Person Re-identification using Semantic Color Names and RankBoost

Cheng-Hao Kuo[1], Sameh Khamis[2*], and Vinay Shet[1]
Imaging and Computer Vision, Siemens Corporation, Corporate Technology[1], Princeton, NJ
University of Maryland[2], College Park, MD
{cheng-hao.kuo|vinay.shet}@siemens.com[1] sameh@umiacs.umd.edu[2]

Abstract

We address the problem of appearance-based person re-identification, which has been drawing an increasing amount of attention in computer vision. It is a very challenging task since the visual appearance of a person can change dramatically due to different backgrounds, camera characteristics, lighting conditions, view-points, and human poses. Among the recent studies on person re-id, color information plays a major role in terms of performance. Traditional color information like color histogram, however, still has much room to improve. We propose to apply semantic color names to describe a person image, and compute probability distribution on those basic color terms as image descriptors. To be better combined with other features, we define our appearance affinity model as linear combination of similarity measurements of corresponding local descriptors, and apply the RankBoost algorithm to find the optimal weights for the similarity measurements. We evaluate our proposed system on the highly challenging VIPeR dataset, and show improvements over the state-of-the-art methods in terms of widely used person re-id evaluation metrics.

1. Introduction

Person re-identification is a critical problem in a video surveillance system. Its goal is to re-identify a person in different locations across multiple, potentially non-overlapping, cameras. Due to the unreliable spatial and temporal information, appearance-based person re-id has been drawing an increasing amount of attention in recent computer vision research. The common assumptions for this task include: a) the finer biometric cues (*e.g.* face, or iris) are not available due to the low image resolution; b) The targets of interest do not change their clothes across different cameras. In other words, appearance-based person re-id relies on the information provided by the visual appearance of human body and clothing. It is a highly challenging problem since human appearance usually exhibits large variations across different cameras. This variation is due to variability in backgrounds, sensor characteristics, lighting conditions, view-points, and human poses. Besides, distinct people may look similar if they wear clothes with the same color, which in turn increases the difficulty of finding correct associations.

Many existing approaches address this problem mainly by two important approaches: descriptor extraction and similarity/distance measurements. In the first approach, the goal is to find the invariant and distinctive representation to describe a person image. Several descriptors have been used, which include color histogram, Histogram of Oriented Gradients(HOG) [5], texture filters [12], Maximally Stable Color Regions(MSCR) [8], and decomposable triangulated model [9]. In the second approach, many existing methods typically use a standard distance measurement, *e.g.* Bhattacharyya distance, correlation coefficient, L1-Norm or L2-Norm. Among these descriptor and similarity measurements, the color histogram followed by Bhattacharyya distance are most widely used since the color information has been found as the most important cue in many person re-id studies. However, the performance of color histogram in any color space is still not satisfactory. There is an example in Figure 1 where HSV color histograms are used for a person re-id problem. It shows results that are counterintuitive to a human operator. For example, why does the lady wearing black have a higher rank than the lady wearing pink, when the lady in the query image is clearly wearing pink?

Inspired by the work in [24], we propose to apply semantic color names on the person re-id problem. As in [24], 11 basic color names are used: black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow. Given the mapping from RGB values to probability distribution over those color names, we can build a semantic histogram as a image descriptor. This representation is well suited for matching a pair of persons or searching for a target with semantic labels like "find a person wearing red shirts and blue pants".

*This author contributed to the work presented in this paper during his internship at Siemens Corporation, Corporate Technology.

(a) (b)

Figure 1. A query image(a) and the gallery images with sorted order(b). These results are generated using HSV color histogram applied in 6 horizontal non-overlapping strips of person images and the similarity score is computed by Bhattacharyya distances. Note that the query is wearing pink but the many candidates who are not wearing pink are ranked higher than the true match.

To better combine the semantic color names with other widely used features including color histogram, texture histogram, and covariance matrix, we first define the appearance-based affinity model as a linear combination of similarity measurements of local descriptors. Unlike other learning-based methods [12, 19, 23] which take a long raw vector with absolute difference as feature pool, ours is based on the similarity measurements of corresponding image descriptors. The weight for each similarity measurement is learned by the RankBoost algorithm. The advantages of this design include: a) the image descriptor does not need to lie in Euclidean space; b) it handles over-fitting better than when working on raw difference vectors; c) the weights clearly indicate the importance of the corresponding local image descriptors.

The rest of the paper is organized as follows. Related work is discussed in Section 2. The image descriptor based on semantic color names is presented in Section 3. The appearance affinity model and RankBoost learning framework are presented in Section 4. The experimental results are shown in Section 5. The conclusion is given in Section 6.

2. Related work

Matching people of interest across a network of non-overlapping cameras is an important task in a surveillance system, which is known as the person re-identification problem. The problem is also called as inter-camera association or multi-camera tracking. To establish the correspondence between objects in different cameras, a typical solution is to fuse two important cues: the spatio-temporal information and target appearance. There has been some early work [14, 6, 16, 10] which focuses on learning the spatio-temporal cues. For the appearance cues, some early work [17, 15, 10, 3, 18] mainly use color information and propose to learn brightness transfer functions(BTFs) or color calibration to handle the changing lighting condition in different cameras.

Since the spatio-temporal information of targets between cameras is unreliable, there has been an increasing interest in appearance-based person re-identification. Apparently, the color information with BTFs does not provide a satisfactory solution to this challenging problem. In past few years, many efforts are made for proposing more advanced descriptors and more sophisticated matching techniques to achieve high re-id accuracy.

Recent re-id approaches can be divided by two categories: a)non-learning based (direct) methods, and b) learning-based methods. The direct methods usually extract a set of hand-crafted descriptive representation and combine their corresponding distance measurements without learning. Gheissari *et al.*[9] develop two person re-identification approaches which use interest operators and model fitting for establishing spatial correspondences between individuals. Wang *et al.*[22] introduce shape and appearance context modeling by co-occurrence matrices. Farenzena *et al.* [7] found the asymmetry/symmetry axes and extracted the symmetry-driven accumulation of local features. Bak *et al.* [1] use body parts detector and spatial pyramid matching. Cheng *et al.* [4] utilize Pictorial Structures (PS) to localize the body parts and match their descriptors. On the other hand, learning-based methods usually extract a bunch of low-level descriptors, concatenate them into a long feature vector, and obtain discriminability by labeled training samples and machine learning techniques. Gray *et al.* [12] present an Adaboost-based method to find the best ensemble localized features sequentially. Schwartz *et al.* [20] established a high-dimensional signature which is then projected into a low-dimensional discriminant latent space by Partial Least Squares reduction. Prosser *et al.* [19] formulate person re-id as a ranking problem and propose the ensemble RankSVM to overcome the scalability problem. Zheng *et al.* [23] reformulate person re-id as a distance learning problem, which maximize the probability of a true match having a smaller distance than that of a wrong match. Hirzer *et al.* [13] propose a two-stage approach to combine the descriptive and discriminative models.

Figure 2. Some examples of pixel-wise assignments on VIPeR dataset. (a) Original examples. (b) Result images. Note that only the color name with highest probability are shown in the result images.

3. Semantic Color Names

Instead of using simple color histogram only, we propose to apply semantic color names to describe a person image in the re-id problem. To choose appropriate color names, we follow the basic color terms which were defined in the famous work on color naming [2]. A basic color term of a language is defined as being not subsumable to other basic color terms and extensively used in different languages. As in [24], we use the 11 basic color names in English language: black, blue, brown, grey, green, orange, pink, purple, red, white, and yellow.

To use color naming as an image descriptor, a mapping from the RGB values of a image pixel to color names is required. Instead of one-to-one matching, a probability distribution over the color names is used since a certain triplet in RGB space could be assigned to multiple color names. This mapping can be represented as:

$$f : \mathbf{x}_{RGB} \rightarrow \mathbf{v} \tag{1}$$

where $\mathbf{v}$ is a 11-element vector, and each element $v_i = P(c_i|\mathbf{x}_{RGB})$ is the probability of the RGB values being assigned to a specific color name c_i. The color name descriptor of region R, K_R, is defined as the summation of probability distribution from the pixels inside the region R:

$$K_R = \frac{1}{N} \sum_{\mathbf{x} \in R} f(\mathbf{x}_{RGB}) \tag{2}$$

The mapping from RGB value to probability distribution over 11 color names is a non-trivial problem. In [24], manually annotated images and hand-segmented the regions corresponding to the color label are used; the mapping is then inferred by Bayes Law assuming the prior probabilities are equal among all color names. In [25], the authors use Google Image instead of manual labelling to collect a training data set. To handle the noisy labels from Google Image, the color names are learned using a PLSA model. In this paper, we take the $16 \times 16 \times 16$ look-up table provided by [25] and do post-processing to cast the probabilities with impossible assignments to zero. Several image examples are shown in Figure 2.

4. Appearance Model and RankBoost Learning for Re-id

Descriptor extraction and matching affinity computation are important elements for a matching systems. In our design, the descriptors are defined as the ensemble of local features extracted in given support regions. The pair-wise similarity measurement from a specific feature over a specific region is computed. The final matching affinity is represented as a linear combination of the corresponding similarity measurements of the local descriptors. The combination coefficient is learned from the training data using RankBoost learning algorithm.

4.1. Local image descriptors and similarity measurements

To establish a strong appearance model, we extract a rich set of local descriptors to describe a person image. A local descriptor d consists of a feature channel λ and a support region r. Given an image sample I, a single descriptor $d_{i,j}$ extracted over r_j via λ_i is denoted as

$$d_{i,j} = I(\lambda_i, r_j) \tag{3}$$

where i and j are the indices of the feature channel and the support region respectively.

In our implementation, the support regions $\{r\}$ are 6 horizontal stripes, which cover the head, upper torso, lower torso, upper leg, and lower leg of human body. The feature channel λ are chosen from five types of features: color name probability distribution, color histogram, texture histogram, maximally stable color regions(MSCR), and covariance matrix. Color name probability distribution is a 11 dimensional vector, as defined in Equation (2). For the color histograms, we use RGB, HSV, and YCbCr color space and each channel of each color space form a 16 dimensional vector. For the texture histogram, Gabor and Schmid texture filter with 21 different parameters in total are applied to the luminance channel. The responses from each texture filter form a 16 dimensional vector. For the MSCR, the mean color and the coordinate of the detected blob region are recorded as descriptors. We use the implementation from the work [7]. For the covariance matrix, the feature set comprises of spatial, color and gradient information. It takes the following form:

$$\mathbf{C} = \frac{1}{n-1}\sum_{k=1}^{n}(\mathbf{z}_k - \boldsymbol{\mu})(\mathbf{z}_k - \boldsymbol{\mu})^T \tag{4}$$

where

$$\mathbf{z}_k = \left[y \; L \; a \; b \; \frac{\partial L}{\partial x} \; \frac{\partial L}{\partial y} \; \frac{\partial^2 L}{\partial x^2} \; \frac{\partial^2 L}{\partial y^2} \right]^T \tag{5}$$

is the vector containing y coordinate, pixel values in Lab color channels, first and second derivatives of image at k-th pixel; $\boldsymbol{\mu}$ is the mean vector and n is the number of pixels.

Given those descriptors, we can compute their corresponding similarity measurement between two image patches. Since the color name probability distribution, color histogram, and texture histogram are histogram-based features, we choose Bhattacharyya coefficient as the similarity measurement. The covariance matrix does not lie in Euclidean space, the distance between two covariance matrix can be determined by solving a generalized eigenvalue problem [21]:

$$\rho(\mathbf{C}_1, \mathbf{C}_2) = \sqrt{\sum_{k=1}^{8} ln^2 \lambda_k(\mathbf{C}_1, \mathbf{C}_2)} \tag{6}$$

where $\{\lambda_k(\mathbf{C}_1, \mathbf{C}_2)\}$ are the generalized eigenvalues of $\mathbf{C}_1$ and $\mathbf{C}_2$, computed from

$$\lambda_k \mathbf{C}_1 \mathbf{x}_k - \mathbf{C}_2 \mathbf{x}_k = 0 \qquad k = 1 \ldots 8 \tag{7}$$

and $\mathbf{x}_k \neq 0$ are generalized eigenvectors.

For MSCR, we employ the distance metric as described in [7]. Note that we adjust the sign and normalize the distance measurement to $[0, 1]$ such that it becomes similarity measurement.

In summary, the similarity score between two image patches based on a certain local descriptor can be written as:

$$s_{i,j} = \rho_i\big(I_1(\lambda_i, r_j), I_2(\lambda_i, r_j)\big) \tag{8}$$

where ρ_i is the corresponding similarity measurement function of feature channel λ_i.

4.2. Matching model definition and RankBoost learning

We define the appearance-based affinity model as an ensemble of local descriptors and their corresponding similarity measurements. It takes any two images of persons as input and computes an affinity score as the output. In our design, the appearance-based affinity models is a linear combination of all similarity measurements on different features and different regions by Equation (8). It takes the following form:

$$H(P_1, P_2) = \sum \alpha_{i,j} s_{i,j} \tag{9}$$

where the coefficients $\{\alpha\}$ represent the importance of local descriptors.

In the re-id problem, the desired model should have the goal of giving correct matches higher ranking than the incorrect ones. Therefore, we propose to formulate the re-id problem as a classic ranking problem. Supposed that we have three person images P_i, P_j, and P_k, where P_i and P_j correspond to the same individual, while P_k is a different individual. The ranking function H should prefer matching P_i and P_j than P_i and P_k. More formally, we seek to train an ideal model such that $H(P_i, P_j) > H(P_i, P_k)$.

Formally put, we define the instance set $\mathcal{X} = \mathcal{P} \times \mathcal{P}$, where $\mathcal{P}$ is the set of person images in our dataset. The ranking sample set is denoted by

$$\mathcal{R} = \{(x_{i,0}, x_{i,1} | x_{i,0} \in \mathcal{X}, x_{i,1} \in \mathcal{X})\} \tag{10}$$

where $x_{i,0}$ and $x_{i,1}$ each represent a pair of person images, and $(x_{i,0}, x_{i,1}) \in \mathcal{R}$ indicates that the association of $x_{i,1}$ should be ranked higher than $x_{i,0}$.

The loss function for boosting is defined as follows:

$$Z = \sum_i w_0(x_{i,0}, x_{i,1}) \mathbf{I}(H(x_{i,0}) - H(x_{i,1})) \tag{11}$$

where $\mathbf{I}$ is a indicator function and $w_0(x_{i,0}, x_{i,1})$ is the initial weight of the i-th sample, which will be updated during boosting. The goal is to find $H(x)$ that minimizes Z. As in traditional boosting, H is obtained by sequentially adding new weak ranker. In the t-th round, we try to find an optimal weak ranker $h_t : \mathcal{X} \to \mathbb{R}$ that minimizes

$$Z_t = \sum_i w_t(x_{i,0}, x_{i,1}) \mathbf{I}((h_t(x_{i,0}) - h_t(x_{i,1})) \tag{12}$$

Algorithm 1 Algorithm of RankBoost for Re-id

Input: ranking sample set $\mathcal{R} = \{(x_{i,0}, x_{i,1}) | x_{i,0} \in \mathcal{X}, x_{i,1} \in \mathcal{X})\}$

Set $w_0(x_{i,0}, x_{i,1}) = \frac{1}{|\mathcal{R}|}$
for $t = 0$ to T **do**
  Choose $k^* = \arg\min_k \sum w_t(x_{i,0}, x_{i,1}) \mathbf{I}((h_k(x_{i,0}) - h_k(x_{i,1})))$
  Set $h_t = h_{k^*}$
  Compute α_t as in Equation (13) and (14)
  Update $w_{t+1}(x_{i,0}, x_{i,1}) \leftarrow w_t(x_{i,0}, x_{i,1}) exp[\alpha_t(h_t(x_{i,0}) - h_t(x_{i,1}))]$
  Normalize $w_{t+1}(x_{i,0}, x_{i,1})$
end for

Output: $H(x) = \sum_{t=1}^{T} \alpha_t h_t(x)$

Once we find the optimal ranker, we compute the weight α_t by:

$$\alpha_t = \frac{1}{2} ln\left(\frac{1+r}{1-r}\right) \tag{13}$$

where

$$r = \sum_i w_t(x_{i,0}, x_{i,1})(h_t(x_{i,0}) - h_t(x_{i,1})) \tag{14}$$

Then we update the sample weights according to h_t and α_t to emphasize difficult ranking samples. The final strong ranking classifier is the weighted combination of the selected weak ranking classifiers: $H(x) = \sum_{t=1}^{n} \alpha_t h_t(x)$, where n is the number of boosting round. The described RankBoost algorithm is shown in Algorithm 1. In our implementation, the weak ranker is the similarity measurements defined in Equation (8).

5. Experimental results

To evaluate the performance of our proposed system, we conducted experiments on the highly challenging VIPeR dataset. The comparison between our system and several state-of-the-art methods is given based on the commonly used evaluation metrics [11]. Additionally, the effectiveness evaluation of the color name and RankBoost respectively is provided. The analysis of best feature channels is also presented.

5.1. Dataset and settings

We use the well-known dataset, VIPeR dataset, for our evaluation. The VIPeR dataset is arguably the most challenging dataset for person re-identification problem in the literature; many state-of-the-art methods report their numbers on it. There are 632 individuals captured in outdoor scenarios with two images for each person. In our experiments, we randomly selected 316 image pairs of people for testing set, and the rest are used for training set. Each test set was composed of a gallery set and a probe set. The probe set consists of one image for each person, and the remaining images are used as the gallery set. During training, a pair of images of each person form a positive pair, and one image of him/her and one of another person in the training set form a negative pair. For evaluation, we use the average cumulative match characteristic(CMC) over 10 trials to show the ranked matching rates. A rank r matching rate indicates the percentage of the probe images with correct matches found in the top r ranks in the 316 gallery images. Note the rank-1 matching rate is the true matching rate in an automatic system. However, in a normal surveillance setting, the top r ranked matching rate with a small r value is important as well since the top matched images will normally be verified by a human operator.

5.2. Performance Comparison

We compare our methods with several state-of-the-art methods, including Ensemble of Localized Features (ELF) [12], Primal-based RankSVM (PRSVM) [19], Symmetry-Driven Accumulation of Local Features (SDALF) [7], Probabilistic Relative Distance Comparison (PRDC) [23], and Pictorial Structure (PS) [4]. The comparison using top ranked matching rate on the VIPeR dataset is shown in Table 1. It is clear that our method achieves better matching rate compared to the state-of-the-art, especially with the rank-1 matching rate of 23.92% and rank-5 matching rate of 45.57%. Note that our proposed method has the general framework, which can be integrated with other work, *e.g.* [4], to achieve better results.

Method	r=1	r=5	r=10	r=20
Ours	23.92	45.57	56.23	68.73
PS [4]	21.84	44.64	57.21	71.23
SDALF [7]	19.87	38.89	49.37	65.73
PRDC [23]	15.66	38.42	53.86	70.09
PRSVM [19]	14.77	36.39	50.81	66.78
ELF [12]	12	31.5	44	61

Table 1. Comparison with the state-of-the-art methods by top ranked matching rate(%) on the VIPeR dataset. r is the rank.

5.3. Effectiveness evaluation of color names and RankBoost

We evaluate the effectiveness of the color name and RankBoost respectively. Several experiments are conducted based on various combinations of whether color names or RankBoost are disabled or not. The cumulative matching characteristic(CMC) curves are presented in Figure 3. The proposed method refers to that both color names and RankBoost are utilized. The method with disabled color names means that the color names descriptors are removed from the feature pool while keeping RankBoost enabled. The method with disabled RankBoost means that the weighting coefficients $\{\alpha\}$ are set to be equal while keeping color names descriptors used. The baseline method means that both components are disabled. The experimental result shows that both color names and RankBoost improve the performance, while RankBoost plays a more important role in the overall contribution.

5.4. Analysis of best feature channels

We also examine which feature channels have the most weights in the RankBoost learning process. The weights of corresponding feature channels are presented in Table 2. The most important features channels include Hue, Color names, and Saturation, and Covariance Matrix. It is shown that the proposed color names provide an important cue in the person re-id problem

6. Conclusion

We propose to use semantic color names in the person re-id problem. Compared to commonly used color histogram, the visual matching results by color names is much closer to what human operators would consider intuitive. We also propose to apply the RankBoost algorithm to learn the weights of similarity measurements of the corresponding local image descriptors. The experiment on challenging dataset shows the effectiveness of our proposed system compared to the state-of-the-art methods.

Feature Channel	Corresponding Weight
Hue	0.961
Color Names	0.704
Saturation	0.302
Covariance Matrix	0.288
Red	0.169
MSCR	0.160
Green	0.148
Y	0.049

Table 2. The first eight feature channels with the most weights learned from RankBoost algorithm. Not surprisingly the color related feature dominates in the feature selection. Meanwhile the proposed color names play an important role.

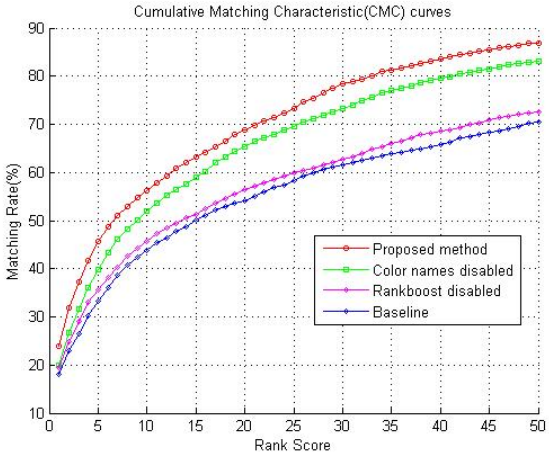


Figure 3. Effectiveness evaluation of two main components in this paper: color names and Rankboost. The results of four different combinations are shown in the CMC curves on the VIPeR dataset. It is shown that both color names and RankBoost contribute to the improvement of overall performance.

References

[1] S. Bak, E. Corvee, F. Bremond, and M. Thonnat. Person re-identification using spatial covariance regions of human body parts. In *AVSS*, 2010. 2

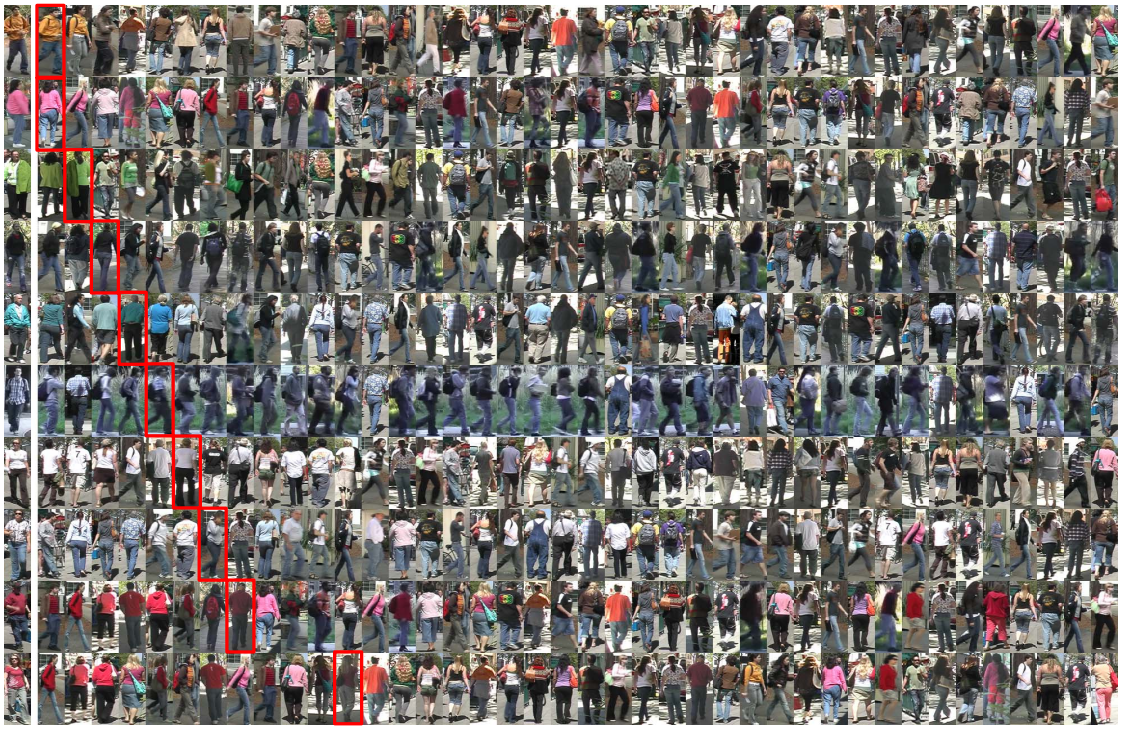
[2] B. Berlin and P. Kay. Basic color terms: Their universality and evolution. *Berkeley: University of California*, 1969. 3

[3] K.-W. Chen, C.-C. Lai, Y.-P. Hung, and C.-S. Chen. An adaptive learning method for target tracking across multiple cameras. In *CVPR*, 2008. 2

[4] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino. Custom pictorial structures for re-identification. In *BMVC*, 2011. 2, 5, 6

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, 2005. 1

[6] A. R. Dick and M. J. Brooks. A stochastic approach to tracking objects across multiple cameras. In *Australian Conference on Artificial Intelligence*, 2004. 2



(a) (b)

Figure 4. Example results on VIPeR dataset.(a) Probe images. (b) Top 40 results sorted from left to right. The images with red boxes represent the correct matches.

[7] M. Farenzena, L. Bazzani, A. Perina1, V. Murino, and M. Cristani. Person re-identification by symmetry-driven accumulation of local features. In *CVPR*, 2010. 2, 4, 5, 6

[8] P.-E. Forssen. Maximally stable colour regions for recognition and matching. In *CVPR*, 2007. 1

[9] N. Gheissari, T. B. Sebastian, P. H. Tu, and J. Rittscher. Person reidentification using spatiotemporal appearance. In *CVPR*, 2006. 1, 2

[10] A. Gilbert and R. Bowden. Tracking objects across cameras by incrementally learning inter-camera colour calibration and patterns of activity. In *ECCV*, 2006. 2

[11] D. Gray, S. Brennan, and H. Tao. Evaluating appearance models for recognition, reacquisition, and tracking. In *Proceedings of International Workshop on Performance Evaluation of Tracking and Surveillance(PETS)*, 2007. 5

[12] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, 2008. 1, 2, 5, 6

[13] M. Hirzer, C. Beleznai, P. M. Roth, and H. Bischof. Person re-identification by descriptive and discriminative classification. In *Scandinavian conference on Image analysis*, 2011. 2

[14] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. Tracking across multiple cameraswith disjoint views. In *ICCV*, 2003. 2

[15] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *CVPR*, 2005. 2

[16] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *CVPR*, 2004. 2

[17] F. Porikli. Inter-camera color calibration by correlation model function. In *ICIP*, 2003. 2

[18] B. Prosser, S. Gong, and T. Xiang. Multi-camera matching using bi-directional cumulative brightness transfer functions. In *BMVC*, 2008. 2

[19] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by support vector ranking. In *BMVC*, 2010. 2, 5, 6

[20] W. R. Schwartz and L. S. Davis. Learning discriminative appearance-based models using partial least squares. In *SIBGRAPI*, 2009. 2

[21] O. Tuzel, F. Porikli, and P. Meer. Region covariance: A fast descriptor for detection and classification. In *ECCV*, 2006. 4

[22] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu. Shape and appearance context modeling. In *ICCV*, 2007. 2

[23] J. Weijer and C. Schmid. Applying color names to image description. In *ICIP*, 2007. 1, 3

[24] J. Weijer, C. Schmid, and J. Verbeek. Learning color names from real-world images. In *CVPR*, 2007. 3

[25] W.-S. Zheng, S. Gong, and T. Xiang. Person re-identification by probabilistic relative distance comparison. In *CVPR*, 2011. 2, 5, 6