An Ensemble Color Model for Human Re-identification

Xiaokai Liu   Hongyu Wang
Dalian University of Technology
xiaokaigirl@mail.dlut.edu.cn, whyu@dlut.edu.cn

Yi Wu   Jimei Yang   Ming-Hsuan Yang
University of California at Merced
{ywu29,jyang44,myang37}@ucmerced.edu

Abstract

Appearance-based human re-identification is challenging due to different camera characteristics, varying lighting conditions, pose variations across camera views, etc. Recent studies have revealed that color information plays a critical role on performance. However, two problems remain unclear: (1) how do different color descriptors perform under the same scene in re-identification problem? and (2) how can we combine these descriptors without losing their invariance property and distinctiveness power? In this paper, we propose a novel ensemble model that combines different color descriptors in the decision level through metric learning. Experiments show that the proposed system significantly outperforms state-of-the-art algorithms on two challenging datasets (VIPeR and PRID 450S). We have improved the Rank 1 recognition rate on VIPeR dataset by 8.7%.

1. Introduction

Human re-identification aims to match images of the same identity over different non-overlapping camera views, without imposing any constraints on spatial or temporal continuity, or requiring any priori knowledge of the viewing conditions. The only assumption is that the clothing of individuals remains unchanged across different scenarios. It is an important and challenging task for video surveillance applications where human activities are monitored. A mixture of texture and color descriptors is most commonly used in the context of re-identification. Texture descriptors, such as texture filters, and have been successfully applied to address the re-identification problem. Compared to texture descriptors, color information attracts more attention [1, 16, 17], because color inconsistency is the most prominent factor that affects the re-identification performance with respect to viewpoint and pose changes. Histograms of different color models, such as RGB, HSV, Lab are selectively used in different algorithms. Using different weightings, features from different color models are then concatenated to form high dimensional vectors in most re-identification algorithm. Most re-identification algorithms follow this route due to its simplicity. However, such approaches raise two main issues. First, the importance of a certain type of feature in re-identification is mostly based on heuristics. Second, the dimension of the concatenated feature vector is increased as more cues are used. One solution addressing this problem would be reducing the dimensionality of the feature vectors using principal component analysis (PCA). However, features from different color spaces are treated equally without considering the respective magnitude and importance when simple dimensionality reduction methods are used.

In this paper, we carry out extensive experiments to evaluate the performance of eight color descriptors in the context of human re-identification, and learn better metric to measure distance between two observed images. Different from other algorithms, which use each color histogram as part of the feature vector, we regard each color descriptor as a separate ranker. Under a structural learning framework, the ranking scores generated by each ranker are integrated on the decision level in an ensemble model such that the invariance property and distinctive strength from different color spaces can be better exploited.

2. Related work

The main challenging factor in the human re-identification problem is that images belonging to different pedestrians may look more similar than images from the same one, due to viewpoint changes and varying lighting conditions. This problem entails effective ranking algorithms to use the most discriminative information from observed images.

Numerous features have been proposed for human re-identification, e.g., color [17, 16], textures [7], shape [28], patch-based features [32] and attributes [18]. Gray and Tao [9] extract a number of localized features to address the problem of viewpoint changes. Symmetric features from pedestrian images are extracted as features for re-identification by Farenzena *et al*. [7]. In [8] Gheissari *et al*. use color and structural information around extracted key-points as signatures to identify humans. Wang *et al*.

[28] exploit co-occurrence metrics of quantized appearance and shape features for re-identification. Recently Ma *et al.* [19] utilize Fisher vectors to capture higher order statistics of visual features.

Color descriptors have recently attracted much attention in human re-identification. Considering the changes that are mainly caused by different lighting and viewing conditions, great efforts have been made to model the camera-dependent photometric transformations. A number of studies have been proposed to estimate the brightness transfer function (BTF) [22, 2, 12, 13, 6]. Prosser *et al.* [23] propose a cumulative brightness transfer function to make better use of color information to alleviate the requirement of an exhaustive set of training examples. The color names [27, 5, 16] has been proposed to describe the chromatic appearance. Van de Weijer *et al.* [27] propose a learning approach to model color names using a large dataset for image retrieval and classification. On the other hand, D'Angelo and Dugelay [5] collect pixel samples from the uniforms of sport teams and adopt a fuzzy k-nearest neighbor classifier to compute color histograms.

An illumination-invariant color feature model for re-identification is developed by Kviatkovsky *et al.* [17] by exploiting discriminative color distribution to describe invariant properties between two matched images from the same pedestrian. Zhao *et al.* [33, 32] design a color-based salience feature model that is invariant to pose and viewpoint variations and achieve state-of-the-art performance. Recent approaches for re-identification typically use histograms of different color models because of their reliable measure.

3. Color features

We categorize the commonly used histogram-based color features into three categories: photometric color feature, invariance color feature and color names feature. The properties of these features vary from photometric invariance to discriminative strength. Typical features in each category, where each would be used as a separate ranker in the proposed ensemble model, are presented.

Photometric color features. Such features are extracted from color spaces which are specially designed to have different photometric properties. For example, the HSV model accords with visual property of human eyes, and the Lab model is a perceptually uniform color space. In this paper, the RGB, HSV and Lab models are used in the proposed algorithm.

Invariant color features. As the lighting conditions in surveillance scenarios are complex, multi-modal and time-varying, the conventional color models are not effective in accounting for appearance change. Numerous color descriptors have been developed to increase the illumination invariance strength. Table 1 shows three types of basic color changes, causes and representative color features for human re-identification. It is clear that all the three types of light changes commonly occur in real-world surveillance scenarios. Although the transformed color distribution (TCD) is invariant to all the three types of lighting changes, the corresponding distinctiveness is low [26]. In this work, we use four color features based on hue, normalized RGB (NormRGB), opponent (Opp) and color moment (Mom) in the proposed ensemble model.

Color names. Color names [27, 5, 16] are linguistic labels that humans use to describe colors. This descriptor is robust to viewpoint and lighting changes. In [27], a fuzzy distribution is used to describe colors for generating a probability distribution map indicating how a specific color is assigned to each color name. The histogram of color names is usually low dimensional (e.g., 11 dimensions) and effective. Among all the color names, we use the model learned using Google Image by Van de Weijer [27] (referred as CN) as it is learned from a large set of real-world images.

4. Representation and color transformation

In this section, we introduce an effective re-identification method which is be applied to each color feature. A decision level ensemble algorithm is proposed to integrate the ranking results given by all the color rankers.

4.1. Representation

We extract color information from the foreground region based on the results using an max-margin segmentation method [30]. As large granular spatial decomposition is likely to cause misalignment due to pose changes, we follow recent schemes for re-identification [24, 35] and partition an image into six horizontal stripes. For each strip, all histograms based color descriptors mentioned above are extracted. Although the foreground regions describe object appearances more accurately, background pixels also provide useful context information. The color descriptors extracted from both the foreground image and the whole image are concatenated to form a feature vector,

$$\mathbf{x}_i^c = (\mathbf{x}_i^{c,Forg\top}, \mathbf{x}_i^{c,Img\top})^\top, c = 1, \ldots, C \qquad (1)$$

where $\mathbf{x}_i^{c,Forg}$ is the feature extracted from the foreground of image i based on color model λ^c, $\mathbf{x}_i^{c,Img}$ refers to that from the whole image, and C is the total number of the color descriptors. It has been shown that colors in the RGB and HSV spaces are scattered and a small number of bins for histograms perform better [4]. In this work, we use an 8-bin (in each dimension) histogram for the RGB and HSV models, and 32-bin for the Lab model. For the NormRGB, opp, Hue, and CN features, 32, 32, 36, 11 bins

Table 1. Basic color changes category. TCD stands for transformed color distribution and Mom represents color moment. O_1, O_2 are the first two channels in Oppenent model [26] (O_3 does not have any invariant property).

Light change	Cause	Typical color features
light intensity changes	a. light source intensity b. (no colored) shadows and shadings	Hue, NormRGB, TCD
light intensity shift	a. scattering of a white light source b. object hightlight under a white light source c. infrared sensitivity of the camera sensor	Hue, O_1, O_2, Mom, TCD
light color changes	a. illuminant color changes b. light scattering	TCD

histograms are used, respectively. For the Mom feature, we use the generalized color moment descriptor up to the second degree and the first order [21].

4.2. Distance metric learning

Given a color feature vector, a proper metric space needs to be learned to make intra-class and inter-class samples more differentiable. In this work, we use the KISSME metric [15] to compute distance between two feature vectors. For a separate ranker $h_c(\cdot,\cdot)$, the distance $d_c(\cdot,\cdot)$ is assumed to take the form of $\sqrt{h_c(\cdot,\cdot)}$. The ranker h_c is parameterized by a rank-one matrix $\boldsymbol{M}$, where $\boldsymbol{M} \succeq 0$ is a positive semidefinite matrix. For a pair of feature points $(\mathbf{x}_i, \mathbf{x}_j)$, $\mathbf{x}_i, \mathbf{x}_j \in \mathbb{R}^d$, the Mahalanobis distance between a pair of samples is measured by $d_c^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^\top \boldsymbol{M}(\mathbf{x}_i - \mathbf{x}_j)$ So the similarity between the feature points is

$$s_c(\mathbf{x}_i, \mathbf{x}_j) = exp(-\frac{d_c^2(\mathbf{x}_i, \mathbf{x}_j)}{2\sigma^2}) \tag{2}$$

where σ is bandwidth of the Gaussian function. We introduce a similarity label y_{ij}, where $y_{ij} = 1$ indicates the images with the same identity, and $y_{ij} = 0$ otherwise. Given two sets of training samples $\mathcal{S}$ and $\mathcal{D}$, where $\mathcal{S} = \{(\mathbf{x}_i, \mathbf{x}_j)\}_{y_{ij}=1}$ and $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{x}_j)\}_{y_{ij}=0}$, from a statistical inference point of view, the Mahalanobis distance matrix $\boldsymbol{M}$ is defined in closed form $\boldsymbol{M} = \Sigma_{\mathcal{S}}^{-1} - \Sigma_{\mathcal{D}}^{-1}$, where $\Sigma_{\mathcal{S}}$ and $\Sigma_{\mathcal{D}}$ are the covariance matrices of $\mathcal{S}$ and $\mathcal{D}$

$$\Sigma_{\mathcal{S}} = \frac{1}{|\mathcal{S}|} \sum_{(\mathbf{x}_i - \mathbf{x}_j) \in \mathcal{S}} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \tag{3}$$

$$\Sigma_{\mathcal{D}} = \frac{1}{|\mathcal{D}|} \sum_{(\mathbf{x}_i - \mathbf{x}_j) \in \mathcal{D}} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^\top \tag{4}$$

Compared to the large margin nearest neighbour (LMNN) [29] and relative distance comparison (RDC) [35] methods, the KISSME metric [15] performs well with an order of magnitude faster in training time [15].

5. Decision level ensemble ranking

For human re-identification, most methods focus on two key problems: exploiting discriminative and robust features, and learning a subspace or distance metric to cope with cross-view variations. However, the important issue on how to effectively integrate descriptors is not addressed. Existing methods typically concatenate all the feature vectors derived from different cues. While straightforward and commonly applied, such approaches may cause several issues. First, the feature dimension is increased when new cues are introduced, thereby entilas more computational load. Second, low dimensional features tend to be dominated by high dimensional ones. However, different cues may carry complementary informations, so some useful information for re-identification would be lost. To address these problems, we propose a decision level integration algorithm based on an ensemble color model (ECM). In the proposed ensemble model, the appearance affinity is defined by a linear combination of similarity measurements where the weight parameters $\mathbf{w}$ are learned by a structural support vector machine (SVM) to sort the relevant ones with the same identity properly.

5.1. Structural learning

Given two datasets: probe set $\mathcal{D}^p = \{\mathbf{x}_t^p\}_{t=1}^N$ and gallery set $\mathcal{D}^g = \{\mathbf{x}_t^g\}_{t=1}^N$, where t indicates the identity label. For a probe image with identify t, a probe gallery set is denoted as $\{(\mathbf{x}_t^p, \mathbf{x}_{t'}^g)\}_{t'=1}^N$.

We aim to learn $\mathbf{w}$ that order relevant gallery images before irrelevant ones. As we only know the orders between the relevant and irrelevant images, but not orders within relevant or irrelevant ones, the probe gallery set of image with identity t can be considered as a partially ordered set $\mathcal{D}_t^{pg} = (\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N; \mathbf{y}_t^p)$, where the partial order $\mathbf{y}_t^p$ is defined by

$$\mathbf{y}_t^p = \{y_{t,t'}^p\}, \quad y_{t,t'}^p = \begin{cases} +1 & \mathbf{x}_t^g \prec \mathbf{x}_{t'}^g \\ -1 & \mathbf{x}_t^g \succ \mathbf{x}_{t'}^g \end{cases} \tag{5}$$

where $\mathbf{x}_t^g \prec \mathbf{x}_{t'}^g$ represents that $\mathbf{x}_t^g$ is ranked before $\mathbf{x}_{t'}^g$, and after otherwise.

In a n-slack structural SVM model [14], the objective function is defined by

$$\min_{\mathbf{w},\xi} \frac{1}{2}\|\mathbf{w}\|^2 + \frac{\lambda}{N}\sum_{t=1}^{N}\xi_t \tag{6}$$

$$\text{s.t. } \forall \hat{\mathbf{y}}_t^p \in \mathcal{Y}_t^p \backslash \mathbf{y}_t^p :$$

$$\mathbf{w}^\top[\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N, \mathbf{y}_t^p) - \Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N, \hat{\mathbf{y}}_t^p)] \tag{7}$$

$$\geqslant 1 - \frac{\xi_t}{\Delta(\mathbf{y}_t^p, \hat{\mathbf{y}}_t^p)}$$

where λ is a trade-off parameter, $\mathbf{y}_t^p$ is a correct partial order that ranks all correct matches before incorrect ones, $\hat{\mathbf{y}}_t^p$ is an incorrect partial order that violates some of the pairwise relations, and $\mathcal{Y}_t^p$ is space consisting of all possible partial orders. The constraints of (7) state that in each probe gallery set, the score $\mathbf{w}\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N; \mathbf{y}_t^p)$ of correct order $\mathbf{y}$ must be greater than the score $\mathbf{w}\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N; \hat{\mathbf{y}}_t^p)$ of all incorrect orders by a margin, which is determined by a loss function Δ and slack variable ξ_t. As discussed in [31], a good ranking can be obtained simply by sorting gallery images by $\mathbf{w}[\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N, \mathbf{y}_t^p)$ in descending order.

5.2. Learning components

There are three components for learning a structural SVM for the proposed ranking algorithm: feature map Ψ, loss function Δ, and an efficient algorithm for separation oracle [20].

Partial order feature. The feature map $\Psi : \mathcal{X} \times \mathcal{Y} \to \mathrm{R}$ measures the compatibility of the partial order $\mathbf{y}$ in a probe candidate set. We use the commonly used *partial order* feature similar to [20]

$$\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N; \mathbf{y}_t^p) = \sum_{\mathbf{x}_t^g \in \mathcal{X}^+_{\mathbf{x}_t^p}} \sum_{\mathbf{x}_{t'}^g \in \mathcal{X}^-_{\mathbf{x}_t^p}} y^p_{t,t'} \frac{\Phi(\mathbf{x}_t^p, \mathbf{x}_t^g) - \Phi(\mathbf{x}_t^p, \mathbf{x}_{t'}^g)}{|\mathcal{X}^+_{\mathbf{x}_t^p}| \cdot |\mathcal{X}^-_{\mathbf{x}_t^p}|} \tag{8}$$

where $\mathcal{X}^+_{\mathbf{x}_t^p}(\mathcal{X}^-_{\mathbf{x}_t^p})$ denotes the subset of relevant (irrelevant) points in the training set. A feature Φ is defined by the similarity measures generated by different rankers in (2)

$$\Phi(\mathbf{x}_t^p, \mathbf{x}_{t'}^g) = [s_1(\mathbf{x}_t^p, \mathbf{x}_{t'}^g), \ldots, s_C(\mathbf{x}_t^p, \mathbf{x}_{t'}^g)]^\top \tag{9}$$

The partial order feature is suitable for our ranking purpose because it only depends on the difference between relevant and irrelevant pairs, not the entire list. By adding vector difference of correct orders and subtracting that of incorrect orders, the partial order feature emphasizes the directions in feature space which are directly related to correct rankings

AUC loss function. Among all the loss functions commonly used in structural SVMs, the area under curve (AUC) measure is appropriate for the partial order ranking. It characterizes the difference between relevant and irrelevant pairs with only partial order available and can be efficiently calculated by counting the portion of incorrect ordered pairs

$$\begin{aligned} \Delta(\mathbf{y}_t^p, \hat{\mathbf{y}}_t^p) &= |N_{incorrect}| / (|\mathcal{X}^+_{\mathbf{x}_t^p}| \cdot |\mathcal{X}^-_{\mathbf{x}_t^p}|) \\ &= \frac{1}{2}\sum_{t'}(1 - \hat{y}^p_{t,t'}) / (|\mathcal{X}^+_{\mathbf{x}_t^p}| \cdot |\mathcal{X}^-_{\mathbf{x}_t^p}|) \end{aligned} \tag{10}$$

which measures the portion of pairs that are not ranked in correct order.

Separation oracle. When using cutting plan approach to optimize the objective function, one key step is the separation oracle. Given a fixed $\mathbf{w}$, the separation oracle aims to find the most violated output $\tilde{\mathbf{y}}_t^p$

$$\tilde{\mathbf{y}}_t^p \leftarrow \arg\max_{\tilde{\mathbf{y}}_t^p \in \mathcal{Y}} \mathbf{w}^\top\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N; \mathbf{y}_t^p) + \Delta(\mathbf{y}_t^p, \tilde{\mathbf{y}}_t^p)$$

(11)

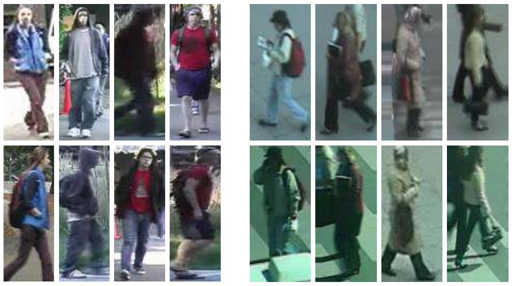
The partial order feature in (8) is attractive as, for a fixed $\mathbf{w}$, the order $\mathbf{y}_t^p$ which maximizes $\mathbf{w}^\top\Psi(\mathbf{x}_t^p, \{\mathbf{x}_{t'}^g\}_{t'=1}^N; \mathbf{y}_t^p)$ is simply sorted by descending $\mathbf{w}^\top\Phi(\mathbf{x}_t^p, \mathbf{x}_{t'}^g)$. Furthermore, as observed by Yue *et al.* [31], optimizing over $\mathbf{y}$ is reduced to finding an optimal interleaving of the relevant and irrelevant sets. So (11) could be efficiently calculated in $O(n \log n)$ time complexity, where $n = |\mathcal{X}^+| + |\mathcal{X}^-|$.

6. Experimental results

To evaluate the performance of the proposed algorithm, we carry out experiments on two challenging datasets: VIPeR [9] and PRID 450S [25]. Each person has one image in either camera view in both datasets. All the quantitative results are reported in form of Cumulated Matching Characteristics (CMC) curves.

6.1. Experimental settings

For fair comparisons, all experiments on both datasets are based on the same evaluation settings [32]. The dataset are randomly partitioned into two parts, half for training and the rest for tests. Images from one camera are treated as probe and those from the other camera are used as gallery. The rank-*k* recognition rate is the expectation of correct matches within rank *k*, and the cumulated values of recognition rate at all ranks is recorded as one-trial CMC result. Each experiment is repeated 10 times and the average results are reported. The feature level fusion (FLF) algorithm is added for comparison. To fairly compare with the proposed ECM algorithm, the dimensionality of the concatenated feature vector is reduced to K via PCA. In all the experiments, we set the dimension K to 70.



(a) (b)

Figure 1. Sample image pairs from VIPeR (a) and PRID 450S (b). The upper and lower rows correspond to different views of the same person, respectively.

6.2. Comparison with State-of-the-Arts

VIPeR Dataset. The VIPeR dataset [1] is arguably the most challenging database for the human re-identification problem due to significant appearance change, drastic illumination difference and large pose variation (See Figure 1 (a)). There are 632 individuals captured in outdoor scenarios with two images from each person (one front/back and one side views). All images are normalized to 128×48 for experiments. In our experiments, we randomly selected half of the dataset for training and the remaining for tests.

Table 2 shows the results of the proposed algorithm with comparisons to two state-of-the-art feature-based algorithms: SDALF [7] and PS [3]. Significant improvements over both methods are achieved by the proposed algorithm. We also compare with seven learning-based methods including PRDC [34], sLDFV [19], CN [16], eSDC-ocsvm [33], KISSME [25], RPLM [10] and salMatch [32]. Overall, the ECM method achieves 38.4% at rank 1, 67.4% at rank 5 and 78.4% at rank 10, with significant improvement over all other methods. Figure 2 shows the performance of the proposed ECM, feature-level fusion (FLF), color rankers (HSV and CN), and salMatch [32] methods. We show the CMC curve of first rank 15 as only the first rank 15 matching results are reported in [33]. Although the HSV ranker performs slightly worse than the salMatch method at rank 1, it achieves comparable or better accuracy from rank 3 (46.6% for HSV ranker and 44.5% for salMatch). Thus the proposed algorithm is effective even using one separate ranker. We also note that the proposed ECM algorithm achieves a 4.46% improvement compared to the FLF method at rank 1.

PRID 450S Dataset. The PRID 450S dataset [2] is a newly constructed dataset which contains 450 single shot person images recorded from two different static cameras. Due to large changes in viewpoint, pose, and illumination, the images of the same persons are significantly different (See Figure 1 (b)). We normalize all the images to 128 × 64 for evaluation.

Table 2. VIPeR dataset: top ranked matching rates in [%] with 316 candidate persons.

Rank	1	5	10	20	50
SDALF [7]	19.9	38.9	49.4	65.7	-
PS [3]	21.8	44.6	57.2	71.2	88.0
PRDC [34]	15.7	38.4	53.9	70.1	-
sLDFV [19]	26.5	56.4	70.9	84.6	-
CN [16]	23.9	45.6	56.2	68.7	-
eSDC-ocsvm [33]	26.7	50.7	62.4	76.4	-
KISSME [25]	27.0	-	70.0	83	95
RPLM [10]	27	-	69 83	83	95
salMatch [32]	30.2	52.3	65.5	-	-
ECM	**38.9**	**67.8**	**78.4**	**88.9**	**96.0**

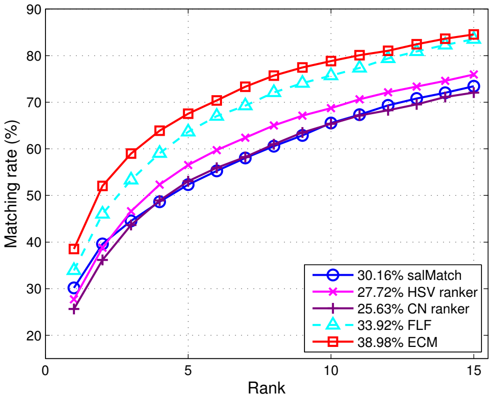


Figure 2. VIPeR dataset: Average CMC curve of the proposed ECM, feature level fusion (FLF), two separate color rankers (HSV and CN) and salMatch [32] methods.

Only a few methods have been evaluated on this dataset, and we compare our algorithm with the best results reported in [25]. The FLF method is also evaluated on this dataset. Table 3 shows that the proposed ECM algorithm performs well against the EIML method [11] (with improvement 6.6% at rank 1). In addition, the proposed algorithm performs wery well against the FLF method especially at low ranks.

Sample re-identification results are shown in Figure 5. Several challenging examples (for which true matches are difficult for humans to distinguish) that are re-identified by the proposed algorithm at rank lower than 3, are shown on the left column, e.g., row 1, 3, 5, 8 in the VIPeR dataset, and row 3, 5, 7 in the PRID 450S dataset. The results can be attributed to the proposed ECM algorithm exploits both color transformation by metric learning and trade-off between invariance and distinctive strength of individual

[1] The VIPeR dataset is available `http://vision.soe.ucsc.edu/?q=node/178`

[2] PRID 450S Dataset is available at: `https://lrs.icg.tugraz.at/download.php`.

Table 3. PRID 450S dataset: top ranked matching rates in [%] with 225 candidate persons.

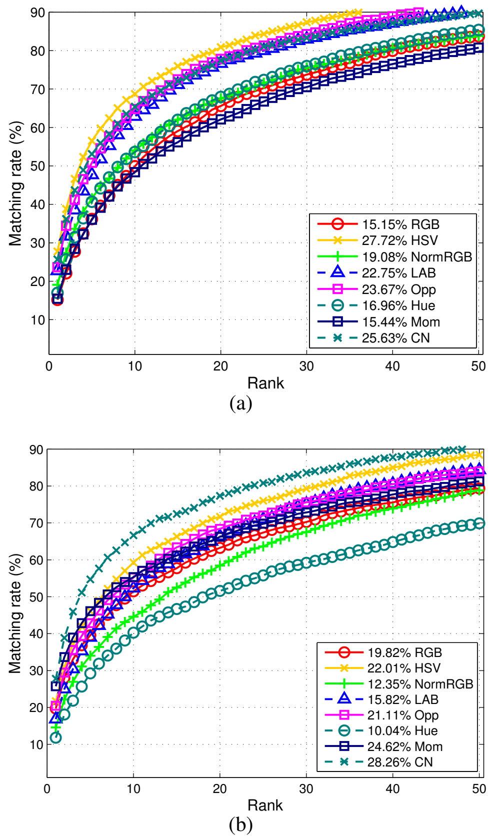
Rank	1	5	10	20	50
KISSME [15]	33.0	-	71.0	79.0	90.0
EIML [11]	35	-	68	77	90
FLF	30.6	60.5	73.6	84.2	93.6
ECM	**41.9**	**66.3**	**76.9**	**84.9**	**94.9**

rankers.

Sample images, in which the true matches rank lower, are shown in the right column of Figure 5. In the VIPeR dataset, appearance changes caused by viewpoint changes are one of the main causes for the failure, especially when one is front view and the other is back view, e.g. row 1, 4 and 5. In the PRID 450S dataset, failures mostly result from heavy overlap between feature distributions of different objects. In such cases, higher level features such as attributes, e.g., gender, accessories, may be exploited for re-identification (when such information can be extracted correctly from images of sufficiently high resolution).

Evaluation of individual ranker and combinations. To evaluate the performance of different rankers, we carry out experiments using individual color features. Performances on both VIPeR and PRID 450S datasets are presented in Figure 3. The re-identification performance of different color rankers are scene specific and no ranker is able to constantly outperform the others. In the VIPeR dataset, the HSV-based ranker performs best, while the one with CN is the best in the PRID 450S dataset. The Mom-based ranker performs differently in these two datasets: 15.44% and 24.62% in the VIPeR and PRID 450S databases. This can be explained by the fact that the Mom features contain rich spatial information and images from two views contain significant changes in space due to viewpoint changes larger than 90 degrees. Overall, the rankers based on HSV, Opp and CN features perform constantly well on both datasets with different lighting conditions. For the Mom features, both O_1 and O_2 are invariant to light intensity shift (O_3 represents intensity information) and thus they are effective to account for appearance change in the PRID 450S dataset.

More visual information can be extracted for re-identification when we combine two color features on the feature or decision level. As there are numerous ways to combine eight rankers, only the representative four are presented in Figure 4. When combining the HSV and CN features on the decision level, significant improvement (6.2%) can be achieved on both datasets (Figure 4(a)), whereas the improvement when features are combined on the feature level is negligible (Figure 4(b)). These results suggest that methods using simple combine on the feature level does not perform better even when each performs well separately. The improvement when RGB and CN features combined are more significant. When features are combined on the feature level, 0.67% improvement is achieved, whereas the gain in performance by the ECM algorithm is 1.58%. On the other hand, when the RGB and NormRGB features are combined, more than 8.3% improvement is achieved on the feature or decision level.



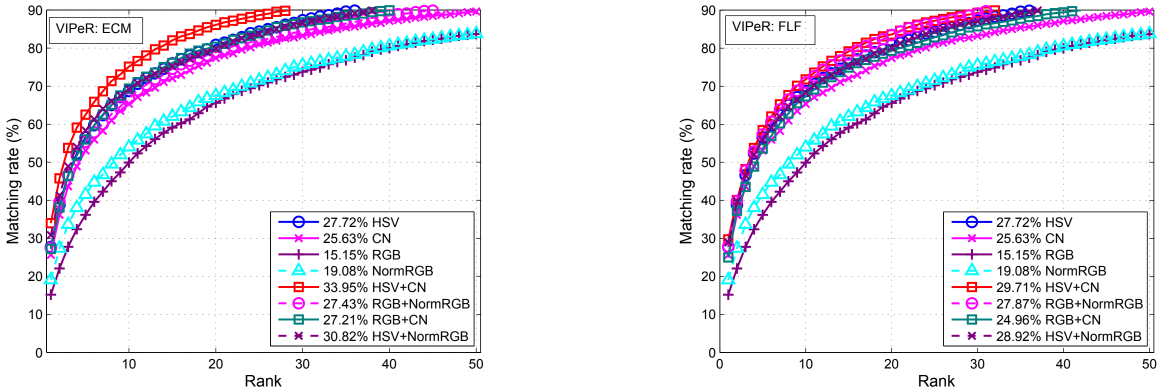
(a)

(b)

Figure 3. The CMC performance comparison of using different color rankers on the VIPeR(a) and PRID 450S datasets (b). Rank-1 recognition rate is marked in front of the ranker name.

7. Conclusion

As color descriptors play an important role in human re-identification, we propose an algorithm to integrate color rankers by exploiting different invariant properties and discriminative strengths of eight color features. For each ranker, the effective KISSME metric learning algorithm is used to account for appearance changes when observed images are acquired at different viewpoints. To address the problems of features with large variation in dimensions and high dimensionality of combined features, we propose a decision level ensemble model using structural support machines, thereby retaining individual distinctiveness and invariant properties. Experimental results on two benchmark datasets demonstrate the proposed algorithm performs favorably against the state-of-the-art methods.



(a)

(b)

Figure 4. CMC curves of using different color combinations on the VIPeR dataset. The performances are tested using two feature fusion strategies: decision level (ECM)(a) and feature level (FLF)(b). For comparison, CMC curves generated by relevant separate rankers are also provided. Rank-1 recognition rate is marked in front of the ranker name.

8. Acknowledgment

X. Liu is sponsored by CSC fellowship. J. Yang and M.-H. Yang are supported in part by the NSF CAREER Grant #1149783 and NSF IIS Grant #1152576. Y. Wu is supported in part by NSFC under Grants #61005027 and #61370036.

References

[1] Y. Cai and M. Pietik. Person re-identification based on global color context. In *ACCV*, 2010.

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

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

[4] A. DAngelo and J.-L. Dugelay. A statistical approach to culture colors distribution in video sensors. In *VPQM*, 2010.

[5] A. DAngelo and J.-L. Dugelay. People re-identification in camera networks based on probabilistic color histograms. In *IS&T/SPIE Electronic Imaging*, 2011.

[6] T. D'Orazio, P. Mazzeo, and P. Spagnolo. Color brightness transfer function evaluation for non overlapping multi-camera tracking. In *ICDSC*, 2009.

[7] M. Farenzena, L. Bazzani, A. Perina, V.Murino, and M. Cristani. Person re-identication by symmetry-rriven accumulation of local features. In *CVPR*, 2010.

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

[9] D. Gray and H. Tao. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In *ECCV*, 2008.

[10] M. Hirzer, P. Roth, M. Köstinger, and H. Bischof. Relaxed pairwise learned metric for person re-identification. In *ECCV*, 2012.

[11] M. Hirzer, P. M. Roth, and H. Bischof. Person re-identification by efficient impostor-based metric learning. In *AVSS*, 2012.

[12] O. Javed, K. Shafique, Z. Rasheed, and M. Shah. Modeling inter-camera spacetime and appearance relationships for tracking across non-overlapping views. *CVIU*, 109(2):146–162, 2008.

[13] K. Jeong and C. Jaynes. Object matching in disjoint cameras using a color transfer approach. *Machine Vision and Applications*, 104(2):154–171, 2008.

[14] T. Joachims, T. Finley, and C.-N. J. Yu. Cutting-plane training of structural SVMs. *Mach. Learn.*, 77(1):27–59, May 2009.

[15] M. Kostinger and M. Hirzer. Large scale metric learning from equivalence constraints. In *CVPR*, 2012.

[16] C.-H. Kuo, S. Khamis, and V. Shet. Person re-identification using semantic color names and RankBoost. In *WACV*, 2013.

[17] I. Kviatkovsky, A. Adam, and E. Rivlin. Color invariants for person reidentification. *PAMI*, 35(7):1622–1634, 2013.

[18] R. Layne, T. Hospedales, and S. Gong. *Attributes-Based Re-identification*. Springer London, 2014.

[19] B. Ma, Y. Su, and F. Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *ECCV*, 2012.

[20] B. Mcfee and G. Lanckriet. Metric Learning to Rank. In *ICML*, pages 775–782, 2010.

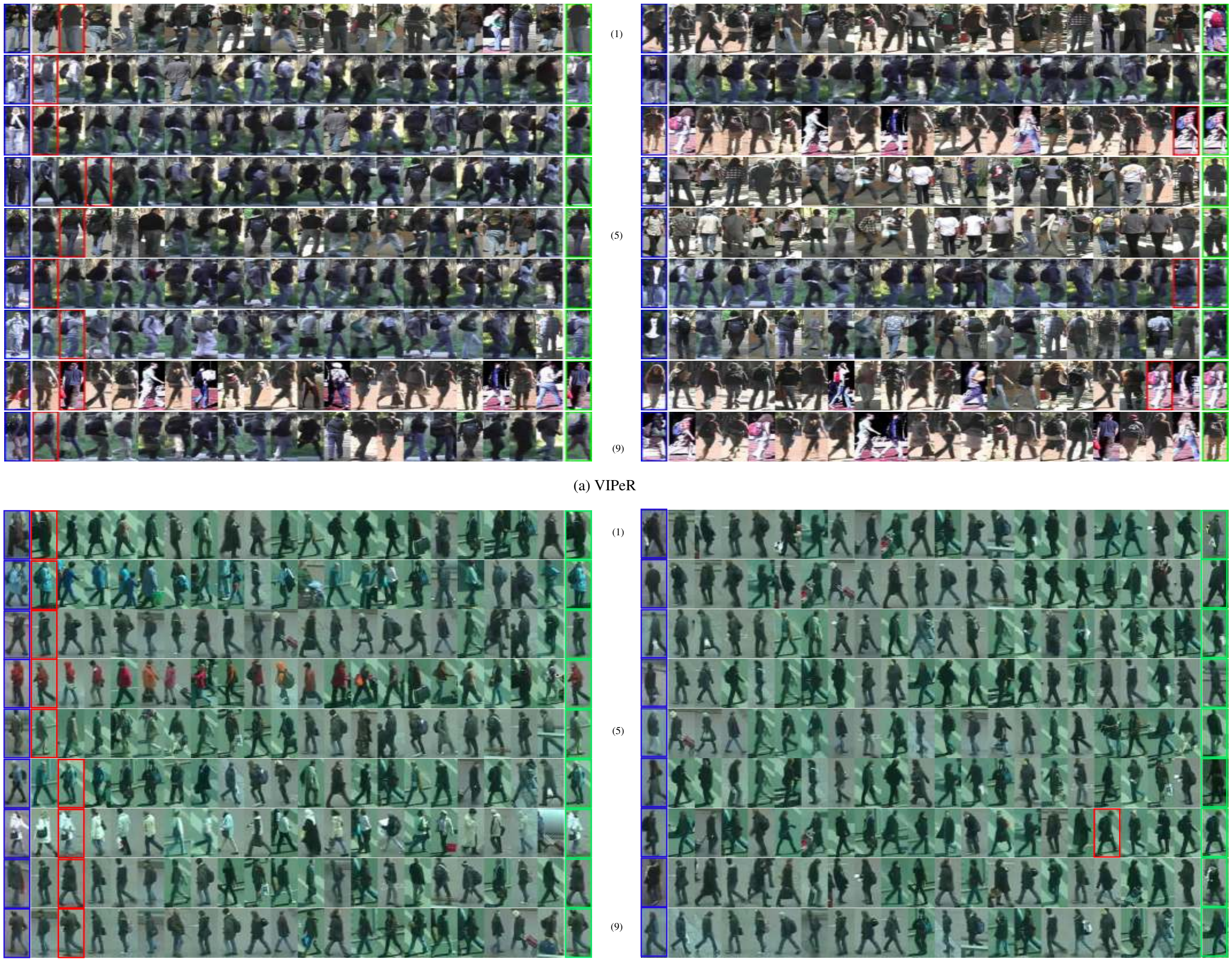
[21] F. Mindru, T. Tuytelaars, L. V. Gool, and T. Moons. Moment invariants for recognition under changing viewpoint and illumination. *CVIU*, 94(1-3):3–27, 2004.

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

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

[24] B. Prosser, W.-S. Zheng, S. Gong, and T. Xiang. Person re-Identification by support vector ranking. In *BMVC*, 2010.

[25] P. M. Roth, M. Hirzer, M. Köstinger, C. Beleznai, and H. Bischof. *Mahalanobis Distance Learning for Person Re-identification*. Springer London, 2014.



(a) VIPeR

(b) PRID 450S

Figure 5. Re-identification results on the (a) VIPeR and (b) PRID 450S datasets. Challenging samples (i.e., true matches are difficult for humans to distinguish) which can be re-identified within rank 3 are shown in the left columns. Sample failures (which the true matches rank lower) are present in the right columns. Probe images, correct matches, and ground truth images are labeled with blue, red, and green boxes, respectively. Row numbers are shown in the middle for clear presentation.

[26] K. E. a. van de Sande, T. Gevers, and C. G. M. Snoek. Evaluating color descriptors for object and scene recognition. *PAMI*, 32(9):1582–1596, 2010.

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

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

[29] K. Weinberger, J. Blitzer, and L. Saul. Distance metric learning for large margin nearest neighbor classification. In *NIPS*, 2006.

[30] J. Yang, S. Simon, and M.-H. Yang. Max-margin Boltzmann machines for object segmentation. In *CVPR*, 2014.

[31] Y. Yue, T. Finley, F. Radlinski, and T. Joachims. A support vector method for optimizing average precision. In *SIGIR*, 2007.

[32] R. Zhao, W. Ouyang, and X. Wang. Person re-identification by salience matching. In *ICCV*, 2013.

[33] R. Zhao, W. Ouyang, and X. Wang. Unsupervised salience learning for person re-identification. In *CVPR*, 2013.

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

[35] W.-S. Zheng, S. Gong, and T. Xiang. Re-identification by relative distance comparison. *PAMI*, 35(3):653–668, 2013.