Re-identification of persons in multicamera surveillance under varying viewpoints and illumination

Bouma, H.; Borsboom, S.; den Hollander, R.J.M.; Worring, M.

DOI
10.1117/12.918576

Publication date
2012

Document Version
Author accepted manuscript

Published in
Proceedings of SPIE, the International Society for Optical Engineering

Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination

Henri Bouma [*a], Sander Borsboom [a,c], Richard J.M. den Hollander [a],
Sander H. Landsmeer [a,b], Marcel Worring [c]

[a] TNO, P.O. Box 96864, 2509 JG The Hague, The Netherlands;
[b] Science and Technology, P.O. Box 608, 2600 AP Delft, The Netherlands;
[c] University of Amsterdam, P.O. Box 94323, 1098 GH Amsterdam, The Netherlands.

ABSTRACT

The capability to track individuals in CCTV cameras is important for surveillance and forensics alike. However, it is laborious to do over multiple cameras. Therefore, an automated system is desirable. In literature several methods have been proposed, but their robustness against varying viewpoints and illumination is limited. Hence performance in realistic settings is also limited. In this paper, we present a novel method for the automatic re-identification of persons in video from surveillance cameras in a realistic setting. The method is computationally efficient, robust to a wide variety of viewpoints and illumination, simple to implement and it requires no training. We compare the performance of our method to several state-of-the-art methods on a publically available dataset that contains the variety of viewpoints and illumination to allow benchmarking. The results indicate that our method shows good performance and enables a human operator to track persons five times faster.

Keywords: Security, surveillance systems, forensics, person re-identification, person matching, tracking, tracing, image retrieval.

1. INTRODUCTION

The capability to track and trace individuals in CCTV cameras is important for surveillance and forensics. However, it is laborious for a camera operator to do this over multiple cameras. Therefore, an automated system is desirable that can assist the operator in his search for specific persons and their whereabouts. For automatic person tracking over multiple cameras without overlapping views, the main component is a person re-identification algorithm. The task of this algorithm is to find the person images in a large collection that are most similar to a query person image. Images of the same person are ranked as high as possible, preferably first.

In this paper, we present our computationally efficient person re-identification method – which is mainly based on multi-dimensional histograms containing color and spatial information. We compare the performance of our method to several state-of-the-art methods. To evaluate the performance of these algorithms the VIPeR dataset (from UC Santa Cruz [19]) is used; this is a publically available benchmark in a realistic setting for viewpoint-invariant person re-identification algorithms. This dataset consists of two different recordings for 632 individuals and each recording is a still image that tightly encloses the person (bounding box). The data contains a wide variety of view-points, poses, backgrounds and lighting conditions, as is typically seen in surveillance systems in an outdoor situation.

The outline of this paper is as follows. Section 2 gives an extensive literature overview, section 3 describes our method, section 4 shows the experiments and results, and finally, section 5 presents the conclusions.

* henri.bouma@tno.nl; phone +31 888 66 4054; http://www.tno.nl

Henri Bouma, Sander Borsboom, Richard den Hollander, Sander Landsmeer, Marcel Worring, "Re-identification of persons in multi-camera surveillance under varying viewpoints and illumination", Proc. SPIE, Vol. 8359, 83590Q (2012). http://dx.doi.org/10.1117/12.918576

2. LITERATURE OVERVIEW

A person-tracing system is composed of multiple components such as moving object detection [13] or static person detection [9][15][26], segmentation, tracking [25][43][45][46], human interaction [6][8] and person re-identification. The scope of our investigation focuses on the person re-identification algorithm which uses still images of the persons' bounding boxes as input and performs two steps: descriptor computation and matching (Figure 1). So, detection and tracking are outside the scope of this paper.

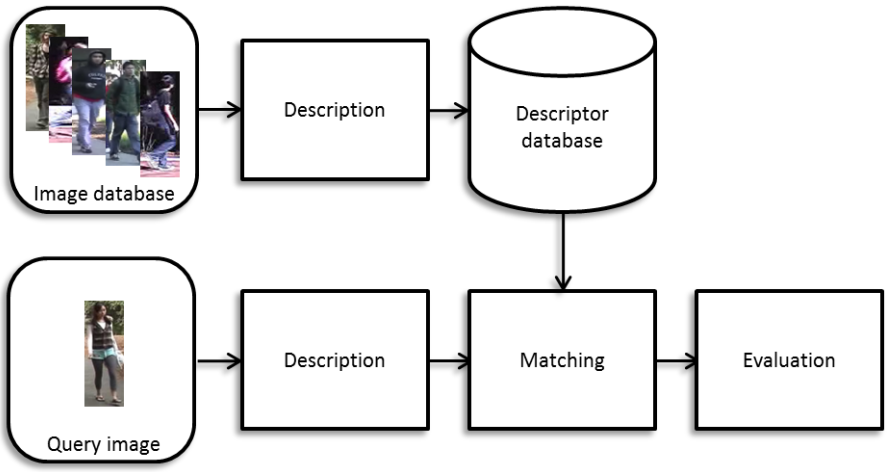

Figure 1: A system for person re-identification includes the descriptor computation and matching.

Many variations of person re-identification algorithms exist and the goal of this section is to give a short overview of these approaches. Figure 1 shows an overview of a generic person re-identification system, which includes the computation of descriptors and the matching of descriptors from different persons to obtain a similarity score. We will investigate the main approaches for descriptor computation and matching seen in literature, starting with color systems (Sec. 2.1), followed by color histograms (Sec. 2.2), Gaussian color models (Sec. 2.3), spatial information (Sec. 2.4), structure and texture (Sec. 2.5), and finish with a discussion (Sec. 2.6).

2.1 Color system

The descriptors that have been proposed in literature use a transformation to a color system that allows separability between persons and sufficient invariance to lighting variation. Many color systems have been defined, each capturing color information in a different way. One of the most commonly used color systems is RGB, which has been tested in many systems [18][19][32][33][42], but in all cases performed comparable or worse than other color systems, because it is not robust to changes in lighting. There are multiple methods to separate brightness information from color information, resulting in, among others, the normalized RGB, Opponent, L*a*b* and HSV color systems. With normalized RGB, the two color channels are calculated by normalizing the R and G channels with the sum of the values of all channels combined. The L*a*b* and Opponent color systems are based on color opponent theory where the first is based on the CIE XYZ and the latter on RGB. The HSV color system further separates color into separate Hue and Saturation channels, but the disadvantage is that Hue becomes unstable at low saturation values. Gray and Tao [18] report that their self-learning approach favors the Hue channel information above the RGB and YCbCr channels, but Wang et al. [42] show lower HSV performance in comparison to RGB and L*a*b. Metternich et al. [32] compared HSV to RGB, normalized RGB and the opponent color system but got mixed results, making it impossible to say which color system performed best. Ranked RGB preserves the color distribution with a transformation from values to percentiles. Lin and Davis [29] and Yu et al. [44] compared the ranked-RGB color system to RGB and normalized RGB and reported significant improvements in performance. In the person-matching literature, only a few color systems are compared at a time, making it hard to infer the best color system for this task. In the object-classification task, the analysis of multiple color systems by van de Sande et al. [37] shows an overview of a larger set of color systems. The benchmark showed RGB, Opponent and Transformed RGB performing best while the Hue and normalized RG color channels performed worst. Although this information cannot be directly used for the person-matching task, the results show that lighting

invariance is important but also that the use of some brightness information is useful in object recognition. Which of the color systems is best for the person-matching task is still hard to say.

2.2 Color histograms

Histograms are a method to capture the distribution of descriptors in an image efficiently. They are simple and fast to calculate and show good matching performance. In many cases, single fixed quantization histograms are used as the benchmark [17][18][42]. The basic histogram contains color information, but can be extended with structural information such as the dominant orientation of edge pixels and the ratio of the color values on each side of the edge [17]. One of the methods used to improve the performance is giving different channels different numbers of bins, e.g. in HSV-space allowing more invariance to intensity (V), while keeping the same level of precision for the hue (H) [42]. Another method proposed to improve performance of histograms is to use variable quantization where the quantization is chosen in such a way as to better follow the average distribution. This can for example be done by recursively splitting at the median intensity [22] or by clustering [42]. In their experiments, Wang [42] varies both the quantization and the local descriptor, which makes it hard to assess the added value of variable quantization without further investigation.

2.3 Gaussian color model

There are several methods that use a Gaussian color model or color covariance matrices. Mensink et al. [31] propose using expectation maximization and a mixture of Gaussians, where each person is represented with one Gaussian. Ma et al. [30] propose computing the centered auto-covariance matrix of all pixels in a person [33]. The distances between covariance matrices are computed as the sum of squared logarithms of the generalized eigenvalues of these matrices [30] or the earth movers distance [27]. Similar to the distance based on covariance matrices is multivariate Kernel Density Estimation (KDE), which tries to capture the underlying distribution in detail by defining the probability of a new point based on its similarity to previously seen points. This in contrast to histograms, where the point is given the probability of a discrete pre-defined area, the bin, closest to it. Yu et al. [44] and Lin and Davis [29] propose using KDE for the person-matching task. The reported results do show large improvements, making it an interesting choice for the person-matching task. Another way to describe an image is to see how it differs from a set of prototypical images. Lin and Davis [29] use the pairwise dissimilarity profiles together with a KDE appearance model and reported a performance increase when both are combined, which shows that combinations of algorithms are another avenue for experimentation. Their dataset shows very small differences in pose and viewpoints. Because of the high precision of their dissimilarity descriptions, it is likely that they are not robust to large changes in viewpoint. Cai [7] also proposed a method inspired by self-similarity. Instead of comparing image descriptors between two images directly, the self-similarity measures how similar they are to neighboring descriptors. The self-similarities of image patterns within the image are modeled in the proposed global color context. The spatial distributions of self-similarities w.r.t. color words are combined to characterize the appearance of pedestrians.

2.4 Spatial information

Spatial information can be used in various ways. The shape context [4] is a location-aware histogram. For a pixel it captures the relative distribution, the context, of a set of other pixels around it by using the pixel as the center of a two-dimensional histogram of the surrounding pixels. Wang et al. [42] used this for person matching by labeling all pixels in the image based on color and gradient information and then modeling the relative distribution of these labels. To further improve this approach, shape labeling is used to model the distribution of appearance information relative to specific parts of the body [12][42]. The performance of this approach is better than basic histogram-based approaches and small variations in pose and viewpoint are handled well. Further investigations are necessary to validate this performance under larger changes in viewpoint and pose.

Another spatially aware histogram is proposed by Gheissari et al. [17] which matches a polygonal outline of a person with a set of images from video. For each area in the polygonal outline a histogram is then calculated over all images. Performance of the approach was reported to be significantly better than a basic histogram based method. The use of temporal information to remove the influence of the background by a better outline is likely to be a major cause of these improvements. This makes it less suitable for the single-image matching. The background in unsegmented person images can have a significant impact on the descriptors of the image, resulting in reduced matching performance. Instead of the polygon based on movement information [17], or a simple Boolean mask, it is also possible to weigh the influence of pixels [34][32], but no changes in performance were reported when weighing was used [32]. These results show that Boolean masking is a good choice for the person-matching task.

Alahi [1] proposed a descriptor which uses a collection of grids. Each grid segments the object into a different number of equally sized sub-rectangles. By combining fine regions with coarse regions, both local and global information is preserved. They show that a sparse selection combined with a cascade of covariances performs best closely followed by a collection of HOGs (histogram of oriented gradients). Farenzena et al. [14] partitioned the person in the different body parts with a vertical separation based on asymmetry and a horizontal center based on symmetry. They used three complementary aspects to encode the visual appearance: the overall chromatic content with a HSV histogram, the presence of recurrent local structures with high entropy, and maximally stable color regions.

A simple way to incorporate the spatial information is by measuring the location in comparison to the height and/or width of the person. Some algorithms use the full *x,y* coordinates of pixels [27][30], but it is likely that this is not robust to pose and viewpoint changes. Because of this, many others [22][29] use only the y (height) value or a coarse separation in two parts (upper and lower body) where the dominant color is computed [2][3]. Dikmen [11] used color histograms extracted from small rectangular overlapping regions to represent the images. For the color histograms RGB and HSV color spaces are used and 8-bin histograms are extracted for each channel separately and concatenated to form a feature vector. PCA reduces this vector to 60 dimensions and a large-margin nearest neighbor with rejection (LMNN-R) classifier is used to re-identify persons. The results seem promising. However they did not show the commonly used average results after multiple iterations, but only the maximal score. Another way to define locations is to measure the distance from the top of the head of the person [44]. This should improve robustness because this location is more stable than the overall length, as this will change because of movement of the legs. The results show that the descriptor performs well, but is not robust to segmentation errors. The shape of specific areas can also be used to recognize specific locations in the images even under pose and viewpoint changes [42].

2.5 Structure and texture

Interest points [17][20][23][41] are located in areas with high information content, mostly located on edges in the image. Gheissari et al. [17] create a set of interest points with a descriptor based on a color and texture histogram over a small neighborhood. The number of matching points can be used to compare images, where a match is defined if the histogram intersection is above a given threshold [17], or the sum of absolute differences [20]. Gheissari et al. compare their interest-point approach to another approach where the body parts of a person are separately described and they reported that the interest points performs worse. The low performance of interest points is probably related to the dependency of textures, which are not invariant to changes in viewpoint.

Because the silhouette and the shape of clothing will change significantly through changes in viewpoint, pose and lighting direction, texture information may result in a large noise burden. Still, clothing contains textural features that can be used to recognize a person. Ma et al. [30] and Le et al. [27] proposed to use the gradient of a pixel as a descriptor but they do not compare its use versus not using it. Wang et al. [42] tested gradient-based linear filter and reported performance comparable to their best performing color system, L*a*b*. Since these filter banks will likely lose some of their color information, these results seem to indicate that the images contain enough textural information to make up for this loss. Stronger indications of usefulness are given by Gray and Tao [18][19] who used filters [16][40] as descriptors in their self-learning matching algorithm. Results show that these descriptors are selected frequently by this algorithm, indicating the use of texture information is beneficial.

Some approaches describe a pixel based on a small region around it. One method uses histograms of edgels [17]. These edgels are based on a spatio-temporal algorithm that finds the edges between major areas of colors that exist over multiple frames, which removes most edges caused by changes in pose, viewpoint and lighting, resulting in more robust descriptors. This approach cannot be applied to single stills. More information can be taken by using SURF descriptors in a dense grid [24], Histogram of Oriented Gradients (HOG) [9][12][42] or Haar features [2]. By calculating HOGs for every color channel and then normalizing these over the channels, color information can be captured. The HOG was used with the Log-RGB color system and reported significant improvements in performance in comparison to their baseline.

2.6 Discussion

To investigate and directly compare state-of-the-art approaches seen in the literature, we will investigate the methods proposed by Wang et al. [42] and Lin and Davis [29] and compare them with our proposed methods. Both obtain very promising results on their own data (recognition rates of 82% and 89% at rank 1) and both already implement a large range of approaches for the parts of a person-matching algorithm and allow simple integration of new approaches. Wang et al. [42] proposed an algorithm including histograms with fixed and variable quantization, the use of texture

information and spatially aware histograms. Lin and Davis [29] use a Kernel Density Estimation (KDE) based approach with dataset-based descriptors and multiple algorithm combination techniques.

A shortcoming of previous work is that many methods in literature use proprietary test sets, making it impossible to compare algorithms from different papers. A further problem with these datasets is that they are very small, containing on average only 50 people, and that most have been created in a lab setting. This results in datasets missing influences which are seen in reality, such as large changes in viewpoints, poses and lighting. To resolve this problem, we will use the public Viewpoint Invariant Pedestrian Recognition (VIPeR) dataset described in Gray et al. [18]. This is a dataset of sufficient size and realism, allowing us to further investigate performance of the different approaches.

The lack of direct comparisons makes it hard to determine the best overall methods or to determine the best combination of parts to create such a method. For example, many papers report large differences in performance between two or three color systems, but very few come close to testing a comprehensive set of color systems, making it hard to select the best color system. Multiple algorithms are proposed that use texture information, but since full algorithms are tested, in every case the performance improvements seen with these algorithms are just as likely to be caused by other parts of the algorithm. To gain an insight in how many of these parts work together, we will apply the approaches proposed by Wang et al. [42] and Lin and Davis [29] on the VIPeR data set and compare them to our proposed person re-identification algorithm.

3. METHOD

3.1 System overview

This section describes two novel histogram-based methods for the re-identification of persons in surveillance video: The multi color-height histogram (MCHH) method and the transformed-normalized color-height histogram (transformed-normalized) method.

3.2 The MCHH method

The multi color-height histograms (MCHH) method consists of two steps: histogram computation and matching. The histogram-computation step of the method consists of building two four-dimensional color-height histograms. The first histogram uses three RGB-Rank [29] color components and height. RGB-rank replaces every color value by the percentage of values in the image lower than the value itself. The rank is calculated for all three R, G and B channels separately. An advantage of this descriptor is its invariance against lighting variations. The other histogram is also four-dimensional and it uses three Opponent color [37] components and height. The Opponent colors are determined from RGB with Equation (1).

$$\begin{pmatrix} O_1 \\ O_2 \\ O_3 \end{pmatrix} = \begin{pmatrix} (R - G)/\sqrt{2} \\ (R + G - 2B)/\sqrt{6} \\ (R + G + B)/\sqrt{3} \end{pmatrix} \quad (1)$$

The Opponent color is a rotation of the RGB color space so that the intensity channel *(O_3)* is separated from the other color information. The benefit of Opponent above HSV is that it does not become unstable in the gray region. A histogram intersection is used to compute the similarity between a histogram of the query image and that of another image. This is performed for both the RGB-Rank and the Opponent histogram. The two histogram-intersection scores are combined to form a similarity measure by summing the squared scores.

3.3 The transformed-normalized method

The transformed-normalized histogram method is similar to the MCHH method. It also consists of two steps: feature computation and matching. The feature-computation step consists of one multi-dimensional color-height histogram. The histogram is four-dimensional and it uses three transformed-normalized RGB color components and height. We use the following definitions for transformed and normalized colors [37]. In a transformed color system, the pixel-values of each color channel are normalized independently by subtracting the average and scaling them with the standard deviation, e.g. for the R-channel $R' = (R - \mu_R) / \sigma_R$. In a normalized color system, the pixel-values of a color channel are normalized

relative to the other color channels by dividing them by the sum of the channels, using: r = R / (R+G+B). The similarity measure between the histogram of the query image and another image is computed with histogram intersection.

4. EXPERIMENTS AND RESULTS

This section describes the data (Sec. III-A), the parameters of methods (Sec. III-B), the experiments (III-C) and the results (III-D).

4.1 Data

The VIPeR dataset [19] is a publically available benchmark for viewpoint-invariant person re-identification algorithms. This dataset consists of two different recordings for 632 individuals (see Fig. 2). The data contains a wide variety of view-points and lighting.

Figure 2: Examples of images from the VIPeR data.

4.2 Parameters of the methods

We computed the performance of earlier mentioned state-of-the-art methods and the proposed methods: pairwise dissimilarity profiles [29], shape and appearance context [42], multi color-height histograms and transposed-normalized histograms. The parameters of the methods from Wang and Lin that were proposed in the original paper, were slightly modified to obtain a fair comparison on the VIPeR data. This section shows the parameters that were used to obtain these results. For a full and detailed description of the methods we refer to the original papers.

The pairwise dissimilarity-profiles method [29] describes each pixel as a 4D feature vector based on 3D color and 1D height. By using multi-variate kernel density estimation (KDE) with a Gaussian kernel, the likelihood of a pixel belonging to an image is modeled. The system consists of two subsystems. The first subsystem directly compares the query and database image using the Kullback-Leibler distance. The second subsystem computes a log-likelihood ratio between the query image and each database image. The ratio is then projected on the vertical axis to create a one-dimensional profile. So, for a database with N images, the profile set that describes a query image consists of N profiles. The distance metric uses voting and it compares the profile set of the query image to the profile set of each database image (where the database consists of N x N profiles) to find the best match. Finally, the scores of the two subsystems are combined in one of several ways (e.g. linear/non-linear). We applied this method on the VIPeR data with the following parameters [5]. It uses the normalized RGB color system, a KDE bandwidth of 2% of the range for the each of the four channels (r, g, brightness and height), the non-linear combination (combine2) with a combination weight of β=0.20.

The shape and appearance context method [12][42] uses local color and texture descriptors. The image descriptors are labeled based on fixed quantization boundaries or learned labels and the labeled images are used to create a global descriptor for the image. The appearance labeling uses k-means clustering and the L1-norm on a specific color system. The shape labeling uses k-means clustering on the combination of HOGs and the average of HOGs in special spatial partitions. In the shape-and-appearance context, shape labels are used together with appearance labels instead of

appearance labels alone. We applied this method on the VIPeR dataset with the following parameters [5]. It uses the shape-and-appearance context. For the appearance it uses the log-RGB color system, 2 orientation bins, a cells size of 5, 70 appearance labels, and vector normalization of the signed weighted HOG. For the shape it uses 8 orientation bins, a cell size of 11, 18 shape labels, and no normalization of the signed weighted HOG.

The MCHH method has the following parameters. In the RankRGB-height histogram the number of histogram bins is (6x6x6)x6 and in the Opponent-height histogram the number of bins is (4x4x4)x8.

The transformed-normalized method has the following parameters. The transformed-normalized RGB color histogram has (7x7x7)x6 bins for the colors and height.

4.3 Experiment

The methods are applied to the VIPeR data. Experiments were performed with ten iterations of 2-fold cross validation; so the performance is estimated with 316 randomly selected queries from one camera in a 316 database from the other camera for 10 times. This was done to estimate the average performance and to allow comparison with available results [14][18][19].

All persons are represented by a bounding box of fixed size. We used a single fixed segmentation mask to separate the foreground from the background for all images and all methods to improve matching scores. This segmentation mask is based on the average location of the head, upper body and upper legs. The lower legs were excluded from the mask because their location varied too much. A test showed that using multiple masks for different poses (based on the principal components, e.g. for frontal and side views) does not result in a performance gain.

4.4 Results

The performance was determined with a Cumulative matching characteristic (CMC). The CMC curves [19] are shown in Figure 3 and the main CMC-results are summarized in Table 1. The figure shows the shape and appearance context [42], KDE and pairwise dissimilarities [29], the novel methods MCHH and Transformed. In the table, we also added the results of the papers of Gray [18], Metternich [32], Farenzena [14] and Dikmen [11] (based on figures in their papers), which were obtained on the same data.

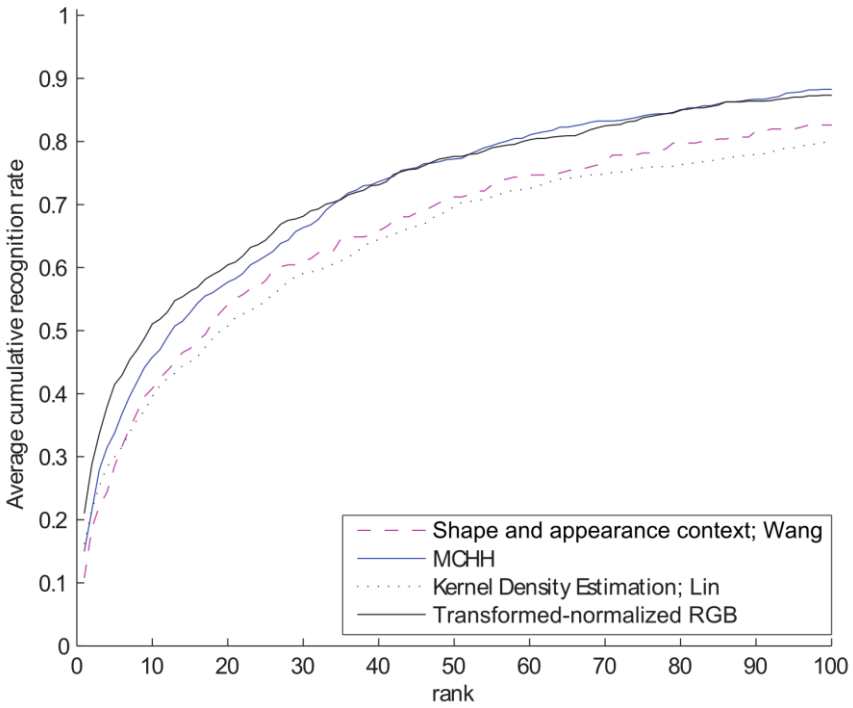

Figure 3: CMC performance curves on the VIPeR dataset of the different methods.

Table 1. CMC performance values on the VIPeR dataset of different methods. (* Most papers retrieved queries in a subset of 316 images, but [32] used only 50 images. ** Most papers averaged the CMC curves of multiple randomly chosen subsets, but [11] only reports the CMC of a single split with the best result.)

Algorithm	rate at rank 10	av. rate of rank 1 to 100
KDE with pairwise dissimilarity profiles [29]	39%	63.8
Shape and appearance context [42]	41%	65.8
Covariance matrix [32]	48% *	-
Ensemble of localized features [18]	46%	-
Symmetry-driven accumulation of local features [14] / MCM [38]	50%	-
Large Margin Nearest Neighbor with Rejection [11]	65% **	-
Single color-height histogram	42%	63.8
Multiple color-height histograms	48%	71.6
Transformed-normalized color-height histogram	51%	72.5

The results show that our methods performs very well. Let us make a few remarks to put the numbers in Table 1 into perspective. First, it should be noted that the computation time of the pairwise dissimilarity profiles [29] is extremely long, and since the results are not extremely good it is not recommended for further use. Second, most papers retrieved queries in a subset of 316 images, but [32] used only 50 images. Achieving a matching rate of 48% at rank 10 in a database of 50 is much easier than in a database of 316. Third, most papers averaged the CMC curves of multiple randomly chosen subsets, but [11] only reports the CMC of a single split with the best result. This will give an overestimation of the performance, since VIPeR contained easy cases and hard cases (e.g., over saturated).
Our average matching rate over rank 1 to 316 is 90%, which results in a 5 times faster search time than a purely manual search in an unordered database (with an average matching rate of 50%).

5. CONCLUSIONS

The capability to track and trace individuals in CCTV cameras is important for surveillance and forensics. However, it is laborious for a camera operator to do this over multiple cameras. Therefore, an automated system is desirable that can assist the operator in his search for specific persons. For automatic person tracking over multiple cameras without overlapping views, the main component is a person-matching algorithm. The task of this algorithm is to find the person images in a large collection that are most similar to a query person image. Images of the same person are ranked as high as possible, preferably first. In this paper, we presented our computationally efficient and training-free person-matching method, which is based on multi-dimensional histograms containing color and spatial information. We compared the performance of our method to several state-of-the-art methods. To evaluate the performance of these algorithms the VIPeR dataset (from UC Santa Cruz) was used, because it is a large publically available benchmark for viewpoint-invariant person re-identification algorithms. This dataset consists of two different recordings of 632 individuals and each recording is a still image that tightly encloses the person. The data contains a wide variety of view-points, poses, backgrounds and lighting conditions, as is typically seen in surveillance systems in an outdoor situation. To estimate the performance and allow comparison, two-fold cross validation was used. The results show that our method performs well. The system is able to retrieve approximately 50% of the images correctly within the best 10 matches. Our system allows a human operator to speed up the tracking process with a factor of 5.

REFERENCES

[1] Alahi, A., Vandergheynst, P., Bierlaire, M., Kunt, M., "Cascade of descriptors to detect and track objects across any network of cameras", Computer Vision and Image Understanding 114(6), 624-640 (2010).

[2] Bak, S., Corvee, E., Bremond, F., Thonnat, M., "Person re-identification using Haar-based and DCD-based signature", IEEE Conf. Advanced Video and Signal Based Surveillance, 1-8 (2010).

[3] Bak, S., Corvee, E., Bremond, F., Thonnat, M., "Person re-identification using spatial covariance regions of human body parts", IEEE Conf. Advanced Video and Signal Based Surveillance, 435 - 440 (2010).

[4] Belongie, S. Malik, J., Puzicha, J., "Shape matching and object recognition using shape contexts", IEEE Trans. PAMI 24(4), 509-522 (2002).

[5] Borsboom, S., [Person matching under large changes in viewpoint and lighting], M.Sc. Thesis University of Amsterdam, Amsterdam The Netherlands, (2011).

[6] Bouma, H., Hanckmann, P., Marck, J.W., Penning, L., Hollander, R., Hove, J.M., Broek, S., Schutte, K., Burghouts, G., "Automatic human action recognition in a scene from visual inputs", Proc. SPIE 8388, (2012).

[7] Cai, Y., Pietikainen, M., "Person re-identification based on global color context", LNCS 6468, 205-215 (2011).

[8] Chen, D., Bilgic, M., Getoor, L., Jacobs, D., Mihalkova, L., Yeh, T., "Active inference for retrieval in camera networks", IEEE Person Oriented Vision, 13-20 (2011).

[9] Dalal, N., Triggs, B., "Histogram of oriented gradients for human detection", IEEE CVPR, (2005).

[10] Dijk, J., Rieter-Barrell, Y., Rest, J. van, Bouma, H., "Intelligent sensor networks for surveillance", Journal of Police Studies: Technology-Led Policing 3(20), 109-125 (2011).

[11] Dikmen, M., e.a., "Pedestrian recognition with a learned metric", ACCV LNCS 6495, 501-512 (2010).

[12] Doretto, G., Sebastian, T., Tu, P., Rittscher, J., "Appearance-based person reidentification in camera networks: problem overview and current approaches", J. Ambient Intell. Humanized Computing 2(2), 127-151 (2011).

[13] Eekeren, A. van, Schutte, K., Dijk, J., Lange, D.J.J. de., "Super resolution on moving objects and background", IEEE Int. Conf. Image Processing, 2709-2712 (2006).

[14] Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M., "Person re-identification by symmetry-driven accumulation of local features", IEEE Conf. Computer Vision and Pattern Recognition, 2360 - 2367 (2010).

[15] Felzenszwalb, P.F, Girshick, R.B., McAllester, D., "Cascade object detection with deformable part models", IEEE Conf. Computer Vision and Pattern Recognition, (2010).

[16] Fogel, I., Sagi, D., "Gabor filters as texture discriminator", Biological Cybernetics 61(2), 103-113 (1989).

[17] Gheissari, N., Sebastian, T., Hartley, R., "Person reidentification using spatio-temporal appearance", IEEE CVPR, 1528-1535 (2006).

[18] Gray, D., Tao, H., "Viewpoint invariant pedestrian recognition with an ensemble of localized features", Proc. Europ. Conf. Computer Vision (1), 262-275 (2008).

[19] Gray, D., Brennan, S., Tao, H., "Evaluating appearance models for recognition, reacquisition, and tracking", IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance, (2007).

[20] Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B., "Person re-identification in multi-camera systems based on interest-point descriptors collected on short video sequences", Int. Conf. Distr. Smart Cam., (2008).

[21] Hu, N., Bouma, H., Worring, M., "Tracking individuals in surveillance video of a high-density crowd", Proc. SPIE 8399, (2012).

[22] Huang, C., Wu, Y., Shih, M., "Unsupervised pedestrian re-identification for loitering detection", Adv. in Im. and Vid. Tech. LNCS 5414, 771-783 (2008).

[23] Jungling, K., Arens, M., "A multi-staged system for efficient visual person reidentification", IAPR Conf. Machine Vision Applications, (2011).

[24] Koppen, P., Worring, M., "Multi-target tracking in time-lapse video forensics", Proc. ACM workshop Multimedia in Forensics, (2009).

[25] Kuo, C, Nevatia, R., "How does person identity recognition help multi-person tracking?", CVPR, (2011).

[26] Laptev, I., "Improving object detection with boosted histograms", Im. Vision Comp. 27(5), 535-544 (2009).

[27] Le, T., Thonnat, M., Boucher, A., Bremond, F., "Appearance based retrieval for tracked objects in surveillance videos," Proc. ACM Int. Conf. Image and Video Retrieval, (2009).

[28] Lian, G., Lai, J., Zheng, W., "Spatial-temporal consistent labeling of tracked pedestrians across non-overlapping camera views", Pattern Recognition 44(5), 1121-1136 (2010).

[29] Lin, Z., Davis, L.S., “Learning pairwise dissimilarity profiles for appearance recognition in visual surveillance”, Adv. in Vis. Comp. LNCS 5358, 23–34 (2008).
[30] Ma, Y., Miller, B., Cohen, I., “Video sequence querying using clustering of objects appearance models”, Adv. in Vis. Comp. LNCS 4842, 328-339 (2007).
[31] Mensink, T., Zajdel, W., Krose, B., “Distributed EM learning for appearance based multi-camera tracking”, ACM/IEEE Int. COnf. Distributed Smart Cameras, 178-185 (2007).
[32] Metternich, M.J., Worring, M., Smeulders, A., “Color based tracing in real-life surveillance data”, Trans. on data hiding and multimedia security LNCS 6010, 18-33 (2010).
[33] Metternich, M.J., Worring, M., “Semi-interactive tracing of persons in real-life surveillance data”, ACM workshop on Multimedia in forensics, security and intelligence, (2010).
[34] Pham, T., Worring, M., Smeulders, A., “A multi-camera visual surveillance system for tracking of reoccurrences of people”, Proc. Int. Conf. Distributed Smart Cameras, (2007).
[35] Prosser, B, Zheng, W, Gong, S, Xiang, T, “Person re-identification by support vector ranking”, BMVC, (2010).
[36] Rest, J., Bovenkamp, E., Eendebak, P., Baan, J., and Van Munster, R., “Sensors and tracking crossing borders," Proc. Conf. Safety and Security Systems in Europe, (2009).
[37] Sande, K., Gevers, T., Snoek, C., “Evaluation of color descriptors for object and scene recognition”, IEEE CVPR, (2008).
[38] Satta, R., Fumera, G., Roli, F., “Exploiting dissimilarity representations for person re-identification”, Similarity-Based Pattern Recognition LNCS 7005, 275-289 (2011).
[39] Satta, R., Fumera, G., Roli, F., Cristani, M., Murino, V., “A multiple component matching framework for person re-identification”, Int. Conf. Image Analysis and Processing LNCS 6979, 140-149 (2011).
[40] Schmid, C., “Constructing models for content-based image retrieval”, IEEE CVPR 2, 39-45 (2001).
[41] Stoettinger, J., Hanburry, A., Sebe, N., Gevers, T., “Do colour interest points improve image retrieval?” IEEE Int. Conf. Image Processing, 169-172 (2007).
[42] Wang, X., Doretto, G., Sebastian, T., e.a., “Shape and appearance context modeling”, IEEE ICCV, (2007).
[43] Withagen, P.J., Schutte, K., Groen, F.C.A., “Likelihood-based object detection and object tracking using a color histograms and EM”, Proc. IEEE Int. Conf. Image Processing (1), 589-592 (2002).
[44] Yu, Y., Harwood, K., Yoon, K., Davis, L., “Human appearance modeling for matching across video sequences”, Machine Vision and Applications 18(3), 139-149 (2007).
[45] Zajdel, W., Zivkovic, Z., Krose, B., “Keeping track of humans: have I seen this person before?”, IEEE Int. Conf. on Robotics and Automation, (2005).
[46] Zheng, W., Gong, S., Xiang, T., “Associating groups of people”, BMVC, (2009).