Color Models and Weighted Covariance Estimation for Person Re-Identification

Yang Yang, Shengcai Liao, Zhen Lei, Dong Yi, Stan Z. Li[*]

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition

Institute of Automation, Chinese Academy of Sciences (CASIA)

Abstract—Due to illumination changes, partial occlusions, and object scale differences, person re-identification over disjoint camera views becomes a challenging problem. To address this problem, a variety of image representations have been put forward. In this paper, the illumination invariance and distinctiveness of different color models including the proposed color model are firstly evaluated. Since color distribution is robust to image scales and partial occlusions, color distributions based on different color models are then calculated and fused in the stage of feature extraction. Different color models obtain robustness to different types of illumination and thus fusing them can compensate each other and contribute to better performance. In the stage of feature matching, a weighted KISSME is presented to learn a better distance metric than the original KISSME. Experimental results demonstrate its feasibility and effectiveness. Finally, image pairs are matched based on the learned distance metric. Experiments conducted on two public benchmark datasets (VIPeR and PRID 450S) show that the proposed algorithm outperforms the state-of-the-art methods.

Keywords—person re-identification, illumination invariance, color models, metric learning.

I. Introduction

Nowadays, person re-identification has attracted much attention from researchers due to its various promising potential applications in video surveillance. It refers to the task of matching the same persons observed from disjoint camera views. Thus, it can be applied to many cases, such as tracking a person from one camera to another non-overlapping camera, human retrieval, and threat detection [1]. Nevertheless, because of the illumination changes, viewpoint variations, partial occlusions, object scale differences and low resolution, person re-identification is a very challenging task.

Color distribution has been widely used in the area of object recognition [3, 4, 5, 6, 7, 8, 9, 18, 21] and shows good performance. There are many approaches to obtain the color distribution, such as color histogram [3, 4, 6, 7], bag-of-features representation [2, 14] and other coding method [15]. Color histogram, which is used in this paper, are robust to scale variation and partial occlusion [5]. However, changes in the illumination of real-word scenes can greatly affect their performance. Fig. 1 shows several examples from the dataset VIPeR [12]. It can be seen that image colors of the same person are quite different from camera A and camera B. This is mainly caused by the illumination variations.

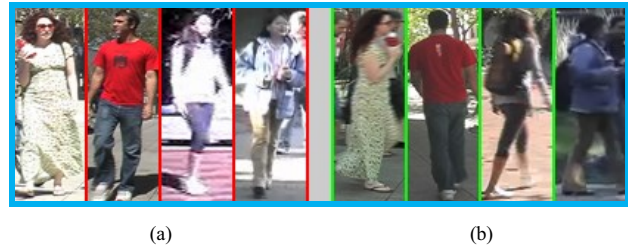Various color models have been employed in the area of the

Fig. 1. Illustration of illumination changes in the dataset VIPeR. (a) Images from camera A. (b) Images from camera B.

object recognition [4, 9] to increase illumination invariance. According to [4], based on the diagonal-offset model, common changes in the image values caused by illumination can be classified as light intensity change, light intensity shift, light intensity change and shift, light color change, and light color change and shift. Different color models are designed to solve these five types of illumination changes. For example, *rgb* (the normalized *RGB* color model) [4, 6, 9], which may be the simplest one of color models is invariant to light intensity change, but is sensitive to the other four types of illumination changes. It is recognized that no single color model has the characteristic of robustness against all the types of illumination changes. Furthermore, the color model which is invariant to more types of illumination changes does not necessarily obtain better results for object recognition [9]. This is due to the fact that the performance of the color models for object recognition does not only depend on their robustness against illumination changes, but also on their discriminative power.

However, there are few reports on comparing the features extracted in different color space for person re-identification. This paper analyzes the invariant properties of the different color models which include *RGB*, *rgb*, $m_1m_2m_3$, $l_1l_2l_3$ and *Hue*. After that, a novel color model named $g_1g_2g_3$ is proposed to increase illumination invariance and distinctiveness. All of them are evaluated on two public benchmark datasets (VIPeR [12] and PRID450S [13]). Extensive experiments demonstrate the usefulness of the invariance and distinctiveness of the proposed color models. In addition, it can also be found that when illumination is controlled, those color models which are invariant to illumination changes does not always perform better than the original *RGB* color model. As a consequence, there should be a tradeoff between invariance and distinctiveness. For the real applications of person re-identification in certain image conditions, these color models, which are less invariant but owning higher discriminative power may gain better results.

---

* Corresponding Author.

IEEE computer society

Since different color models compensate each other, it can be expected that fusing them is a good choice [2, 7, 8, 10, 11] and should be better than a single color model. Reference [2] uses a histogram of *RGB*, *HSV* and *YCbCr* as a color feature. *HSV* and *Lab* histograms are extracted to describe color cues in [7, 8, 11]. In [10], 12 color channels including *R*, *G*, *B*, *r*, *g*, *RG*, *YB*, *L*, *a*, *b*, *S* and *V* are selected and fused for image feature detection. Although there are a large number of existing color spaces such as *CMY*, *Lab*, *Luv*, *UVW*, *XYZ*, *YIQ* [9], some of them are correlated with each other. Therefore, not all the existing color models are included in the fusion process. Once all the features are obtained and concatenated to form the final feature, the principal component analysis (PCA) is applied to reduce the dimension of the features.

In the stage of feature matching, a Mahalanobis metric is learned from the similar and dissimilar pairs of features in the training set. Reference [7] presents a simple and efficient metric learning method, named KISSME. Compared to other approaches, one of the biggest advantages is that it does not need an iterative optimization procedure. However, it is not an optimal way to compute the covariance matrix proposed in [7] when the pairs of features are insufficient. In this work, an easy and effective algorithm is presented to calculate the covariance matrix, which is inspired by expectation-maximization (EM) based density estimation. Experimental results demonstrate the feasibility and validity of the proposed strategy.

Contributions of this paper are summarized as follows: 1) the illumination invariance and distinctiveness of several existing color modes are evaluated for person re-identification and some useful conclusions are made; 2) a novel and proper color model is presented. Even though it does not show the best performance among the existing color models, it is better than original RGB color model in the image condition of various illumination changes and better results can be obtained when it is fused with other color models; 3) the original KISSME is improved and experimental results demonstrate that the weighted KISSME algorithm can learn a better Mahalanobis metric than the original KISSME; 4) the proposed method is fast, stable and easy to implement. Though only color information is employed, some promising results are obtained on the public datasets (VIPeR and PRID 450S).

II. COLOR MODELS

In this section, some of the existing color models are firstly reviewed and then a novel color model ($g1g2g3$) is presented. To measure the illumination invariance, reference [16] adopts a diagonal-offset model which uses a diagonal matrix to map colors under illumination o to the corresponding colors under illumination c with an offset $(o_1, o_2, o_3)^T$:

$$\begin{pmatrix} R^c \\ G^c \\ B^c \end{pmatrix} = \begin{pmatrix} a & 0 & 0 \\ 0 & b & 0 \\ 0 & 0 & c \end{pmatrix} \begin{pmatrix} R^o \\ G^o \\ B^o \end{pmatrix} + \begin{pmatrix} o_1 \\ o_2 \\ o_3 \end{pmatrix}. \qquad (1)$$

Based on (1), five types of illumination changes are defined in [4]. For example, if $a = b = c$ and the offset is a null vector, then this is a condition of light intensity change. Table I shows all conditions of five types of illumination changes.

	(a, b, c)	$(o_1, o_2, o_3)^T$
Light Intensity Change	$a = b = c$	null
Light Intensity Shift	$a = b = c = 1$	$o_1 = o_2 = o_3$
Light Intensity Change and Shift	$a = b = c$	$o_1 = o_2 = o_3$
Light Color Change	$a \neq b \neq c$	null
Light Color Change and Shift	$a \neq b \neq c$	$o_1 \neq o_2 \neq o_3$

A. Analysis of existing color models

Although there exist a great quantity of well-known color models which, as is shown in [9], include *RGB*, *CMY*, *YIQ*, *YUV*, *XYZ*, $I_1I_2I_3$, *rgb*, *xyz*, *UVW*, *Lab*, *Luv*, *HIS*, some of them are correlated to each other. Thus, similar to [4, 9], this paper only concentrates on the following color models:

- **rgb.** It is a normalized *RGB* color model:

$$\begin{pmatrix} r \\ g \\ b \end{pmatrix} = \begin{pmatrix} R/(R+G+B) \\ G/(R+G+B) \\ B/(R+G+B) \end{pmatrix}. \qquad (2)$$

If $a = b = c$ and the offset is a null vector in (1), then

$$\left(r^c, g^c, b^c\right)^T = \left(r^o, g^o, b^o\right)^T, \qquad (3)$$

can be obtained. Thus, it is invariant to light intensity change.

- **l1l2l3.** It is proposed in [9], which solely determines the direction of the triangular color plane in RGB color space. It is defined as follows:

$$\begin{pmatrix} l_1 \\ l_2 \\ l_3 \end{pmatrix} = \begin{pmatrix} (R-G)^2 / ((R-G)^2 + (R-B)^2 + (G-B)^2) \\ (R-B)^2 / ((R-G)^2 + (R-B)^2 + (G-B)^2) \\ (G-B)^2 / ((R-G)^2 + (R-B)^2 + (G-B)^2) \end{pmatrix}. \qquad (4)$$

It can be found that when $a = b = c$ and $o_1 = o_2 = o_3$, the following equality holds well:

$$\left(l_1^c, l_2^c, l_3^c\right)^T = \left(l_1^o, l_2^o, l_3^o\right)^T. \qquad (5)$$

Therefore, the *l1l2l3* color model is invariant to light intensity change, light intensity shift and light intensity change and shift.

- **Hue.** In [9], the color model *Hue* is defined as

$$H = \arctan(\sqrt{3}(G-B)/(2R-G-B)). \qquad (6)$$

If $a = b = c$ and $o_1 = o_2 = o_3$, then (7) can be obtained:

$$H^c = H^o. \qquad (7)$$

It can be seen that the performance of *Hue* is similar with that of *l1l2l3* color model. Specifically, *Hue* is invariant to light intensity change, light intensity shift and light intensity change and shift.

- **m1m2m3.** Reference [9] also defined the *m*1*m*2*m*3 color model, which is based on the color ratio between two neighboring pixels:

$$\begin{pmatrix} m_1 \\ m_2 \\ m_3 \end{pmatrix} = \begin{pmatrix} (R^{x_1}G^{x_2})/(R^{x_2}G^{x_1}) \\ (R^{x_1}B^{x_2})/(R^{x_2}B^{x_1}) \\ (B^{x_1}G^{x_2})/(B^{x_2}G^{x_1}) \end{pmatrix}. \quad (8)$$

where x_1 and x_2 denote the locations of two neighboring pixels. Suppose that the color of illumination is constant over the two neighboring locations, which means the diagonal matrix in (1) is the same for two neighboring pixels, so (9) can be achieved as long as $o_1 = o_2 = o_3$.

$$\left(m_1^c, m_2^c, m_3^c \right)^T = \left(m_1^o, m_2^o, m_3^o \right)^T. \quad (9)$$

Consequently, the *m*1*m*2*m*3 color model is invariant to light intensity change, light intensity shift, light intensity change and shift and light color change.

B. Proposed g1g2g3 color model

Motivated by the log-chromaticity color space, which is proposed in [17] and adopted in [6], a novel color mode, named g*1*g*2*g*3* mode is presented. It is defined as (10):

$$\begin{pmatrix} g_1 \\ g_2 \\ g_3 \end{pmatrix} = \begin{pmatrix} \ln(R/G) \\ \ln(G/B) \\ \ln(B/R) \end{pmatrix}. \quad (10)$$

According to (1), if $a = b = c$ and $o_1 = o_2 = o_3 = 0$, then,

$$\left(g_1^c, g_2^c, g_3^c \right)^T = \left(g_1^o, g_2^o, g_3^o \right)^T. \quad (11)$$

Hence, similar to *rgb*, the proposed *g*1*g*2*g*3 model is only invariant to light intensity change. It can be seen that the relationship between *R*, *G* and *B* is preserved in *g*1*g*2*g*3 model. For example, if *R* is smaller than *G*, then g_1 is negative, and vice versa. All negative numbers are set to zero. It should also be noted that the sum of g_1 , g_2 and g_3 equals to zero. Such relationship can be regarded as some constraint for *g*1*g*2*g*3 model. Similar constraint can also be in *rgb* model that the sum of *r*, *g* and *b* is 1. The difference between *rgb* and *g*1*g*2*g*3 lies in the fact that in *rgb* each channel is divided by a same value while in *g*1*g*2*g*3 the difference between every two of *R*, *G* and *B* is compared. Thus, *rgb* and *g*1*g*2*g*3 should have different distinctiveness. Fig.2 (b) and (c) also compare the color distribution based on *rgb* and *g*1*g*2*g*3 color models respectively. It is obvious that their distributions are different and thus they should have different performance. Experimental results find that they are not highly correlated and fusing them can achieve better results than a single color model of them.

In conclusion, different color models are reviewed and presented, which are designed to deal with the illumination problem. According to the diagonal-offset model introduced by (1), their illumination invariant properties are also analyzed. Because of distinctiveness, although the *Hue* and *l*1*l*2*l*3 have
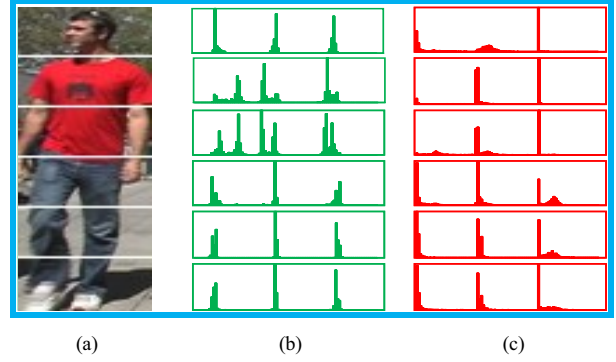

(a)　　　　　　(b)　　　　　　(c)

Fig. 2. An example of part-based body model. (a) Divide an image into six parts. (b) Color distribution of each part based on the normalized color histogram (*rgb* color model). (c) Color distribution of each part based on the normalized color histogram (*g*1*g*2*g*3 color model).

similar invariance, the features based on them may have different performance. For example, in (4), *l*1*l*2*l*3 totally depends on the absolute difference between *R*, *G*, and *B*, and thus may have less distinctiveness than *Hue* which is determined by their relative distance. Similar to *rgb* model which preserves the rank ordering of *R*, *G* and *B*, the distinctiveness of *g*1*g*2*g*3 model can be assured by the fact that the rank ordering of *R*, *G* and *B* is also preserved in *g*1, *g*2 and *g*3. In the experimental section, a comparison is further made among these color models.

III. FEATURE EXTRACTION

Different color models have been utilized to tackle with, at least partially, the illumination problems. In this section, how to extract features based on these color models is the main task.

A. Part-based body model

Color distribution has been widely employed as a feature to represent the whole body appearance in the area of person re-identification. Color histogram may be the mostly used because it is a simple yet effective approach and invariant to object scale and partial occlusions. In this paper, color histogram is also adopted to represent the images.

Since the human body is not rigid, it is better to choose a part-based model instead of treating the body as a whole. Different mechanisms have been exploited [5, 6, 18]. Reference [5] divides them into fixed models, adaptive models and learned models. More accurate model implies more time and efforts. For simplicity, a fixed model is adopted. Specifically, an image is divided into six horizontal stripes of equal size. Fig. 2 (a) gives an example of division. It can be seen that six parts including the head, upper and lower torso, upper and lower legs and the feet are roughly captured.

Assume that $\mathbf{H}_i$, *i*= 1,2,...,6 denotes the color histogram of the *i*-th part of an image, then for simplicity, they are normalized by

$$\mathbf{H}_i^o = \mathbf{H}_i / \mathbf{H}_{i\max} , \quad i = 1,2,...,6, \quad (12)$$

where $\mathbf{H}_{i\max}$ refers to the maximum value of $\mathbf{H}_i$. The benefits

are twofold: 1) the maximum in $\mathbf{H}_i^o$ is 1 for each part and the relative magnitudes of elements in $\mathbf{H}_i^o$ are preserved; 2) the effect of the number of pixels in each part can be eliminated. This step is of vital importance, especially when the images cannot be divided equally or the mask information is used. Fig. 2 (b) and (c) shows the color distribution of each part based on (12). Different distributions can be found between the torso and legs. Finally, all the six color histograms are concatenated into a vector to represent the whole body.

B. Fusion of features

Since different features have been obtained on the basis of different color models, a combination technique is required to fuse them. There are mainly two methods [5]: 1) feature-level fusion (e.g. concatenating them [6, 7, 8, 11]) and 2) score-level fusion (e.g. maxpooling[2], sum of weighted features[10]). In this paper, feature-level fusion is employed, i.e. different features are concatenated into the final feature.

IV. FEATURE MATCHING

In person re-identification, the task is to find a person image selected from the probe set in all the images from the gallery set. Recently, a great deal of attention has been paid to metric learning to search for the most similar pairs [7, 11]. Numerous experimental results have shown good performance of Mahalanobis distance learning. Specifically, given a pair of data points x_i and x_j, its goal is to learn the metric matrix $\mathbf{M}$ in

$$d_{\mathbf{M}}^2(\mathbf{x}_i, \mathbf{x}_j) = (\mathbf{x}_i - \mathbf{x}_j)^T \mathbf{M}(\mathbf{x}_i - \mathbf{x}_j), \qquad (13)$$

where $\mathbf{M}$ is a positive semi-definite matrix. When the metric matrix is learned from the labeled image pairs, the distances between the probe image and all the gallery images are calculated using (13). The gallery image with smallest distance is recognized as the similar pair with the probe image.

Thus, the following mainly concerns how to learn the metric matrix $\mathbf{M}$ in (13) with the labeled image pairs.

A. KISSME

Motivated by a statistical inference perspective based on a likelihood-ratio test, reference [7] offers a KISS metric learning. In [7], whether a pair is similar or not can be determined by a likelihood ratio test. Then, the learned matrix $\mathbf{M}$ is:

$$\mathbf{M} = (\Sigma_{y_{ij}=1}^{-1} - \Sigma_{y_{ij}=0}^{-1}), \qquad (14)$$

where $y_{ij=1}$ denotes the similar pair, $y_{ij=0}$ refers to dissimilar pair and,

$$\Sigma_{y_{ij}=1} = \sum_{y_{ij}=1}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \qquad (15)$$

$$\Sigma_{y_{ij}=0} = \sum_{y_{ij}=0}(\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T. \qquad (16)$$

The most advantage of KISSME is that it does not depend on an iterative optimization procedure, which meets the KISS principle (keep it simple and straightforward).

B. Weighted KISSME

Extensive experimental results [7, 8] have demonstrated the effectiveness of metric matrix learned by KISSME. However, when the similar and dissimilar pairs are insufficient, it may not be an optimal way to calculate the covariance matrix $\Sigma_{y_{ij}=0}$ and $\Sigma_{y_{ij}=1}$. A novel method is presented to compute them in this section.

Since $\Sigma_{y_{ij}=0}$ and $\Sigma_{y_{ij}=1}$ denote the covariance matrix of similar pairs and dissimilar pairs respectively, it can be estimated more accurately by an expectation-maximization (EM) [19] algorithm when the pairs are insufficient. To be specific, for the pairwise differences $\mathbf{x}_{ij}$, which means $\mathbf{x}_i - \mathbf{x}_j$ are symmetric, the mean of all pairwise differences is zero. Then, considering the j-th iteration, the covariance matrix is measured by (17) in maximization (M) step,

$$\Sigma_j = \sum_{i=1}^m \omega_j^i \mathbf{x}_{ij} \mathbf{x}_{ij}^T \Big/ \sum_{i=1}^m \omega_j^i ; \qquad (17)$$

where ω_j^i is estimated by the posteriori probability of latent variables label in expectation (E) step.

However, although EM can estimate the covariance matrix more accurately, the tedious iteration step is time-consuming. In accordance with the KISS principle, the posteriori probability of latent variables label in E step is assumed to be a Gaussian density function with zero mean, then the covariance matrix is

$$\Sigma = \sum_{i=1}^m \omega^i \mathbf{x}_{ij} \mathbf{x}_{ij}^T \Big/ \sum_{i=1}^m \omega^i, \qquad (18)$$

where $\omega^i = \exp(-\text{dis}(\mathbf{x}_{ij}, \mathbf{\mu})/\sigma^2)$, $\text{dis}(\mathbf{x}_{ij}, \mathbf{\mu})$ denotes the Euclidean distance between $\mathbf{x}_{ij}$ and the mean $\mathbf{\mu}$ which is zero here. σ is predefined. ω^i can be seen as how much the contribution of $\mathbf{x}_{ij}$ to the final covariance matrix, i.e., the farther $\mathbf{x}_{ij}$ is from the zero mean measured by the Euclidean distance, the less contribution $\mathbf{x}_{ij}$ makes to the final covariance matrix. Consequently, it can be seen as a weighted covariance matrix. ω^i is the weight lowering the contribution of those that are farther from the zero mean. σ is used to adjust weight decay speed. Then, a weighted KISS metric can be learned by $\Sigma_{y_{ij}=0}$ and $\Sigma_{y_{ij}=1}$ which are obtained by (18). This metric learning method is called weighted KISSME in this paper.

V. EXPERIMENTS

In this section, two public benchmark datasets: VIPeR and PRID 450S are used for evaluation. They are shown in Fig. 3. Different color models are firstly compared in two conditions: extracted features with and without mask. Mask is automatically obtained by the method [21] with the parameter settings used in [18]. Then the proposed weighted KISSME is also compared with original KISSME. The results are shown in form of Cumulated Matching Characteristic (CMC) curve [20]. Finally, they are compared with the results achieved by state-of-the-art approaches.
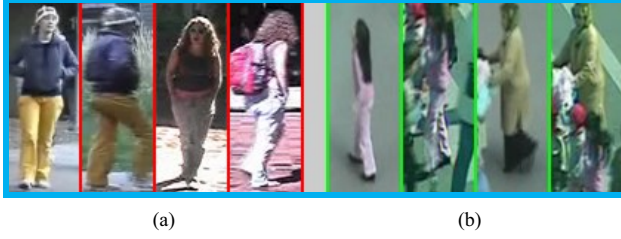
Fig. 3. Examples from the datasets：VIPeR (a) and PRID 450S (b).

A. Datasets and settings

Datasets: Two public benchmark datasets are selected:

1) VIPeR. This is a commonly used dataset for person re-identification, which contains 632 image pairs observed from two different camera views. Each person has one image pair. It is one of the most challenging person re-identification dataset because of illumination changes and viewpoint variations. As is shown in Fig.1, the clothing color of the same person looks quite different due to the illumination changes.

2) PRID 450S. It contains 450 image pairs observed from two different and static cameras. Each person has one image pair. Even though the illumination changes of this dataset are not as severe as that of VIPeR, it is also a challenging person re-identification dataset due to the background interference, partial occlusion and viewpoint changes.

Experiment settings: For both datasets, half of the overall image pairs, i.e. 316 image pairs for VIPeR and 225 image pairs for PRID 450S, are randomly selected for training and the rest is for test. During the test, images from one camera are regarded as probe and the others as gallery, and then the probe and gallery are switched. The average of their CMC curves is taken as one trial. As in [7, 8, 11], 100 trials of evaluation are repeated to achieve a statistical significance. Then, the average of the first 50 ranks are reported. Principal component analysis (PCA) is applied to reduce the computational effort. In all experiments, σ is set to 0.4 and 0.8 respectively for $\Sigma_{y_{ij}=0}$ and $\Sigma_{y_{ij}=1}$. Unless otherwise specified, the number of bins of each channel in different color models is set to 32. Resizing the images to the same scale is unnecessary.

B. Evaluation of different color models

In this section, features based on different color models including *RGB*, *rgb*, *m1m2m3*, *l1l2l3*, *Hue* and *g1g2g3* are evaluated on both datasets. As in [7], the dimension of them is reduced to 34. Weighted KISSME is used to calculate the covariance matrix. Experimental results are shown in Fig. 4.

Fig. 4 (a) compares the performance of different color models on VIPeR. In general, features based on the *rgb*, *g1g2g3* and *Hue* color models are better than on the *RGB* color model. However, in Fig. 4 (b), features based on the *RGB* color model using the mask yield the best results on the PRID 450S dataset. This may because PRID 450S dataset contains less illumination variations than VIPeR does, then the illumination invariant color models on PRID 450S are less effective than on VIPeR. Due to the fact that there are many background or other
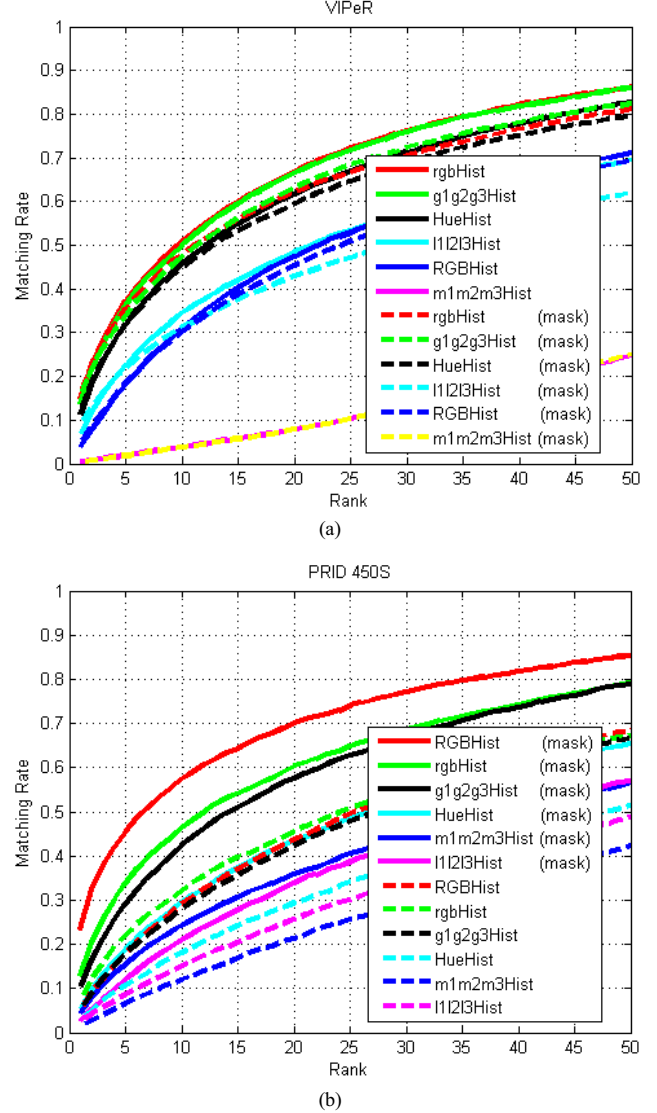

(a)

(b)

Fig. 4. Comparison of features based on different color models. (a) Features using mask and without mask on VIPeR. (b). Features using mask and without mask on PRID 450S.

interference factors in PRID 450S, features using mask are better than those without mask. It should also be noted that the performance of *l1l2l3* and *m1m2m3* color models is poor on both datasets, even though they are more robust to illumination changes. This demonstrates that features does not only require invariance but also distinctiveness. Thus, there should be a balance between them.

C. Fusion and comparison to the state-of-the-art resultss

Different features are fused in this section and the results based on weighted KISSME and KISSME are compared on both datasets. In experiments, it is found that features including *l1l2l3* and *m1m2m3* color models do not improve the results and that when S and V of *HSV* color space as well as *YCbCr* color space are fused, the performance will be enhanced. Thus,

TABLE II. VIPeR DATASETS: TOP RANKED MATCHING RAETES (%)

Method	R=1	R=5	R=20	R=25	R=30	R=50
EIML [11]	22	-	78	-	-	93
ALL [8]	11.7	38.0	71.5	78.2	82.0	91.3
SCEFA [23]	26.5	49.8	73.5	-	-	-
Comb [6]	24.2	44.9	69.4	73.8	77.2	87.4
WBTF [3]	22.0	46.8	75.6	-	-	-
KISSME [7]	19.6	-	-	80.7	-	91.8
Salience [1]	30.2	52.3	-	-	-	-
Fusion + KISSME	**33.1**	**65.6**	**89.0**	**91.2**	**93.3**	**95.9**
Fusion + Weighted	**37.6**	**68.1**	**90.2**	**92.4**	**94.0**	**96.3**

TABLE III. PRID 450S DATASETS: TOP RANKED MATCHING RAETES (%)

Method	R=1	R=10	R=20	R=25	R=30	R=50
MDL [22]	29	62	73	-	-	86
MDL+Seg [22]*	35	68	77	-	-	90
Fusion + KISSME	26.6	**70.0**	**82.3**	86.0	87.8	92.9
Fusion + Weighted	**35.0**	**73.6**	**85.8**	88.1	89.5	93.8
Ours + Seg*	**40.6**	**80.7**	**90.0**	91.3	92.4	96.0

a. *Seg refers to have used the manually segmented masks

those features using mask and without mask are fused, which are based on *RGB*, *rgb*, *g1g2g*3, *HSV* (*Hue*, *S* and *V*) and *YCbCr* color space. The number of bins of each channel in *HSV* and *YCbCr* are set to 16. The dimensionality of fused features is reduced to 60 by PCA.

As is shown in Table II and III, features based on weighted KISSME show better performance than that on KISSME for both datasets. Note that there is a small fluctuation for each average result of 100 trials. From table II, it can be seen that the proposed features based on KISSME achieves significantly better results than [7] and [8], both of which are also based on KISSME for the datasets VIPeR. In Table III, to fairly compare with "MDL + Seg" in [22], the manually segmented masks are also used (named "Ours + Seg"). Apparently, the proposed method in this paper achieves new state-of-the-art results on VIPeR and PRID 450S at rank 1.

VI. CONCLUSION.

In this paper, person re-identification is formulated as a color distribution matching problem. In order to deal with the illumination problems, the color invariance and discriminative power of different color models are evaluated and a novel color model is then proposed to increase illumination invariance and distinctiveness. However, none of these color models is robust to all illumination changes. Consequently, to compensate each other, color histograms based on those color models which have both invariance and distinctiveness are fused to form the final feature. There are two advantages of this approach: 1) there is no need to resize the images and 2) it is very fast to implement. To match the features, Mahalanobis distance, whose metric matrix is learned by the weighted KISSME is adopted. Simple though the proposed approach is, experimental results demonstrate its feasibility and validity and greatly boost the performance of person re-identification.

REFERENCES

[1] R. Zhao, W. L. Ouyang, and X. G. Wang, "Person re-identification by salience matching," in ICCV, 2013.

[2] X. R. Zhang, Y. Yang, L. C. Jiao, and F. Dong, "Manifold-constrained coding and sparse representation for human action recognition," Pattern Recognition, 2013.

[3] A. Datta, L. M. Brown, R. Feris, and S. Pankanti, "Appearance modeling for person re-identification using weighted brightness transfer function," in ICPR, 2012.

[4] K. E. A. Van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluating color descriptors for object and scene recognition," PAMI, 2010.

[5] R. Satta, "Appearance descriptors for person re-identification: a comprehensive review," in CoRR, 2013.

[6] I. Kviatkovsky, A. Adam, and E.Rivlin, "Color invariants for person re-identification," PAMI, 2013.

[7] M. Köstinger, M. Hirzer, P. Wohlhart, P. M. Roth, and H. Bischof, "Large scale metric learning from equivalence constraints," in CVPR 2012.

[8] Z. Yang, L.W. Jin, and D. P. Tao, "A comparative study of several feature extraction methods for person re-identification," in CCBR, 2013.

[9] T. Gevers and A. W . Smeulders, "Color-based object recognition," Pattern Recognition, 1999.

[10] H. Stokman and T. Gevers, "Selection and fusion of color models for image feature detection," PAMI, 2007.

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

[12] D. Gray, S. Brennan, and H. Tao, "Evaluating apperance models for recognition, reacquisition and tracking," in PETS, 2007.

[13] P. M. Roth, M. Hirzer, M. Kostinger, C. Beleznai, and H. Bischof, "Evaluation of mahalanobis distances for single-shot person re-identification," in ICG-TR, 2012.

[14] G. Doretto, T. Sebastian, P. Tu, and J. Rittscher, "Appearance-based person reidentification in camera networks: problem overview and current approaches," JAIHC, 2011.

[15] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in CVPR, 2009.

[16] G. D. Finlayson, S. D. Hordley, and R. Xu, "Convex programming colour constancy with a diagonal-offset Model," in ICIP, 2005.

[17] D. Berwick and S. Lee, "A chromaticity space for specularity, illumination color- and illumination pose-invariant 3-D object recognition," in ICCV, 1998.

[18] M. Farenzena, L. Bazzani, A. Perina, V. Murino and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in CVPR, 2010.

[19] M. Collins, "The EM algorithm," in fulfillment of Written Preliminary Exam II requirement, 1997.

[20] X. G. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," in ICCV, 2007.

[21] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey, "Stel component analysis: modeling spatial correlations in image class structure," in CVPR, 2009.

[22] S. G. Gong, M. Cristani, S. C. Yan, and C. C. Loy, "Mahalanobis distance learning for person re-identification," advances in Computer Vision and Pattern Recognition, in press.

[23] Y. H, S. C. Liao, Z. Lei, D. Yi, and S. Z. Li, "Exploring structural information and fusing multiple features for person re-identification," in CVPRW, 2013