People Re-identification Based on Bags of Semantic Features

Zhi Zhou[1], Yue Wang[2], Eam Khwang Teoh[1]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore 639798, Singapore
[2]Visual Computing Department, Institute for Infocomm Research (I2R), Singapore 138632, Singapore

Abstract. People re-identification has attracted a lot of attention recently. As an important part in disjoint cameras based surveillance system, it faces many problems. Various factors like illumination condition, viewpoint of cameras and occlusion make people re-identification a difficult task. In this paper, we exploit the performance of bags of semantic features on people re-identification. Semantic features are mid-level features that can be directly described by words, such as hair length, skin tone, race, clothes colors and so on. Although semantic features are not as discriminative as local features used in existing methods, they are more invariant. Therefore, good performance on people re-identification can be expected by combining a set of semantic features. Experiments are carried out on VIPeR dataset. Comparison with some state-of-the-art works is provided and the proposed method shows better performance.

1 Introduction

Multi-camera based surveillance system is more and more popular in our daily life. How to monitor the environment and collaboration with such a surveillance system becomes a challenge. Considering there are usually 20-40 cameras in a surveillance system, it is costing and inefficient if only rely on human visual inspectors. Thus, a system which is able to automatically detect and identify the people appears in cameras and understand his/her behavior is highly desiered. One of the tasks is to identify the person when he/she disappears from one camera and appears in another one, known as people re-identification. Efficient people re-identification is an important part for people tracking in a surveillance system with disjoint located cameras and it has become a popular research topic.

Some problems exist in people re-identification, such as illumination change and view-point change. Inappropriate selection of features could lead to poor performance. In this paper, we exploit the performance of semantic features on people re-identification. A set of semantic features are selected from all over the body. Experiment is carried out on VIPeR dataset [1]. Some image samples in VIPeR dataset are shown in Fig. 1. Comparison with some state-of-the-art methods is provided.

Fig. 1. Some examples of image pairs from VIPeR dataset [1].

The remaining structure of this paper is arranged as follows. Section 2 gives a brief introduction on related works. The proposed method is detailed in Section 3. Experimental results and discussions are presented in Section 4. Finally, the conclusion of this paper is given out in Section 5.

2 Related Works

Lots of works have been done on people re-identification in the past several years. Some works tried to represent the target by extracting color information as feature. One of the earliest works is from Javed [2], one color histogram is used to represent a person. Then information combining color, position and time is feed into a Bayesian model to track a person in disjoint cameras. Bird [3] divided the human body into ten horizontal regions. The descriptor is formed by cascading values of median color in HSV color spaces from each region. Kao [4] employed color information from several images to form a hierarchical color structure. Then the similarity of two images is measured based on Bayesian decision. In Madden's work [5], an on-line clustering method is used to classify pixels belonging to a person. Then color spectrum histogram is used to represent the person.

Besides color, some works exploited texture information. Farenzena [6] partitioned the person into head, torso and leg. Then for each part, local features describing texture like Maximally Stable Color Regions and Recurrent High-Structured Patches are used together with color to represent this person. Berdugo [7] combined background subtraction method and saliency map to segment the people first. Three texture features–oriented gradients, color ratio and color saliency were employed to obtain a discriminative descriptor.

Other local features such as key-points are used as well. Gheissari [8] combined interest point matching and model fitting in corresponding body parts for people re-identification. Hamdoun [9] extracted interest points from body and use descriptors of interest points to model this person in a K-D tree. People re-identification is done by matching interest points and searching in the K-D tree. Wang [10] introduced local shape information and appearance context for the representation of people. Histogram of Gradient (HOG) in the log-RGB color space and spatial distribution of the appearance are used as local descriptors. In Azizs [11] method, SIFT, SURF and Spin images are used as descriptors to represent people. Bak [12] extracted Mean Riemannian Covariance (MRC) patches during the track to model the people detected by HOG. The combination of MRC patches was then used for re-identification. Zheng [13] made use of information from close neighborhood to identify a person in a group of people. Salient features are used with unsupervised method in Zhao's method [14].

Some researchers treat people re-identification as a recognition problem, they improved the performance by enhancing the similarity calculation or by exploiting machine learning methods. Zheng [15] proposed a Probabilistic Relative Distance Comparison (PRDC) model to optimize the calculation of similarity between a pair of images. Gray [16] selected a set of local features including eight color channels and texture features like Schmid and Gabor. AdaBoost is used to train samples to get the optimal weight for each feature. Similarly, Prosser [17] used SVM to train samples and same features are used. In Baks [18] method, two kinds of local features– haar-like features and dominant color descriptor (DCD) are extracted. AdaBoost is used to choose the most distinctive features.

Recently, some semantic features like soft biometrics have been used to solve face identification problems. Dantcheva and Dugelay [19] proposed a face re-identification method based on three soft biometric traitshair, skin and clothes. Vaquero [20] introduced a new people identification method by using semantic features in the head area such as facial hair type, type of eyewear and hair type. Classifier is used to train selected attributes. Layne [21] [22] defined soft-biometrics as attributes and used machine learning methods like matric learning or SVM. Selected attributes are trained and binary results are given for a test image.

3 The Proposed Method

In this paper, we conduct people re-identification with bags of semantic features. Features used cover the whole body area, including hair length, colors of clothes and pants, length of sleeves and pants, clothes patterns, intensity contrast between clothes and pants. Though each feature is not as discriminative as local features, a discriminative descriptor can be obtained by combining these features. In this section, details of extracting these semantic features will be introduced, followed by the similarity calculation when we conduct the re-identification experiment on the VIPeR dataset [1].

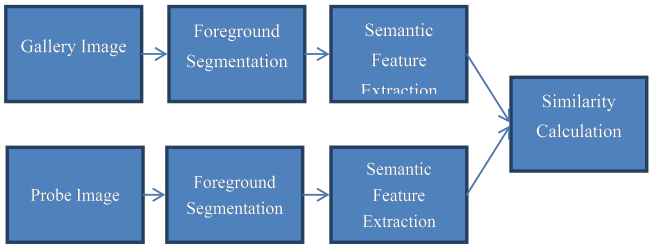

Fig. 2. Flow chart of the proposed method.

Fig. 3. Examples of foreground extraction.

A flow chart of the proposed method is shown is Fig. 2. Images in both gallery set and probe set are first processed to segment the foreground, which means the human body. Then semantic features are extracted from three parts– head, torso and leg. Then similarity between two images are calculated based on extracted semantic features.

3.1 Pre-processing

Since images in VIPeR [1] contain some background information around the people. The involvement of background information always downgrade the extraction of human descriptors. Therefore, Stel Component Analysis [23] is used to segment the human body and remove background for better performance. Some segmentation results are shown in Fig. 3. As we can see, human body can be well segmented from clustering background, though minor background still exist.

3.2 Semantic Feature Extraction

To obtain a more discriminative descriptor, semantic features that cover the whole body of people are selected. In the head area, soft biometrics [24] can be used. The rest of body is divided into torso and leg areas. Semantic features related to the apparel of people can be extracted in both two areas.

Fig. 4. Examples of hair segmentation.

Hair Length A lot of soft biometric traits are listed in [24], such as glasses, hair length, hats, facial hair and so on. However, due to the low resolution in images, especially in images from surveillance cameras, some of those traits are hard to be extracted and do not perform well in people re-identification. In this paper, only hair length is selected because it is observable, even in images with low resolution and various view angles. To extract hair length, Wang's method [25] is adopted, it's a morphological method and demands the segmentation of hair first. Mean shift segmentation [26] is used to separate hair from the other parts of head area. The hair length is then calculated and categorized into short hair and long hair. Some examples of the segmentation result are shown in Fig. 4.

Color from Torso and Leg Color is considered in both torso and leg areas. Color is the most common used feature in computer vision tasks. However, it usually suffers from varying factors like illumination. To constrain the effect of illumination change and different color specifications in cameras, colors are represented in nine predefined categories in the proposed method, instead of using color histograms. Nine common color categories are defined as black, gray, white, red, green, blue, cyan, magenta and yellow. Color histogram with nine bins is calculated in HSV color space, and the color bin with maximum vote is picked.

Color Intensity Contrast Even represented in pre-defined categories, the same object may still appear in different colors in different cameras. Therefore, as a complementary feature to color, the color intensity contrast between torso area and leg area is selected as one of the semantic features. Color intensity contrast is not discriminative, but the link between two color intensities is invariant to those changing factors. The color intensity histograms from torso area and leg area are compared and the contrast is classified into four categories– High contrast with brighter torso, High contrast with darker torso, Low contrast with brighter torso and Low contrast with darker torso.

Clothes Pattern Besides color, clothes pattern is also used in torso area. Under circumstances like illumination change, the color may change, but the pattern of clothes remains the same. Similar with color, five predefined categories are used– plain color, horizontal stripe, vertical stripe, checker and complex patterns. The

Fig. 5. Examples of clothes pattern. From left to right are plain colors, horizontal stripes and checkers. Vertical stripe sample does not exist in VIPeR.

Fig. 6. Examples of skin detection.

first four are common clothes patterns in our daily life while the rest is classified into the fifth category. As shown in Zhou's work [27], clothes pattern can be obtained by using histograms of magnitude and orientation of gradient in the torso area. Plain color is decided only by the histogram of magnitude of gradient. Since plain color clothes usually form few edges in the image, large proportion of low magnitude pixels exist in the histogram of magnitude of gradient, while in histograms of the other categories are not. Furthermore, by using the histogram of orientation, we can detect horizontal stripes and vertical stripes because they cause intensive response in horizontal and vertical directions respectively. The rest categories are classified by both of the two histograms. Some examples of defined categories are shown in Fig. 5.

Length of Sleeves and Pants One of the common ways for human to describe the apparel of a people is the length of sleeves or pants. Therefore, the length of sleeves and pants are used in the proposed method as semantic features. A simple way to detect the length of sleeves or pants is not to detect directly. Since color of skin is usually gathering in a small range of color space [28] and easy to be detected, we can extract the length of sleeves or pants by detect nude arms and legs instead. HSV color space is used in the proposed method and a compact range in H and S plane [29] is used to detect skins. Some examples of skin detection is shown in Fig. 6. Then the length of sleeves or pants can be decided by extracting the length of nude arms or legs. Two categories are used in both the detection of sleeve length and pant length.

3.3 Similarity calculation

After all semantic features extracted, similarity between two images are calculated as weighted combination of features' similarities

$$S\left(I_p, I_q\right) = \sum_s h_s\left(s_p, s_q\right) \tag{1}$$

where $h_s(s_p, s_q)$ represents the weighted similarity of one feature between two images s_p and s_q, and it is calculated as

$$h_s\left(s_p, s_q\right) = \begin{cases} w_c, if & s_p = s_q = c \\ 0, & else \end{cases} \tag{2}$$

Here, w_c is the weight for a specific category of a semantic feature. Weights are calculated beforehand with some samples, which will not be included into the images used for people re-identification. For each sample, images in both gallery set and probe set of the dataset are used. For each feature, every category is assigned a weight, based on the probability of its appearance in sample images. It is calculated as

$$w_c = \frac{n_{ab}}{n_a + n_b - n_{ab}} \times \frac{n_s}{\sqrt{n_a n_b}} \tag{3}$$

where n_{ab} is the number of samples with both images in gallery set and probe set belong to this category, n_a and n_b are the number of samples at least one of the images belongs to this category, n_s is the number of samples used. In this way, those common categories will receive lower weights while categoreis seldom appear receive high weights. Finally, weights within one semantic feature are normalized.

4 Experiment and Discussion

To test the performance of the proposed method on people re-identification, we use the public dataset VIPeR [1]. VIPeR contains image pairs of 632 people. For each people, two images are captured from cameras with different viewpoint and put into two subsets respectively. This is a challenging dataset since these images suffer from severe illumination changes and changes of viewpoint. Even worse, some occlusion is included. In the experiment, images are grouped into gallery set and probe set. For each person, one image is in gallery set and the other is in probe set. Then, for one image in the probe set, the similarities with images in gallery set are calculated, and an ID is assigned to it

$$ID_j = arg \max_{i \in 1,\ldots N} S\left(g_i, p_j\right) \tag{4}$$

To evaluate the people re-identification performance, we use Cumulative Matching Characteristic (CMC) curves and Synthetic Recognition Rate (SRR) curves. For an image in probe set, a rank score is obtained by matching with

images in gallery set. As the rank score increase, CMC curve records the probability of finding right matches within the rank score. It is a cumulative curve that increases as the rank score increases. For SRR curve, a specific number of people are first selected, then matching is done on these people. SRR curve records the probability that finds the right matches in the first rank. This curve decreases as the number of people selected increases. Besides these two curves, Area Under Curve (AUC) for CMC curve and some statistics at lower rank scores are also provided.

Table 1. Detection rates for using single features (in percentage).

Features	Hair Length	Clothes Color	Clothes Pattern	Sleeve Length	Pant Color	Pant Length	Color Contrast
Detection Rate	53.8	51.42	59.34	76.58	56.01	79.59	51.58

First, we compare the performance of the combination of semantic features with performances of single features. Detection rates of using only single features are shown in Table 1. The detection rate of sleeve length and pant length achieve 70% above accuracy, while other features just have detection rate slightly higher than 50% due to the dramatic changes in illumination and viewpoint. Table 2 listed the number of categories of each feature, and weights of semantic features contributed to the combined descriptor matching.

Table 2. Number of categories of each feature and weights (in percentage) contributed to the descriptor matching.

Features	Hair Length	Clothes Color	Clothes Pattern	Sleeve Length	Pant Color	Pant Length	Color Contrast
Categories of feature	2	9	4	2	9	2	4
Weights contributed	10.34	28.43	6.07	9.1	25.47	4.85	15.74

In Fig. 7, the CMC curves show that the recognition rates of using single features are low. However, when these features are combined, the recognition rate is significantly improved. AUCs are listed in Table 3. Fig. 8 shows the SRR curve. The features combined method performs better than single feature method.

In addition, we compare the proposed method with some state-of-the-art works. SDALF [6] and PRDC [15] are selected. All codes of these methods are provided by authors. Compared with these methods, the proposed method does

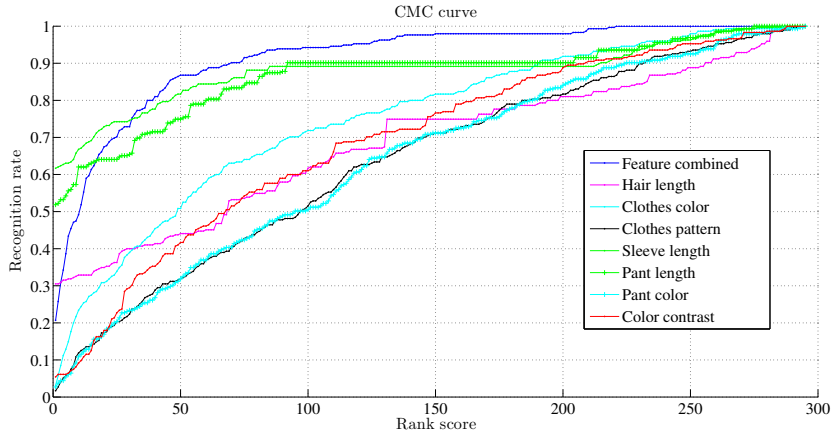

Fig. 7. Comparison on CMC curves between features combined and single features.

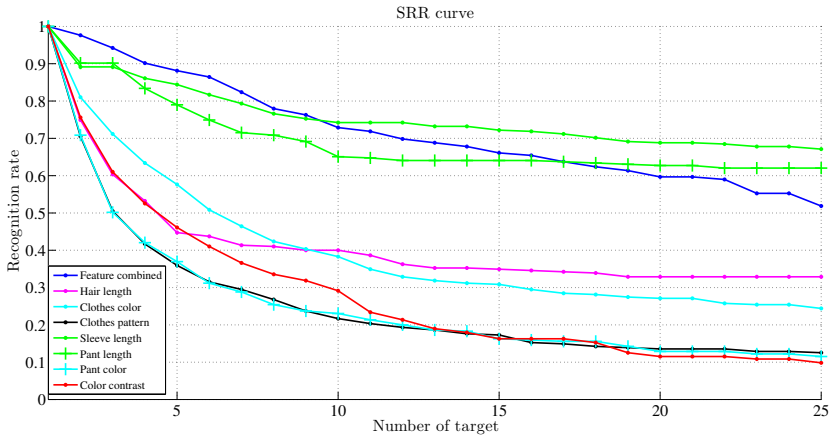

Fig. 8. Comparison on SRR curves between features combined and single features.

Table 3. Comparison of AUCs (in percentage), the best is show in bold.

Features	AUCs	Features	AUCs
Combined	**87.85**	Hair Length	68.73
Clothes Color	75.24	Clothes Pattern	63.89
Sleeve Length	86.04	Pant Color	64.11
Pant Length	84.92	Color Contrast	67.78

not need training beforehand for a specific database, because all semantic features selected can be extracted with simple methods. Since training is needed

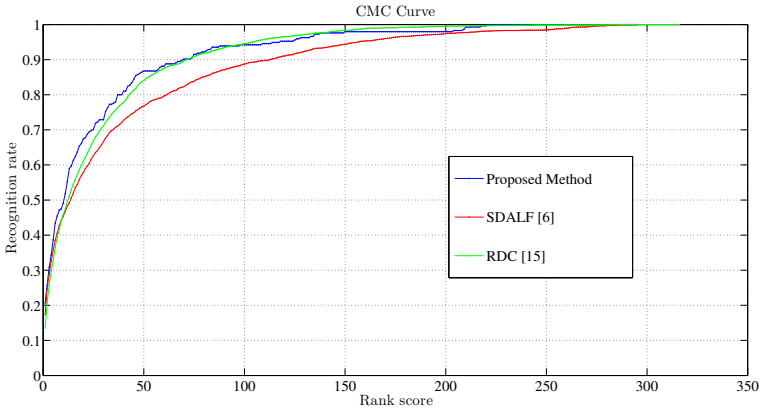

Fig. 9. Comparison on CMC curves between the proposed method and SDALF [6], PRDC [15].

for RDC, half of 632 peopl are used as training samples. The rest are used for people re-identification.

Fig. 9 shows the CMC curves. 316 pairs of images are used in people re-identification. The proposed method shows superior performance than the other two. Area Under Curve (AUC) for CMC curve is shown in Table 4 . The proposed method achieves the best AUC. Some statistics at lower rank scores of CMC curve are shown in Table 5. At lower rank scores, the proposed method achieves the best recognition rate. SRR curves are shown in Fig. 10.

Table 4. Comparison of AUCs (in percentage), the best is shown in bold.

	SDALF [6]	PRDC [15]	Proposed Method
AUCs	87.85	91.71	**91.83**

Table 5. Comparison of recognition rate (in percentage) in lower rank scores, the best is shown in bold.

	SDALF [6]	PRDC [15]	Proposed Method
Rank 1	13.54	17.32	**20.68**
Rank 5	32.82	35.69	**38.64**
Rank 10	45.95	45.39	**49.15**
Rank 25	66.87	62.68	**70.17**
Rank 50	84.05	76.78	**86.78**

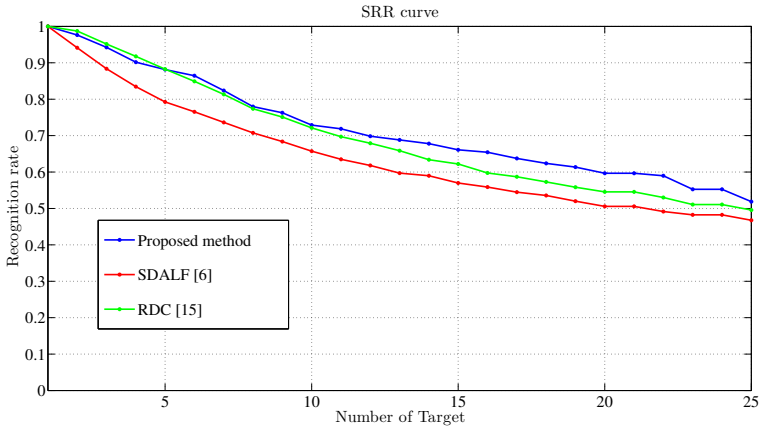

Fig. 10. Comparison on SRR curves between the proposed method and SDALF [6], PRDC [15].

Fig. 11. Examples of image pairs suffer from severe illumination change.

Some of the reasons undermine the performance of the proposed method are discussed as below.

(1) The detection of hair length is mostly affected by different viewpoints of cameras. For example, the hair length could be different when observing from the front view and from the back view. Because in the front view, long hair could be occluded by head and body and be mistaken as short hair. Since segmentation is performed to divide face and hair in the proposed method, the performance of segmentation may also undermine the detection rate of hair length.

(2) The detections of clothes color and pant color are severely affected by the dramatic illumination change in VIPeR dataset. In some images, the illumination is too dark or too bright to differentiate colors (Fig. 11). The severe lighting condition also affects the extraction of length of sleeves and pants, because skin can not be detected under some extreme lighting conditions and lead to the failure on detecting short sleeves and pants.

(3) Different viewpoints of cameras raise the problem that, some people carry backpack and have different colors or patterns with the clothes. In this case, features from front view and back view are mismatched.

5 Conclusion

In this paper, we propose a people re-identification method with bags of semantic features. 7 features are selected, which cover the whole body. Without using learning method, each feature can be extracted with simple method. The experiment shows that the combination of these semantic features outperforms single feature. Experimental results also show that the proposed method achieves good performance compared with state-of-the-art methods.

Semantic features used in the proposed method are far from enough for robust and discriminative description. In our future work, we will try to include more semantic features. Also, we will exploit the combination of robustness from semantic features and discriminative from local features. Another way to improve the description is adopting an advanced method to conduct more detailed and accurate segmentation on the human body.

References

1. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International workshop on performance evaluation of tracking and surveillance. (2007)
2. Javed, O., Rasheed, Z., Shafique, K., Shah, M.: Tracking across multiple cameras with disjoint views. In: Proceedings of Ninth IEEE International Conference on Computer Vision (ICCV). (2003) 952–957
3. Bird, N.D., Masoud, O., Papanikolopoulos, N.P., Isaacs, A.: Detection of loitering individuals in public transportation areas. IEEE Transactions on Intelligent Transportation Systems **6** (2005) 167–177
4. Kao, J.H., Lin, C.Y., Wang, W.H., Wu, Y.T.: A unified hierarchical appearance model for people re-identification using multi-view vision sensors. In: Advances in Multimedia Information Processing-PCM. (2008) 553–562
5. Madden, C., Cheng, E.D., Piccardi, M.: Tracking people across disjoint camera views by an illumination-tolerant appearance representation. Machine Vision and Applications **18** (2007) 233–247
6. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2010) 2360–2367
7. Berdugo, G., Soceanu, O., Moshe, Y., Rudoy, D., Dvir, I.: Object reidentification in real world scenarios across multiple non-overlapping cameras. In: Proc. Euro. Sig. Proc. Conf. (2010) 1806–1810
8. Gheissari, N., Sebastian, T.B., Hartley, R.: Person reidentification using spatiotemporal appearance. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Volume 2. (2006) 1528–1535
9. Hamdoun, O., Moutarde, F., Stanciulescu, B., Steux, B.: Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences. In: Second ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC). (2008) 1–6
10. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.: Shape and appearance context modeling. In: In: Proceedings of IEEE International Conference on Computer Vision (ICCV). (2007)

11. Aziz, K.E., Merad, D., Fertil, B.: Person re-identification using appearance classification. In: Image Analysis and Recognition. (2011) 170–179
12. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Boosted human re-identification using riemannian manifolds. Image and Vision Computing **30** (2012) 443–452
13. Zheng, W.S., Gong, S., Xiang, T.: Associating groups of people. In: Proc. British Machine Vision Conference (BMVC). (2009)
14. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2013) 3586–3593
15. Zheng, W.S., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence **35** (2013) 653–668
16. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Proc. European Conference on Computer Vision (ECCV). (2008) 262–275
17. Prosser, B., Zheng, W.S., Gong, S., Xiang, T., Mary, Q.: Person re-identification by support vector ranking. In: Proc. British Machine Vision Conference (BMVC). Volume 1. (2010) 5
18. Bak, S., Corvee, E., Brémond, F., Thonnat, M.: Person re-identification using haar-based and dcd-based signature. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). (2010) 1–8
19. Dantcheva, A., Dugelay, J.L.: Frontal-to-side face re-identification based on hair, skin and clothes patches. In: IEEE International Conference on Advanced Video and Signal-Based Surveillance (AVSS). (2011) 309–313
20. Vaquero, D.A., Feris, R.S., Tran, D., Brown, L., Hampapur, A., Turk, M.: Attribute-based people search in surveillance environments. In: Workshop on Applications of Computer Vision (WACV),. (2009) 1–8
21. Layne, R., Hospedales, T.M., Gong, S., et al.: Person re-identification by attributes. In: BMVC. Volume 2. (2012) 3
22. Layne, R., Hospedales, T.M., Gong, S.: Towards person identification and re-identification with attributes. In: Computer Vision–ECCV 2012. Workshops and Demonstrations, Springer (2012) 402–412
23. Jojic, N., Perina, A., Cristani, M., Murino, V., Frey, B.: Stel component analysis: Modeling spatial correlations in image class structure. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (2009) 2044–2051
24. Dantcheva, A., Velardo, C., Dangelo, A., Dugelay, J.L.: Bag of soft biometrics for person identification. Multimedia Tools and Applications **51** (2011) 739–777
25. Wang, Y., Zhou, Z., Teoh, E.K.: Human hair segmentation and length detection for human appearance model. In: Internation Conference on Pattern Recognition (ICPR). (2014)
26. Comaniciu, D., Meer, P.: Mean shift: A robust approach toward feature space analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence, **24** (2002) 603–619
27. Zhou, Z., Wang, Y., Teoh, E.K.: People apparel model for human appearance description. In: International Conference on Information, Communications and Signal Processing (ICICS). (2013) 1–5
28. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: Proc. Graphicon. Volume 3. (2003) 85–92
29. Sobottka, K., Pitas, I.: A novel method for automatic face segmentation, facial feature extraction and tracking. Signal processing: Image communication **12** (1998) 263–281