Salient Color Names for Person Re-identification

Yang Yang[1], Jimei Yang[2], Junjie Yan[1],
Shengcai Liao[1], Dong Yi[1], and Stan Z. Li[1,*]

[1] Center for Biometrics and Security Research & National Laboratory of Pattern
Recognition, Institute of Automation, Chinese Academy of Sciences
[2] University of California, Merced
{yang.yang,jjyan,scliao,dong.yi,szli}@nlpr.ia.ac.cn, jyang44@ucmerced.edu

Abstract. Color naming, which relates colors with color names, can
help people with a semantic analysis of images in many computer vision
applications. In this paper, we propose a novel salient color names based
color descriptor (SCNCD) to describe colors. SCNCD utilizes salient
color names to guarantee that a higher probability will be assigned to
the color name which is nearer to the color. Based on SCNCD, color
distributions over color names in different color spaces are then ob-
tained and fused to generate a feature representation. Moreover, the
effect of background information is employed and analyzed for person
re-identification. With a simple metric learning method, the proposed
approach outperforms the state-of-the-art performance (without user's
feedback optimization) on two challenging datasets (VIPeR and PRID
450S). More importantly, the proposed feature can be obtained very fast
if we compute SCNCD of each color in advance.

Keywords: Salient color names, color descriptor, feature representa-
tion, person re-identification.

1 Introduction

Person re-identification is an important topic in visual surveillance. Its goal is
to recognize an individual over disjoint camera views. It is a very challenging
task because the appearance of an individual can be of significant difference in
different viewpoints, illumination, poses, etc. Partial occlusions, low resolution
and background interference add to the intractability of person re-identification.

To address these challenges in person re-identification, many researchers have
proposed different strategies which can be summarized as two stages: (1) **feature
representation** (e.g. [1,4,18,15,13,31,12,7]), which is our main concern in this
paper and (2) **person matching** (e.g. [30,19,28,6,11,5,32,14]).

Color and texture are the most commonly used appearance based features
for person re-identification. Texture descriptors such as Maximally Stable Color
Regions (MSCR) [1], Local Binary Patterns (LBP) [6,11,5] and 21 texture filters
(8 Gablor filters and 13 Schmid fiters) [15] have been successfully applied to

---

* Corresponding Author.

address the problem of person re-identification. But color information, in comparison with texture information, seems to be a more important cue due to the fact that in most cases, only low-resolution images can be obtained. Traditional color information such as color histogram, a simple yet effective feature representation, is most widely used in [1,4,18,15,13,31,12,19,6,11,5,32]. With the consideration of the influence of illumination variations, we calculate color histograms in different color spaces separately and fuse them to make the final feature more robust to illumination changes. However, the performance of feature representation by means of color histograms is sill not satisfactory. Since color names show good robustness to photometric variance [26], an alternative approach is to apply color names to describe colors [12,27,10,26].

In this paper, we propose a novel salient color names based color descriptor (SCNCD) for person re-identification. An example of SCNCD is illustrated in Fig. 1. Different from [27] which is based on Google images, we employ 16 colors[1] from 16-color palette in RGB color space as color names in SCNCD, including fuchsia, blue, aqua, lime, yellow, red, purple, navy, teal, green, olive, maroon, black, gray, silver and white. Inspired by the idea of saliency [8] which is also reflected in other classic coding strategies (e.g. locality-constrained linear coding (LLC) [24], salient coding (SC) [8], local manifold-constrained coding (LMC) [29] and localized soft-assignment coding (LSC) [16]) in image classification, we assign the color's salient color names with nonzero values. Salient color names indicate that one color only have a certain probability of being assigned to several nearest color names, and that the closer one owns a higher probability. For the purpose of making the SCNCD relatively less sensitive to small RGB value changes caused by variations of incident illumination, we employ index to make colors owning the same index have the same color descriptor. The role of index is similar to that of bins to color histogram or that of clusters to bag-of-words model.

To achieve the feature representation, we choose a part-based model [23] which divides each image into six horizontal stripes of equal size, shown in Fig. 2(a) and (c). On the basis of SCNCD, we can obtain the color distribution over the color names (named as color names distribution in this paper) in each part. Examples are shown in Fig. 2(b) and (d). Then, color names distributions of all parts are fused to form an image-level feature. In addition, due to the fact that the background can provide scene context for classification [21], the effect of background information is employed and analyzed for person re-identification. In the stage of person matching, we adopt a fast and effective approach - Keep It Simple and Straightforward MEtric (KISSME) [11,20]. Experimental results show that our proposed method greatly outperforms the state-of-the-art performance on two challenging datasets (VIPeR and PRID 450S).

Contributions. The main contributions of this paper can be summarized as follows. (1) A novel salient color names based color descriptor is proposed for person re-identification. Experimental results demonstrate that SCNCD has shown better

---

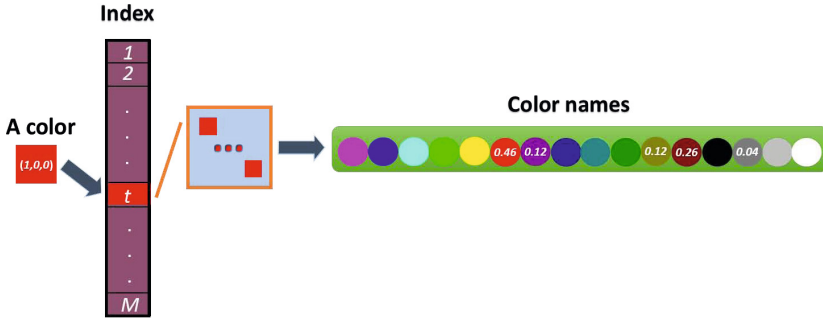[1] Refer to: `http://www.wackerart.de/rgbfarben.html`

Fig. 1. An example of salient color names based color descriptor. The value corresponding to a color name denotes the probability of the set of colors, the indexes of which are the same, being assigned to this color name. It is noted that only several color names have nonzero values.
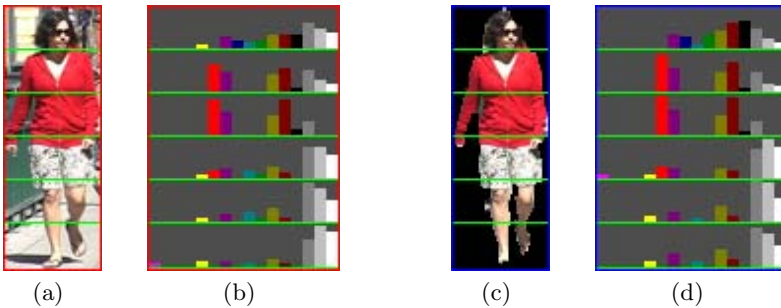

(a)   (b)   (c)   (d)

Fig. 2. An example of the color names distribution of a person image from VIPeR dataset. (a) Divide an image into six parts; (b) Color names distribution of each part of the image based on the SCNCD; (c) Divide the foreground (object of interest) into six parts; The mask used is automatically extracted by using the approach in [9]; (d) Color names distribution of each part of the foreground based on the SCNCD.

performance than the previous color descriptor. (2) Background information is exploited to enrich the feature representation for person re-identification. With it, we can obtain an image-foreground feature representation which is of good robustness against background interference and partial occlusion. (3) Since there is no single color model or descriptor which has the characteristic of robustness against all the types of illumination changes [2,22], features based on color names distributions and color histograms are fused to compensate each other which are computed in four different color spaces including original RGB, rgb, $l_1 l_2 l_3$ [2] and HSV.

2   Related Work

To tackle the problem of person re-identification, many researchers have proposed different approaches the focus of which can be roughly divided into **feature representation** and **person matching**.

Feature Representation. For the sake of describing a person's appearance, many of the existing approaches try to learn a stable as well as very distinctive feature representation. To address the problem of viewpoint changes, Gray *et al.* [4] propose an ensemble of localized features (ELF) to obtain a better representation. Farenzena *et al.* [1] extract three types of features to model the complementary aspects of human appearance, including weighted color histograms, maximally stable color regions (MSCR) and recurrent high-structured patches (RHSP). The algorithm reported in [1] achieves certain robustness against very low resolution, occlusions and pose, viewpoint and illumination changes. The drawback of this feature representation is that it is very time-consuming to extract these three types of features.

Features based on different color spaces and textures are all employed to represent the images, but what features are more important? In [15], Liu *et al.* present a novel unsupervised method to weigh the importance of different features. Experimental results show that the importance of features including different color spaces and textures is different under different circumstances and that instead of treating all features equally, endowing informative feature with a larger weight when different features are fused can lead to better results. The problem of person re-identification is revisited by means of color distribution in [13]. Kviatkovsky *et al.* [13] propose a novel illumination-invariant feature representation based on logchromaticity (log) color space and demonstrate that color as a single cue has a relatively good performance in identifying persons under greatly varying imaging conditions. In consideration of many existing approaches neglecting valuable salient information in matching persons, Zhao *et al.* [31] put forward an unsupervised framework to extract discriminative features for person re-identification and then patch matching is employed with adjacency constraint. The salience in [31] is specially designed to match persons and is robust to different pose, viewpoint variations and articulation. However, traditional color information may not be the optimal way of describing color. Thus, Kuo *et al.* [12] employ semantic color names, which are learned in [27] to describe color and achieve improvements over the state-of-art methods on VIPeR dataset.

Person Matching. Another line of research pays more attention to how to match persons efficiently. For instance, Zheng *et al.* [32] formulate person re-identification as a distance learning problem regardless of the choice of representation. A novel probabilistic relative distance comparison (PRDC) model is proposed which aims to maximise the probability of similar pairs having a smaller distance than that of dissimilar pairs. To solve the problem caused by different camera angles, Hirzer *et al.* [5] learn a Mahalanobis metric learning by employing similar pairs from different cameras. Then a linear projection is obtained that keeps similar pairs together whilst pushes impostors. In [6], a relaxed pairwise metric learning is presented to learn a discriminative Mahalanobis for matching persons from different cameras. It should be noted that a simple yet effective strategy named KISSME is introduced in [11] to learn a distance metric from equivalence constraints from a statistical inference perspective.

Recently, Zhao *et al.* [30] exploit salience matching, which is tightly integrated with patch matching in a unified structural RankSVM learning framework, to match persons over disjoint camera views. In [19], the local fisher discriminant analysis (LFDA) is applied to learn a distance metric for person re-identification problem. After dimensionality reduction, the obtained features can be classified by the nearest neighbor method. Different from the afore-mentioned matching approaches which refer to the target individual as a reference template, Xu *et al.* [28] represent an person image as a compositional part-based template, which introduces flexibility to the matching formulation of person re-identification.

3   Proposed Method

In this section, we first introduce salient color names to describe colors. Color names distribution of a person image is then obtained based on the SCNCD. In addition, background information is employed to form different feature representations that are then fused to obtain the final feature representation. At the end of this section, we will briefly review a simple metric learning method - KISSME.

3.1   Salient Color Names Based Color Descriptor

Color distribution [15,13,6,11,5,2,22] has been widely used to describe a person image in person re-identification. However, it is a challenging task to describe colors because many factors can lead to variations in RGB values, such as variations in illumination and viewpoints. To increase photometric invariance, different color models and descriptors have been presented and evaluated in [2,22]. But no single color model or descriptor has the characteristic of robustness against all the types of illumination changes. Worse still, photometric invariance is often increased at the cost of lowering discriminative ability. To make up the deficiency of RGB values, color names are employed as an alternative way of describing colors in [12,27,26,17]. Experimental results in [17] demonstrate that color description based on color names has a good robustness against photometric variance. Thus, the objective of this subsection is to present a novel approach of describing colors.

To describe colors based on color names, an appropriate mapping from RGB values of a image to color names is required. In this paper, we choose a probability distribution over the color names as a mapping method. Motivated by the idea of saliency, we put forward a novel concept of salient color names. To be specific, for each color to be named, salient color names indicate that this color only has a certain probability of being assigned to several nearest color names, and that the closer the color name is to the color, the higher probability the color has of being assigned to this color name. Fig. 1 gives an example of the salient color names representation of a color. Similar to color histogram or bag-of-words model which assigns a color (or a feature) to bins or 'words' respectively instead of all elements, we introduce index for our SCNCD. Through this way, we can

assign multiple similar colors to the same index with the same color description. In the following, we explain in detail how to compute color description of these similar colors.

Throughout the paper, each channel in all color space is normalized to the range $[0, 1]$. The initial RGB color space is discretized into M indexes. In our case M is $32 \times 32 \times 32 = 32768$ of equally spaced grid points in the RGB cube. Therefore, there are $8 \times 8 \times 8 = 512$ colors for each index. We define $\boldsymbol{d} = \{\boldsymbol{w}_1, ..., \boldsymbol{w}_{512}\}$ as a set of colors the indexes of which are the same. The remaining question is how to calculate the salient color names representation of $\boldsymbol{d}$.

Assume $\boldsymbol{Z} = [\boldsymbol{z}_1, \boldsymbol{z}_2, ..., \boldsymbol{z}_{16}]$ denotes a set of 16 color names defined in the introduction, then the probability of assigning $\boldsymbol{d}$ to a color name $\boldsymbol{z}$ is

$$p(\boldsymbol{z}|\boldsymbol{d}) = \sum_{n=1}^{512} p(\boldsymbol{z}|\boldsymbol{w}_n)p(\boldsymbol{w}_n|\boldsymbol{d}), \tag{1}$$

with $p(\boldsymbol{z}|\boldsymbol{w}_n) =$

$$\begin{cases} \dfrac{\exp\left(-\|\boldsymbol{z}-\boldsymbol{w}_n\|^2/\frac{1}{K-1}\sum_{\boldsymbol{z}_l \neq \boldsymbol{z}}\|\boldsymbol{z}_l-\boldsymbol{w}_n\|^2\right)}{\sum_{p=1}^{K}\exp\left(-\|\boldsymbol{z}_p-\boldsymbol{w}_n\|^2/\frac{1}{K-1}\sum_{\boldsymbol{z}_s \neq \boldsymbol{z}_p}\|\boldsymbol{z}_s-\boldsymbol{w}_n\|^2\right)} & , \quad if \ \boldsymbol{z} \in KNN(\boldsymbol{w}_n) \\ 0 & , \quad otherwise \end{cases} \tag{2}$$

and

$$p(\boldsymbol{w}_n|\boldsymbol{d}) = \frac{\exp\left(-\alpha\|\boldsymbol{w}_n-\boldsymbol{\mu}\|^2\right)}{\sum_{l=1}^{512}\exp\left(-\alpha\|\boldsymbol{w}_l-\boldsymbol{\mu}\|^2\right)}. \tag{3}$$

where K means the number of nearest neighbors, $\boldsymbol{\mu}$ refers to the mean of $\boldsymbol{w}_n$ ($n = 1, ..., 512$). In Eq. (2), $\boldsymbol{z}_p$, $\boldsymbol{z}_l$ and $\boldsymbol{z}_s$ ($p, l, s = 1, ..., K$) belong to K nearest color names of $\boldsymbol{w}_n$. To reflect the saliency of the salient color names for $\boldsymbol{d}$, we first use the KNN algorithm to find K nearest color names of $\boldsymbol{w}_n$ in Euclidean space. Then, the difference between the one of K nearest color names to the other K - 1 color names is utilized to embody the saliency as in [8]. To calculate the saliency degree, we employ a better function $\Phi(t) = \exp(-t)$ instead of [8] which uses $\Phi(t) = 1-t$. After normalization, the probability distribution of $\boldsymbol{w}_n$ over 16 color names is defined as Eq. (2). To further obtain the final probability of $\boldsymbol{d}$ being assigned to color names, Eq. (3) is employed to weigh the contribution of w_n to the $\boldsymbol{d}$. It can be seen in Eq. (3) that the nearer of $\boldsymbol{w}_n$ to $\boldsymbol{\mu}$, the more it contributes to $\boldsymbol{d}$. With Eq. (1), we can describe each set of colors based on their salient color names. We refer to this type of color description as SCNCD in this paper. The biggest difference between salient coding and our SCNCD lies that SCNCD is a description of the probability distribution over its salient color names while salient coding has no relationship with probability distribution. Besides, based on SCNCD, multiple similar colors have the same color description, which increases its illumination invariance. In section 4, we will compare our SCNCD with salient coding which we take as a mapping method from RGB to color names in this paper.

Because all colors in the same set (or have the same index) possess the same salient color names, the salient color names representation of $\boldsymbol{d}$ is also that of the color belonging to the set $\boldsymbol{d}$. Moreover, it is easy to prove that the sum of the distribution of $\boldsymbol{d}$ over all color names $\boldsymbol{z}_m$, $m = 1, ..., 16$ is 1, i.e. $\sum_{m=1}^{16} p(\boldsymbol{z}_m|\boldsymbol{d}) = 1$.

SCNCD has the following advantages:

1. Each color in RGB color space is represented by the probability distribution over its salient color names. Furthermore, to get the salient color names representation, we not only compare the difference among salient color names, but also compare the probability of a color being assigned to each salient color name with that of a color being assigned to overall salient color names. In this way, a relatively reasonable probability distribution can be achieved.

2. It can achieve a certain amount of illumination invariance. Because small RGB value changes caused by illumination will have the same color description if only their indexes are the same.

3. It does not rely on complex optimization and is easy to implement. More importantly, it is very fast because all salient color names representation can be computed offline. Then, we just need to compute each color's index and assign it with its corresponding set's salient color names representation.

3.2 Feature Representation

Once we have achieved the distribution of each color over color names, we can employ them to describe all colors. To capture the color information of an image, we compute color names distribution of an image with the aid of SCNCD. In the following, we first explain in details how to calculate the color names distribution and then show different feature representations.

Color Names Distribution. Because the human body is not rigid, a part-based model [23] is selected instead of taking a person image as a whole. Similar to [15], we partition an image into six horizontal stripes of equal size, as is shown in Fig. 2(a) and (c).

We can find that six parts including the head, upper and lower torso, upper and lower legs and the feet are roughly captured. Let $\boldsymbol{H} = [\boldsymbol{h}_1, ..., \boldsymbol{h}_6]^T$ be the color names distribution of a person image, then the m-th, $m = 1, ..., 16$ element of the distribution of i-th part $\boldsymbol{h}_i = [h_{i1}, ..., h_{i16}]$ is defined as

$$h_{im} = \frac{\sum_{k=1}^{N} p(\boldsymbol{z}_m|\boldsymbol{x}_{ik})}{\sum_{m=1}^{16} \sum_{k=1}^{N} p(\boldsymbol{z}_m|\boldsymbol{x}_{ik})}, \tag{4}$$

where $\boldsymbol{x}_{ik}$, $k = 1, ..., N$, means the k-th color (or pixel) in part i, and N denotes the total number of colors in part i. An example of the color names distribution in each part of a person image is shown in Fig. 2(b) and (d). The bin of color name m denotes the probability of all colors in the corresponding part being assigned to color name m. Similar color names distribution can be obtained except the parts of head and feet between the image and foreground.

Foreground and Background Based Feature Representation. In image classification, it is demonstrated in [21] that background can provide scene context and improve classification accuracy. However, due to the fact that the background in person re-identification is not constant and may even include disturbing factors, background feature representation combined directly with the foreground feature representation will reduce the classification accuracy. To address this problem, we introduce image-foreground feature representation, which can be seen as that the foreground information is employed as the main information while the background information is treated as the secondary one. It alleviates the negative influence of noisy background.

We first introduce image-only and foreground feature representations. (1) **Image-only.** Inspired by the weighted color histograms [1], we endow each pixel x_{ik} with a different weight ω_{ik}:

$$\omega_{ik} = exp(-\frac{(y_{ik} - \mu)^2}{2\sigma^2}), \tag{5}$$

where $\mu = L/2$ and $\sigma = L/4$. In Eq. (5), y_{ik} denotes the column of x_{ik} in the image matrix whose column equals to the image width L. Then, h_{im} defined as Eq. 4 is transformed into

$$h_{im} = \frac{\sum_{k=1}^{N} \omega_{ik} p(z_m | x_{ik})}{\sum_{m=1}^{16} \sum_{k=1}^{N} \omega_{ik} p(z_m | x_{ik})}, \tag{6}$$

where ω_{ik} means the weight of the color x_{ik}. (2) **Foreground.** To obtain the foreground representation, we need a mask to extract the object of interest. In this paper, we use the mask which is automatically obtained by the method [9] with the parameter settings used in [1]. It is a commonly used mask (or a revised mask) in person re-identification [1,18,23,20]. Color names distribution can be obtained according to Eq. (6) for foreground feature representation. Then, image-only and foreground feature representations are concatenated to form the **image-foreground** feature representation.

Fusion of Different Features. Since there is no single color model or descriptor which has the characteristic of robustness against all types of illumination changes [2,22], features based on four different color models including original RGB, normalized rgb, $l_1 l_2 l_3$ [2] and HSV are selected and fused to compensate each other. Because the range of pixel values of them in each channel are from 0 to 1, we can take them as a type of transformation of RGB values.

$$(\theta) = \mathcal{T}(\theta_o), \tag{7}$$

where θ_o (or θ) denotes the original RGB value (or the transformed RGB value) while $\mathcal{T}$ means a transformation approach. For example, the transformation for normalized rgb is

$$\mathcal{T}(a, b, c) = (a/(a + b + c), b/(a + b + c), c/(a + b + c)). \tag{8}$$

In addition, color histogram is also fused with color names distribution to improve the accuracy. Then, the final image-foreground feature representation is obtained by concatenating all image-foreground feature representations which are based on color names distribution and color histograms over the four different color models.

3.3   Person Matching

Mahalanobis distance learning has attracted considerable attention in computer vision. Given a pair of samples x_i and x_j $(x_i, x_j \in \mathcal{R}^d)$, the Mahalanobis distance between them is

$$d_M^2(x_i, x_j) = (x_i - x_j)^T M (x_i - x_j). \tag{9}$$

where $M \geqslant 0$ is a positive semidefinite matrix. From a statistical inference point of view, KISSME defines the Mahalanobis distance matrix M by

$$M = \Sigma_{\mathcal{S}}^{-1} - \Sigma_{\mathcal{D}}^{-1}. \tag{10}$$

where

$$\Sigma_{\mathcal{S}} = \frac{1}{|\mathcal{S}|} \sum_{x_i, x_j \in \mathcal{S}} (x_i - x_j)(x_i - x_j)^T, \tag{11}$$

$$\Sigma_{\mathcal{D}} = \frac{1}{|\mathcal{D}|} \sum_{x_i, x_j \in \mathcal{D}} (x_i - x_j)(x_i - x_j)^T. \tag{12}$$

denote the covariance matrices for similar pairs $\mathcal{S}$ and dissimilar pairs $\mathcal{D}$ respectively. Then, M can be learned easily from the training samples. More details can be found in KISSME [11,20].

4   Experiments

In this section, we evaluate our method on two publicly available datasets (VIPeR dataset [3] and PRID 450S dataset [20]). VIPeR dataset is commonly employed for single-shot re-identification while PRID 450S dataset is a recently published dataset and is more realistic than VIPeR dataset. Each person has one image pair in both datasets. Therefore, the single-shot evaluation strategy [1] (described specifically in the following experimental settings) can be used. All the results are shown in form of Cumulated Matching Characteristic (CMC) curve [25].

4.1   Settings

In our experiment, we randomly choose half image pairs for training and the remaining half image pairs are used for test. In the stage of test, images from one camera are treated as probe and those from the other camera as gallery.

Then, we switched the probe and gallery. The average of the results is regarded as one-trial CMC result. Similar to [19,11], we repeat 100 trials of evaluation and report the average result to achieve a more stable results in the following. When we calculate the SCNCD, the number of nearest neighbor of intermediate variable w_n in Eq. (2) is set to 5. α is set to 1 in Eq. (3). As in [11,20], the principal component analysis (PCA) is employed to reduce the computational efforts before KISSME is applied. When we compute the color histogram, the number of bins of each channel is set to 32 for all color models. In the following, we name **im**age-only, **for**eground and **im**age-**f**oreground feature representations as *Img*, *Forg* and *ImgF* respectively.
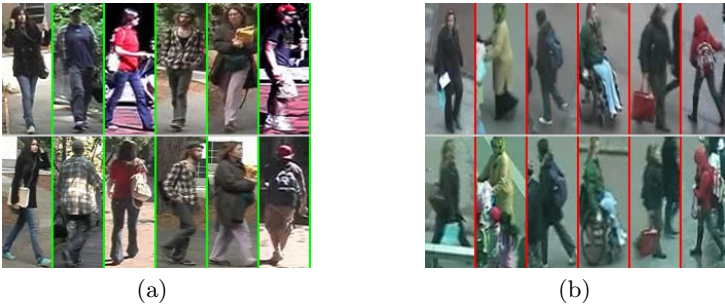

Fig. 3. Some examples from the datasets: (a)VIPeR and (b) PRID 450S

4.2 VIPeR Dataset

This is a challenging dataset[2] for viewpoint invariant pedestrian recognition (VIPeR), suffering from arbitrary viewpoints, pose changes and illumination variations between two camera views. It contains 632 image pairs, which correspond to 632 persons and are captured by two cameras in outdoor academic environment. Images from Camera A are mostly captured from 0 degree to 90 degree and that from Camera B mostly from 90 degree to 180 degree. In experiments, we normalize all images to 128×48 pixels. Some examples from VIPeR are shown in Fig. 3(a).

Comparison with the State-of-the-art Methods. We compare our methods ($SCNCD_{all}(ImgF)$ and Final($ImgF$)) with the state-of-the-art methods on VIPeR dataset. $SCNCD_{all}(ImgF)$ refers to *ImgF*s of SCNCD over four color models while Final($ImgF$) means all the *ImgF*s of color names distributions and color histograms over four color spaces are concatenated. The compared methods are following the same evaluation protocol as ours. The dimensions of the features are reduced to 70 by PCA.

Table 1 shows that both $SCNCD_{all}(ImgF)$ and Final($ImgF$) outperform others. We can also find that when we fuse the $SCNCD_{all}(ImgF)$ with different

---

[2] Available at: http://vision.soe.ucsc.edu/?q=node/178

Table 1. Comparison with the state-of-the-art methods on VIPeR dataset

Rank	1	5	10	15	20	25	30	50
PRDC[32]	15.7	38.4	53.9	-	70.1	-	-	-
Fusing+PRDC[15]	16.1	37.7	51.0	-	66.0	-	-	-
RPLM[6]	27	-	69	-	83	-	-	95
EIML[5]	22	-	63	-	78	-	-	93
KISSME[11]	19.6	-	62.2	-	-	80.7	-	91.8
KISSME*[20]	27.0	-	70.0	-	83.0	-	-	95
eSDC-ocsvm[31]	26.7	50.7	62.4	-	76.4	-	-	-
RankBoost[12]	23.9	45.6	56.2	-	68.7	-	-	-
LF[19]	24.2	-	67.1	-	-	85.1	-	94.1
Salience[30]	30.2	52.3	-	-	-	-	-	-
SCNCD$_{all}$($ImgF$)	**33.7**	**62.7**	**74.8**	**81.3**	**85.0**	**87.7**	**89.6**	**93.8**
Final($ImgF$)	**37.8**	**68.5**	**81.2**	**87.0**	**90.4**	**92.7**	**94.2**	**97.0**

color histograms, the obtained Final($ImgF$) leads to a 4.1% improvement compared to $SCNCD_{all}(ImgF)$ at rank 1. In addition, by comparing our approaches with approaches used in [11] and [20], our feature representation shows more effectiveness.

4.3   PRID 450S Dataset

PRID 450S dataset[3] is a new and more realistic dataset. It contains 450 single-shot image pairs captured over two spatially disjoint camera views. Fig 3(b) shows some examples from PRID 450S dataset. It is also a challenging person re-identification dataset due to different viewpoint changes, background interference, partial occlusion and viewpoint changes. In experiments, each image is normalized to 168×80 pixels.

Comparison with the State-of-the-art Methods. Because the PRID 450S is a new dataset, few methods have been tested on it. We only compare our approach with the best results reported in [20] which uses the existing methods. We use $SCNCD_{all}(ImgF)$ and Final($ImgF$). The dimensions of the features are reduced to 70 by PCA.

It is shown in Table 2 that our proposed methods outperform KISSME [11] and EIML [5] both of which employ the precise masks (generated manually). Specifically, the results of $SCNCD_{all}(ImgF)$ and Final($ImgF$) are at least 6.0% higher than the best result EIML[5] at rank 1. On the PRID 450S dataset, the improvement from $SCNCD_{all}(ImgF)$ to Final($ImgF$) is not as great as that in VIPeR dataset. This is because there are background noise and partial occlusion in PRID 450S dataset, which influences the performance of the color histograms.

---

[3] Available at: https://lrs.icg.tugraz.at/download.php

Table 2. Comparison with the state-of-the-art methods on PRID 450S dataset. KISSME* and EIML employ precise masks (generated manually). Our proposed methods employ the masks (generated automatically).

Rank	1	5	10	15	20	25	30	50
KISSME*[20]	33.0	-	71.0	-	79.0	-	-	90.0
EIML[5]	35	-	68	-	77	-	-	90
SCNCD$_{all}$($ImgF$)	**41.5**	**66.6**	**75.9**	**81.1**	**84.4**	**86.7**	**88.4**	**92.4**
Final($ImgF$)	**41.6**	**68.9**	**79.4**	**84.9**	**87.8**	**90.0**	**91.8**	**95.4**

4.4   Analysis of SCNCD

The performance of the proposed SCNCD is analyzed in the following:

Comparison with Other Color Descriptions. We compare our proposed SC-NCD with several existing color descriptions on both VIPeR dataset and PRID 450S dataset, including color histogram, discriminative descriptor (DD) [10], the color names (CN) [27], semantic color names (SCN) [17] and salient coding representation (SCR) [8].

Table 3. Comparison with different color descriptions on VIPeR dataset. All six color descriptions are calculated in RGB color space. *Img* is employed for them.

Rank	1	5	10	15	20	25	30	50
Hist(RGB)	6.5	22.8	34.8	43.4	50.5	55.9	60.3	72.6
SCN[17]	11.9	32.3	45.9	55.0	61.8	67.1	71.4	83.7
SCR[8]	12.5	32.9	45.9	54.3	60.4	65.0	68.9	79.1
DD[10]	17.6	40.3	52.4	60.2	66.0	70.3	73.6	82.7
CN[27]	19.6	44.2	58.1	66.3	72.3	76.9	80.4	88.8
SCNCD(Ours)	**20.7**	**47.2**	**60.6**	**68.8**	**75.1**	**79.1**	**82.4**	**90.4**

For DD, we choose the best setting, namely 25 clusters. To obtain SCR, we employ salient coding [8] to map the color to color names and treat the mapping coefficients the color's description. To evaluate the performance of those six color descriptions fairly, we compute all of them based on *Img* in RGB color space while KISSME is employed as a matching method. The dimension is reduced to 34 (the same as in [11,20]) using PCA. It can be seen from Tables 3 and 4 that our proposed SCNCD outperforms all of other color descriptions on both datasets at all ranks.

***Img* v.s. *Forg* v.s. *ImgF*.** Three types of feature representations are shown in section 3.2, including *Img*, *Forg* and *ImgF*. We compare the performances of them on VIPeR and PRID 450S datasets. RGB color model is selected for

Table 4. Comparison with different color descriptions on PRID 450S dataset. All six color descriptions are calculated in RGB color space. *Img* is employed for them.

Rank	1	5	10	15	20	25	30	50
Hist(RGB)	4.9	17.6	28.7	36.7	43.6	49.2	54.0	68.0
SCN[17]	6.6	20.7	31.6	39.3	46.0	51.6	56.2	70.0
SCR[8]	9.6	26.2	37.0	44.4	49.8	54.8	59.1	70.4
DD[10]	17.6	40.3	52.4	60.2	66.0	70.3	73.6	82.7
CN[27]	20.4	42.6	53.3	60.3	65.3	68.9	71.8	79.3
SCNCD(Ours)	**26.9**	**52.9**	**64.2**	**70.4**	**74.9**	**78.0**	**80.4**	**87.3**

SCNCD. The dimension of *Img* (or *Forg*) is reduced to 34 by PCA while the dimension of *ImgF* 50.

In Fig. 4(a) and (b), *ImgF* shows better performance than traditional *Img* and *Forg* on the VIPeR and PRID 450S datasets. In addition, we can see from Fig. 4(a) that *Forg* yields similar results as *Img* on VIPeR dataset while Fig. 4(b) shows that *Forg* significantly outperforms *Img* on PRID 450S dataset. This phenomenon demonstrates that there is much more background noise in PRID 450S than in VIPeR. This is why the improvement in PRID 450S is not as great as in VIPeR when the background information is added in *ImF*.
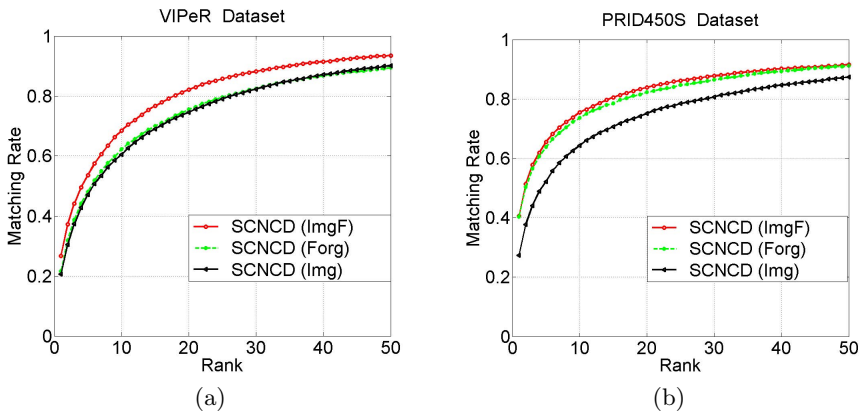
Fig. 4. Performances of different feature representations on: (a)VIPeR and (b) PRID 450S based on SCNCD in RGB color space.

SCNCD with Color Models. We employ four color models for SCNCD when we calculate the color names distribution of a person in section 3.2. Since these color modes are used to address the problems of illumination changes in [2,22], we choose VIPeR dataset to test our approach which suffers from illumination variations between two cameras. Features based on different color models including original RGB, normalized *rgb*, $l_1 l_2 l_3$ and HSV as well as feature obtained
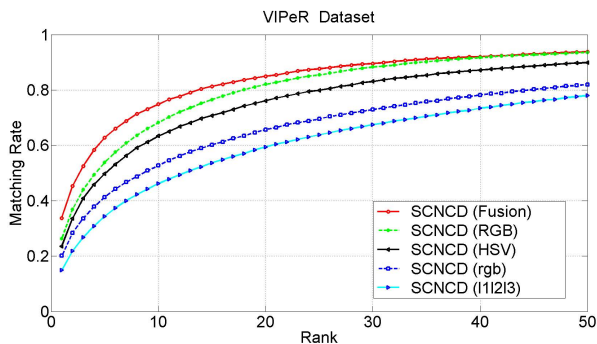
Fig. 5. Different color models are compared based on SCNCD

by fusing them are compared. *ImgF* is selected as the feature representation. The dimensions of features based on different color models are reduced to 50 by PCA while the dimension of the fusing feature 70. Fig. 5 shows the experimental results. It can be seen that among the four color models, SCNCD based on RGB achieves the best results. Thus, SCNCD shows a certain amount of illumination changes. Moreover, when we fuse these features computed under different color models, the recognition accuracy is improved significantly. It benefits from that these color models are invariant to different types of illumination [2].

5   Conclusion

In this paper, we propose a novel method to describe a color by its salient color names. It is very fast because each color can be represented by its corresponding index's color names representation which is precomputed. Then, color names distributions are computed over different color models and fused to address the illumination problem. Background information is added in the image-foreground feature representation. To improve the recognition accuracy, other color distribution methods based on different color histograms are also fused with color names distribution. Finally, we formulate the person re-identification problem as a color distribution matching problem. Experiments demonstrate that our proposed SCNCD possesses a certain robustness with background interference and partial occlusion and that the final image-foreground feature representation significantly improves the recognition accuracy of person re-identification.

References

1. Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: Person re-identification by symmetry-driven accumulation of local features. In: Proc. CVPR (2010)
2. Gevers, T., Smeulders, A.W.: Color-based object recognition. Pattern Recognition 32(3), 453–464 (1999)
3. Gray, D., Brennan, S., Tao, H.: Evaluating appearance models for recognition, reacquisition, and tracking. In: IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (2007)
4. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part I. LNCS, vol. 5302, pp. 262–275. Springer, Heidelberg (2008)
5. Hirzer, M., Roth, P.M., Bischof, H.: Person re-identification by efficient impostor-based metric learning. In: Proc. AVSS (2012)
6. Hirzer, M., Roth, P.M., Köstinger, M., Bischof, H.: Relaxed pairwise learned metric for person re-identification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part VI. LNCS, vol. 7577, pp. 780–793. Springer, Heidelberg (2012)
7. Hu, Y., Liao, S., Lei, Z., Li, S.Z.: Exploring structural information and fusing multiple features for person re-identification. In: Proc. CVPRW (2013)
8. Huang, Y., Huang, K., Yu, Y., Tan, T.: Salient coding for image classification. In: Proc. CVPR (2011)
9. Jojic, N., Perina, A., Cristani, M., Murino, V., Frey, B.: Stel component analysis: modeling spatial correlations in image class structure. In: Proc. CVPR (2009)
10. Khan, R., de Weijer, J.V., Khan, F.S., Muselet, D., Ducottet, C., Barat, C.: Discriminative color descriptors. In: Proc. CVPR (2013)
11. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: Proc. CVPR (2012)
12. Kuo, C.H., Khamis, S., Shet, V.: Person re-identification using semantic color names and rankboost. In: Proc. WACV (2013)
13. Kviatkovsky, I., Adam, A., Rivlin, E.: Color invariants for person reidentification. IEEE Trans. on PAMI 35(7), 1622–1634 (2013)
14. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, L., Smith, J.R.: Learning locally-adaptive decision functions for person verification. In: Proc. CVPR (2013)
15. Liu, C., Gong, S., Loy, C.C., Lin, X.: Person re-identification: What features are important? In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 391–401. Springer, Heidelberg (2012)
16. Liu, L., Wang, L., Liu, X.: In defense of soft-assignment coding. In: Proc. ICCV (2011)
17. Liu, Y., Zhang, D., Lu, G., Ma, W.Y.: Region-based image retrieval with high-level semantic color names. In: Proceedings of the 11th International Multimedia Modelling Conference (2005)
18. Ma, B., Su, Y., Jurie, F.: Local descriptors encoded by fisher vectors for person re-identification. In: Fusiello, A., Murino, V., Cucchiara, R. (eds.) ECCV 2012 Ws/Demos, Part I. LNCS, vol. 7583, pp. 413–422. Springer, Heidelberg (2012)
19. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: Proc. CVPR (2013)
20. Roth, P.M., Hirzer, M., Kostinger, M., Beleznai, C., Bischof, H.: Mahalanobis distance learning for person re-identification. Advances in Computer Vision and Pattern Recognition (2014)

21. Russakovsky, O., Lin, Y., Yu, K., Fei-Fei, L.: Object-centric spatial pooling for image classification. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part II. LNCS, vol. 7573, pp. 1–15. Springer, Heidelberg (2012)
22. van de Sande, K.E., Gevers, T., Snoek, C.G.: Evaluating color descriptors for object and scene recognition. IEEE Trans. on PAMI 32(9), 1582–1596 (2010)
23. Satta, R.: Appearance descriptors for person re-identification: a comprehensive review. In: Proc. CoRR (2013)
24. Wang, J., Yang, J., Yu, K., Lv, F., Huang, T., Gong, Y.: Locality-constrained linear coding for image classification. In: Proc. CVPR (2010)
25. Wang, X., Doretto, G., Sebastian, T., Rittscher, J., Tu, P.: Shape and appearance context modeling. In: Proc. ICCV (2007)
26. van de Weijer, J., Schmid, C.: Applying color names to image description. In: Proc. ICIP (2007)
27. van de Weijer, J., Schmid, C., Verbeek, J., Larlus, D.: Learning color names for real-world applications. IEEE Trans. on Image Processing. 18(7), 1512–1523 (2009)
28. Xu, Y., Lin, L., Zheng, W.S., Liu, X.: Human re-identification by matching compositional template with cluster sampling. In: Proc. ICCV (2013)
29. Zhang, X., Yang, Y., Jiao, L., Dong, F.: Manifold-constrained coding and sparse representation for human action recognition. Pattern Recognition 46(7), 1819–1831 (2013)
30. Zhao, R., Ouyang, W., Wang, X.: Person re-identification by salience matching. In: Proc. ICCV (2013)
31. Zhao, R., Ouyang, W., Wang, X.: Unsupervised salience learning for person re-identification. In: Proc. CVPR (2013)
32. Zheng, W.S., Gong, S., Xiang, T.: Person re-identification by probabilistic relative distance comparison. In: Proc. CVPR (2011)