) trajectories belong to the same individual. Therefore, we present in this section an efficient whole-body appearance-based human re-identification method for low-resolution videos.

The input of our proposed approach are image regions (rectangles) generated from our tracking method. An image region contains exactly one person - image regions containing multiple persons are detected automatically by the tracker and are not considered in the re-identification. Additionally, the persons are segmented in the images by the tracker and the background are blackened (details about the tracker can be found in [21]). The last step before performing the re-identification is the scaling of the images to a fix width and height.

The re-identification procedure is performed by comparing means of covariance descriptors that represent the

whole-body appearance (for every track at least one mean is computed). To this, a set of covariance descriptors $\mathbf{\Sigma} = \{\Sigma_1, \ldots, \Sigma_n\}$ - one descriptor for each image region of a sequence - is computed as described in Section II. Then k empirical mean values $\bar{\Sigma}_i$, $i = 1 \ldots k$ are computed from these descriptors for a sequence as described in Section II-C in accordance with Equation 9.

The number of means k is determined by an eigengap heuristic, where the common goal is to choose the number k such that all eigenvalues $\lambda_1 \ldots \lambda_k$ are very small, but λ_{k+1} is relatively large. As in Section III described, in the ideal case of k completely disconnected clusters, the eigenvalue 0 has multiplicity k, and then there is a gap to the $(k+1)th$ eigenvalue $\lambda_{k+1} > 0$. If there are no ideal disconnected clusters, we compare the eigenvalue gap to a bias that we empirically determined from our data set.

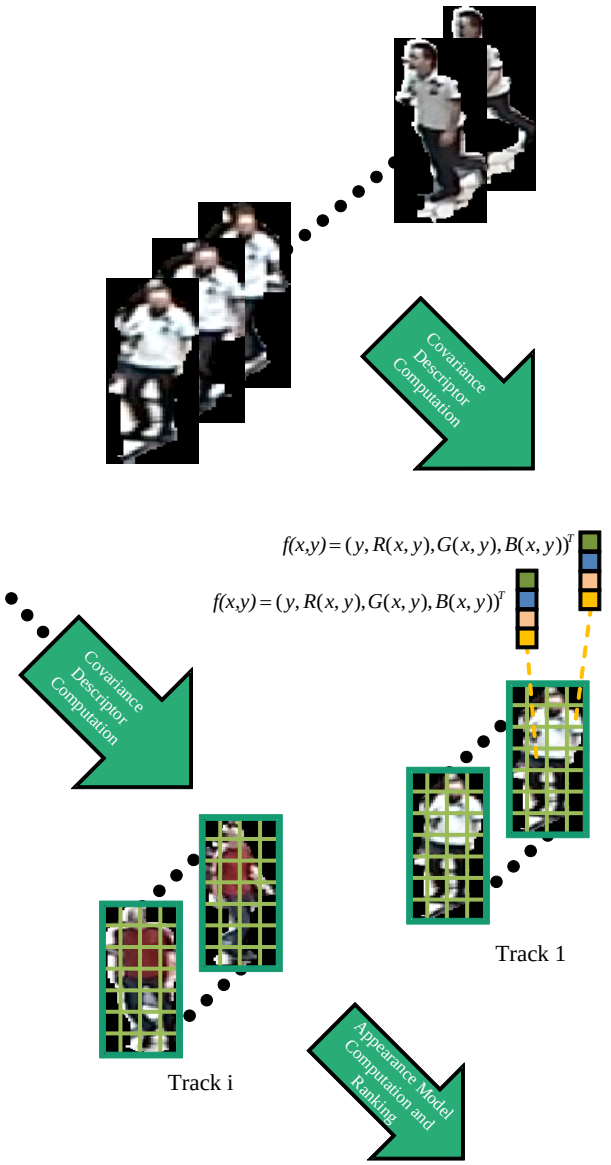
The result are k means of covariance descriptors for a sequence. This procedure is then done for every sequence. After that, the means representing different sequences are compared pairwise using the geodesic metric as defined by Equation (8). The final decision of "who is who" results directly from these comparisons. In the context of this paper the objective is to get a ranking by the distances of the means of covariance descriptors. Figure 1 gives an overview of the proposed approach.

V. Experimental Results

In this section, the experimental results of our proposed approach is summarized. The camera network is installed in an atrium at a height of 3 meters and consists of 3 IP cameras with a resolution of 4CIF. For the experiments a test data set consisting of 96 connected tracks were considered (16, 25 respectively 55 from each one camera). Each track consists of minimum 10 and up to 1000 images. The image regions containing the individuals are scaled to a width of 64 pixels and a height of 128 pixels. Figure 2 shows some examples of the data set. The low resolution of the persons is the biggest challenge in this data set, so that especially similar appearances - as exemplarily pointed out in the first two rows of the figure - make the re-identification difficult.

Covariance descriptors were computed from the image coordinates y and the R, G, B color values. The x-coordinates were excluded from the feature set as they increase the invariance in cases where individuals are seen from different sides.

The re-identification experiments were performed in three cameras, whereby their different fields of view are non-overlapping. Most individuals passed all cameras as well as re-entered one or more same cameras' field of view after a longer period of time. We evaluated our proposed approach on this data set and compared it to a histogram based multi-shot re-identification procedure, whereby - in both procedures - pre-processing methods such as histogram equalization were deliberately not applied.



Rank	1	2	3
Mean 1	Mean i	Mean n	Mean j
⋮	…	…	…
Mean n	…	…	…

Figure 1. Overview of the proposed approach.

The histogram based method uses RGB histograms containing 64 bins for each color channel. For each image of a connected reference track a histogram was computed and the histogram with the smallest chi-square distance of the gallery tracks was determined. The ranking was then directly concluded from the number of track hits (normed to the track length).

In our proposed approach, all means of covariance de-

Table I
RANKING RESULTS OF THE RE-IDENTIFICATION EXPERIMENTS. THE TABLE COLUMNS 2 (HISTOGRAM BASED APPROACH) AND 3 (OUR APPROACH) SPECIFY FOR EVERY RANK THE PERCENTAGE OF THE CORRECT CORRESPONDING TRACKS RESPECTIVELY RE-IDENTIFICATIONS. RANKINGS HIGHER THAN RANK FIVE ARE NOT CONSIDERED.

Rank	Histogram approach	Our approach
1	38.94%	64.56%
2	47.34%	71.48%
3	55.06%	76.73%
4	62.61%	80.13%
5	64.32%	83.85%

scriptors were computed first. Then for each mean of a reference track the distances to means of the gallery tracks were determined and ranked by their distance with the result that the closest one was put on the first rank. At it, for every gallery track a seperate test run was performed, where only one track of the wanted person was kept in the gallery. This procedure avoids random hits in cases there are several tracks or means of one person in the gallery. The same procedure was done if there were several means of covariance descriptors for one track. The results are summarized in Table I that specifies for every rank the percentage of the correct re-identifications.

VI. CONCLUSION

We have proposed an appearance-based human re-identification method for low-resolution videos using means of covariance descriptors and spectral clustering techniques. The main contribution is an appearance model based on means of covariance descriptors which significantly decreases the number of comparisons necessary for re-identification. The approach was evaluated on a data set of a representative surveillance scenario and it was shown that it has the capability to outperform color histogram approaches.

ACKNOWLEDGMENT

The research reported in this contribution is partly funded by the German Federal Ministry of Education and Research (BMBF) within the program "Research for Civil Security", in conjunction with the project *CamInSens* (www.caminsens.org).

Figure 2. Sample images from our data set.

REFERENCES

[1] M. Baeuml, K. Bernardin, M. Fischer, H. K. Ekenel, and R. Stiefelhagen, "Multi-pose face recognition for person retrieval in camera networks," *Proc. of International Conference on Advanced Video and Signal-Based Surveillance*, 2010.

[2] M. S. Nixon, T. Tan, and R. Chellappa, "Human identification based on gait," *Proc. of International Series on Biometrics*, vol. vol. 4, 2006.

[3] R. Chellappa, A. K. Roy-Chowdhury, and A. A. Kale, "Human identification using gait and face," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, 2007.

[4] A. Dantcheva, C. Velardo, A. D'Angelo, and J.-L. Dugelay, "Bag of soft biometrics for person identification," *Journal of Multimedia Tools and Applications*, vol. vol. 51-2, 2011.

[5] A. K. Jain and A. Kumar, "Biometrics of next generation : An overview," *Second Generation Biometrics, Springer*, 2010.

[6] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," *Journal of Ambient Intelligence and Humanized Computing*, vol. vol. 2, pp. 127–151, 2011.

[7] O. Tuzel, F. Porikli, and P. Meer, "Region covariance: A fast descriptor for detection and classification," *Proc. of European Conference on Computer Vision ECCV*, vol. vol. 2, pp. 589–600, 2006.

[8] F. Porikli, O. Tuzel, and P. Meer, "Covariance tracking using model update based on means on riemannian manifolds," *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, 2005.

[9] ——, "Human detection via classification on riemannian manifolds," *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition CVPR*, 2007.

[10] S. Bak, E. Corvee, F. Bremond, and M. Thonnat, "Boosted human re-identification using riemannian manifolds," *Journal of Image and Vision Computing, Elsevier*, 2011.

[11] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof, "Person re-identification by descriptive and discriminative classification," *Scandinavian Conference on Image Analysis, Lecture Notes in Computer Science*, vol. vol. 6688, pp. 91–102, 2011.

[12] F. Porikli and O. Tuzel, "Fast construction of covariance matrices for arbitrary size image windows," *Proc. of Intl. Conf. on Image Processing ICIP*, 2006.

[13] W. Foerstner and B. Moonen, "A metric for covariance matrices," *Technical report, Department of Geodesy and Geoinformatics, University of Stuttgart*, 1998.

[14] X. Pennec, P. Fillard, and N. Ayache, "A riemannian framework for tensor computing," *International Journal of Computer Vision IJCV*, vol. vol. 64, pp. 41–66, 2006.

[15] X. Pennec, "Intrinsic statistics on riemannian manifolds: Basic tools for geometric measurements," *Journal of Mathematical Imaging and Vision JMIV*, vol. vol. 25, pp. 127–154, 2006.

[16] H. Karcher, "Riemannian center of mass and mollifier smoothing," *Comm. Pure Appl. Math.*, vol. vol. 30, pp. 509–541, 1977.

[17] W. S. Kendall, "Probability, convexity, and harmonic maps with small image i: uniqueness and fine existence," *Proc. London Math. Soc.*, vol. vol. 61-2, pp. 371–406, 1990.

[18] L. Skovgaard, "A riemannian geometry of the multivariate normal model," *Scandinavian Journal of Statistics*, vol. vol. 11, pp. 211–223, 1984.

[19] U. von Luxburg, "A tutorial on spectral clustering," *Technical Report, Max Planck Institute for Biological Cybernetics*, vol. vol. TR-149, 2007.

[20] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: analysis and an algorithm," *Advances in Neural Information Processing Systems, MIT Press*, vol. vol. 14, pp. 849–856, 2002.

[21] C. Herrmann, D. Manger, and J. Metzler, "Feature-based localization refinement of players in soccer using plausibility maps," *Proc. of International Conference on Image Processing, Computer Vision, and Pattern Recognition IPCV*, vol. vol. 2, 2011.