A Hybrid Approach to Pedestrian Clothing Color Attribute Extraction

Mu Gao, Yuning Du, Haizhou Ai
Computer Sci. & Tech. Dept., Tsinghua University
Beijing 100084, P. R. China
gaom13@mails.tsinghua.edu.cn

Shihong Lao
OMRON Social Solutions Co. Ltd.
Japan
lao_shihong@oss.omron.co.jp

Abstract

Clothing attributes, of which color plays an important role, are receiving more and more interests in machine vision researches and applications because of their uses and effectiveness in tasks like pedestrian analysis. However, color description is a challenging problem due to complex environments such as illumination variations. Most prior works describe color attributes using only low-level features or mid-level descriptors, which results in a marked drop of the discriminative power or photometric invariance. In this paper we introduce a new efficient joint representation that aims to overcome the shortcomings of using low-level features or mid-level descriptors alone and present a novel hybrid approach to pedestrian clothing color attribute extraction. As a necessary preprocessing step, a novel processing pipeline is also proposed. We evaluate our approach on the task of color classification on both the public dataset VIPeR and our own newly-built pedestrian dataset. Experimental results have demonstrated the effectiveness of our approach and have shown its great potential for further researches and applications.

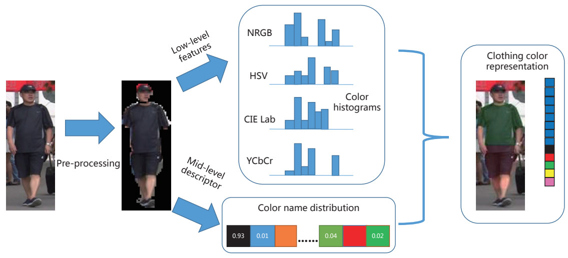

Figure 1. Overview of our approach. The main motivation of our approach is to propose a novel clothing color attribute representation combining low-level features (color histograms) with mid-level semantic descriptor (color name distribution).

1 Introduction

Over recent years the study of clothing appearance is receiving more and more interests in machine vision researches and applications since clothing can be used as an important contextual cue to improve the performance of related tasks. Attributes, which serve as mid-level semantic and human-understandable properties, have been widely applied to more and more pedestrian related researches. Numerous and various pedestrian attributes [7, 9, 10, 13] have been proposed for pedestrian analysis. As one of the most useful attributes, clothing appearance is of special practical interest due to its high application potential in areas such as image classification, pedestrian re-identification, clothing based or textual based person retrieval, clothing recognition and online shopping, etc. [1, 5, 7, 9, 10, 11, 12, 13].

In this paper we focus on clothing color attribute, which is of vital importance because it is one of the most obvious and easily perceived attributes that similar to descriptions provided by human eyewitness. Despite the great variety of clothes, the clothing colors are always discriminative and important, which have similar and limited characteristics. Though the effectiveness and distinctiveness of color attribute have been demonstrated by different methods, how to develop a robust clothing representation is still an unsolved and challenging problem due to the influence of complex environments such as illumination variation in pedestrian images.

Most previous works in pedestrian analysis utilize only color histograms to describe color attribute. Different color spaces have different characteristics such as some are more close to human perception while others are better for industrial uses. Histograms in various color spaces such as RGB, HSV and YcbCr are adopted by most existing methods of feature extraction in pedestrian oriented tasks [1, 7, 8]. They are common in using only histograms as color features while they are different in choosing color spaces and bins of color channels. Low-level color histogram features have good generalization capability but are lack of discriminative power i.e. there exists a semantic gap between the low-level features and high-level attributes. Another alternative way of color description is by means of color names. Weijer [2, 3] proposed a new approach to learn color names from real-world images. They used Google images and PLSA method to learn a color name (CN) model, which mapped RGB values to 11 pre-defined colors (black, blue, brown, grey, green, orange, pink, purple, red, white and yellow). Similarly, Yang et.al [6] proposed another mapping model between RGB values and 16 salient color names and applied it to person re-identification, the probability for each RGB value over certain color names is pre-calculated. Color naming models can be seen as a new mid-level color descriptor in form of a probability distribution over certain number of color names. They are discriminative in semantic viewpoints but can be easily influenced by photometric invariance. To overcome the drawbacks of using low-level features or mid-level descriptors alone, we introduce a new efficient joint representation, combining complementary characteristics of each other. We aim at achieving a balance between discriminative power and photometric invariance. In order to extract efficient color attribute, pre-processing steps mainly include two aspects: pedestrian segmentation and body part subdivision. We also present a novel processing pipeline, utilizing an efficient silhouette generating method followed by an effective body part subdivision model. Our main contributions in this paper include: (1) We propose a new processing pipeline and a hybrid approach to pedestrian clothing color attribute extraction; (2) We develop a new pedestrian clothing color representation, combining high-dimensional low-level features (color histograms) and low-dimensional mid-level semantic descriptors (color names distribution); (3) We build a new pedestrian dataset with annotated clothing color attribute. Our representation and approach are demonstrated to perform well in both the public dataset VIPeR and our own dataset. The paper is organized as follows. In Section 2, the framework of our approach is explained in detail. Section 3 describes the experiments on two datasets. The conclusion summarizes the contents of this paper.

2 Our Approach

2.1 Pre-processing

Our pre-processing procedures include two main steps: pedestrian segmentation and body part subdivision, as shown in Figure 2.

For a robust pedestrian clothing representation, the respective regions have to be segmented first and the irrelevant non-person pixels need to be removed, thus an efficient pedestrian segmentation method is the prerequisite. Previous approaches [1, 10] in pedestrian analysis usually adopted empirically prior knowledge and fixed rectangles to localize the foreground. Farenzena [8] used a generative model named STEL for background elimination in still images. However, the main limitation of stated methods is that pedestrian segmentation results are often not accurate enough, resulting in influence on following feature extraction. As far as we know, no effective algorithms have been proposed for pedestrian segmentation. Recently from another point of view Luo [4] applied deep learning method to pedestrian parsing in still images and introduced a new Deep Decompositional Network for parsing pedestrian images into semantic regions.

Inspired by the work of Luo [4], we develop a lightweight variant of its method and found that it is very effective in doing pedestrian segmentation in various datasets. We made a little adaptation of the method to do two parts parsing i.e. foreground-background segmentation. The masks after background removal are then used on original images to generate pedestrian silhouettes. The results on various images are found to be better compared with prior knowledge based fixed rectangles location and other existing methods.

After segmentation, we need to locate different body descriptors typically exploit a part-based body model because of the non-rigid structure of the human body i.e. relative positions of body parts are not fixed a-priori but are inferred from the image. Satta [5] summarizes the part-based body models used in current appearance descriptors and divides them into three categories: fixed, adaptive and learned models. Fixed

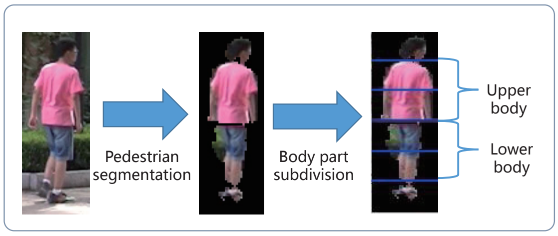

Figure 2. Pre-processing of pedestrian clothing color attribute extraction.

models, in which the sizes and positions of body parts are chosen a-priori, are simple yet proven to be effective in many pedestrian related tasks [5, 6, 10]. We utilized a fixed part model, which divides each pedestrian image into six horizontal stripes of equal size. As a result, the human body can be divided into four main parts: head, upper body, lower body and foot, by using the first, second and third, fourth and fifth, sixth stripe respectively. Thus the upper body and lower body can be obtained as shown in Figure 2. When it comes to the lower body part, short lower garment must be taken into account. This issue can be easily settled by utilizing the fourth stripe alone for the lower body clothing color attribute extraction. Results show that the upper and lower body clothing can be almost exactly obtained in this approach.

2.2 Color Attribute Representation

After upper body and lower body parts are obtained, we can form a joint clothing color attribute representation which aims to gain both discriminative power and photometric invariance. In order to build a part-based representation, each pedestrian $\mathbf{P}$ is defined as a sequence of m stripes (here m is 6 in our representation):

$$\mathbf{P} = [P_1, ..., P_m]$$

We only consider the upper body and lower body i.e. P_2, P_3, P_4, P_5 here. For each stripe, we extract a fused representation, combing low-level features with mid-level descriptors. The feature vector of the stripe j is denoted as:

$$P_j = [CH_{1j}, ..., CH_{kj}, CN_{1j}, ..., CN_{sj}]$$

where CH_k is the color histograms of the k-th color space and CN_s is the color name probability for color s. Though color histograms are effective features in pedestrian analysis, no single color model or descriptor is robust enough against all illumination changes, so we decided to use four color spaces to strengthen the robustness and the discriminative power. For each stripe of upper body and lower body separately, we calculate color histograms in four different color spaces (namely RGB, HSV, YCbCr and CIE Lab) , each with 16 bins per channel and then concatenate them into one combined color descriptor as the color histogram feature vector ($2\times4\times3\times16 = 384$ dimensions). We use PCA to do dimension reduction (50 dimensions). Then we calculate the color name distribution for each stripe using the model in [2, 3]. For the stripe j (j = 2, 3, 4, 5 in our representation), the color name distribution CN_s is defined as:

$$CN_{sj} = \frac{1}{N} \sum_{x_{RGB} \in R_j} p(CN_s | x_{RGB})$$

where $p(CN_s|x_{RGB})$ is the probability of a certain pixels RGB value (x_R, x_G, x_B) being assigned to a specific color name CN_s, R_j is the foreground region of the stripe j, N denotes the summation of pixels that are in R_j. Then the final clothing representation is obtained by fusing color histograms with CN into one feature vector. We expect our color attribute can be more representative by taking advantage of the photometric invariance of color histograms and the discriminative power of color name descriptor.

2.3 Learning Algorithm

We employ multi-class SVM classifiers to learn the clothing color categories. The effectiveness of SVM in discriminative learning has been widely proved. After comparing different kernels in experiments, we choose linear kernel for our learning algorithm.

3 Experiments

Datasets and Experimental settings. The first dataset we used for evaluating our approach is VIPeR, one of the most commonly used and challenging datasets for person re-identification, containing 1264 images of 632 different pedestrians. Satta [9] has manually tagged all VIPeR images for their pre-defined attributes. We chose a subset of their label results including the upper and lower body parts colors which are related to our task as ground-truth for our experiments. After discarding the samples whose color cannot be determined, they divided 1264 images into 8 categories for upper body clothing colors: *black* (298), *blue* (84), *green* (72), *pink* (42), *red* (73), *white* (277), *brown* (71), *grey* (70). Lower body clothing colors are labeled into 5 categories: *blue* (568), *white* (112), *black* (178), *grey* (52), *brown* (40).

We also built our own pedestrian dataset after having examined some public datasets we can get and found most of them (including VIPeR) unbalanced in samples because of lack of clothing color diversity, i.e. some clothing color attributes have little positive examples. To avoid bias and better test our approach, we built our own pedestrian dataset. These images are not only of different clothing colors but also of various poses and illumination, most persons in our dataset are captured with other clothing attributes (e.g. carrying a bag). With a broad range of clothing with large appearance variations, our dataset, containing over 2800 images of nearly 600 different pedestrians under different conditions, makes the problem more challenging and close to reality. We manually accomplished clothing color attribute annotation, classifying the images to different color categories according to two multi-class attributes: upper body clothing color and lower body clothing color. Upper body clothing color attribute contains 7 color categories: *black*, *blue*, *green*, *pink*, *red*, *white*, *yellow*, with 600, 600, 256, 298, 252, 600, 289 images respectively. Lower body clothing color attribute contains 3 color categories: *black*, *blue*, *white*, with 600 images for each color. The size of each pedestrian image is 160*400. To guarantee annotation results are as close to human perception as possible, the annotation procedure is made by different persons independently and the ambiguous images are discarded, thus the remaining images are in agreement to human-perception in color attribute and can serve as ground truth in our task. This dataset can facilitate future pedestrian oriented research for tasks such as recognition, identification, classification and so on. For every category in each dataset, half of the images are used for training and the rest half are for testing. The division of training and testing samples ensures there are no overlapping persons in two parts because most persons appear in more than one image in the dataset.

Figure 3. Examples of labelled images taken from VIPeR dataset. (a) Examples for upper body clothing colors (from left to right: *black*, *blue*, *brown*, *green*, *grey*, *pink*, *red*, *white*); (b) Examples for lower body clothing colors (from left to right: *blue*, *white*, *black*, *grey*, *brown*).

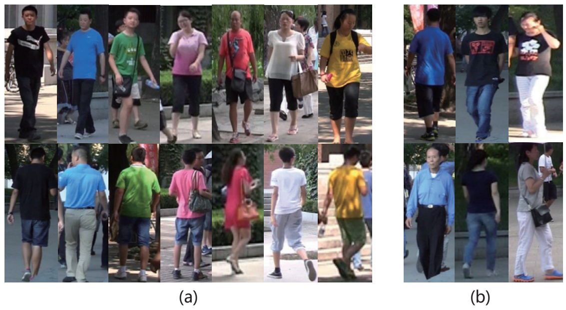

Figure 4. Examples of images taken from our dataset. (a) Examples for upper body clothing colors (from left to right: *black*, *blue*, *green*, *pink*, *red*, *white*, *yellow*); (b) Examples for lower body clothing colors (from left to right: *black*, *blue*, *white*).

Evaluation of different representations. Firstly we evaluate different representations on two datasets. The results are shown in Table 1, where CH and CN respectively denotes using color histograms features only and using color name distribution only, [U] and [L] respectively denotes upper body part and lower body part. As shown below, our joint representation obtained obviously better results compared to color histogram features only or CN descriptor only in almost all color categories.

Evaluation of different approaches. We compared our approach with the one stated in [9] following

Table 1. Evaluation of different approaches for clothing color classification on VIPeR.

Colors	Classification Accuracy	
	Best result in [9]	Ours
Black[U]	74%	**95.33%**
Blue[U]	48%	**94.46%**
Green[U]	59%	**83.74%**
Pink[U]	51%	**80.64%**
Red[U]	69%	**95.4%**
White[U]	77%	**94.77%**
Brown[U]	46%	**84.39%**
Grey[U]	38%	**91.21%**
Overall[U]	65.62%	**88.06%**
Blue[L]	90%	**95.81%**
White[L]	68%	**96.13%**
Black[L]	75%	**95.68%**
Grey[L]	30%	**86.09%**
Brown[L]	62%	**65%**
Overall[L]	80.13%	**93.05%**

Table 2. Evaluation of different representations for clothing color classification.

Datasets	Colors	Classification Accuracy		
		CH	CN	Joint
Ours	Black[U]	96.92%	96.37%	**98.23%**
	Blue[U]	96.68%	93.93%	**96.82%**
	Green[U]	93.03%	79.27%	**93.28%**
	Pink[U]	96.26%	90.08%	**98.21%**
	Red[U]	98.73%	97.58%	**99.49%**
	White[U]	89.27%	81.88%	**89.83%**
	Yellow[U]	94.21%	91.49%	**96.70%**
	Overall[U]	91.37%	84.11%	**93.09%**
	Black[L]	97.33%	93.83%	**99.33%**
	Blue[L]	96.33%	87.5%	**97.67%**
	White[L]	96.67%	87.67%	**99.33%**
	Overall[L]	96.78%	86.22%	**98.78%**
VIPeR	Black[U]	92.62%	85.91%	**95.33%**
	Blue[U]	76.19%	47.62%	**94.46%**
	Green[U]	58.33%	55.56%	**83.74%**
	Pink[U]	57.14%	38.10%	**80.64%**
	Red[U]	81.08%	56.76%	**95.4%**
	White[U]	90.58%	68.84%	**94.77%**
	Brown[U]	44.44%	25%	**84.39%**
	Grey[U]	74.71%	5.71%	**91.21%**
	Overall[U]	80.78%	61.35%	**88.06%**
	Blue[L]	94.85%	77.16%	**95.81%**
	White[L]	94.28%	79.64%	**96.13%**
	Black[L]	93.39%	83.24%	**95.68%**
	Grey[L]	80.21%	50%	**86.09%**
	Brown[L]	**69.56%**	50%	65%
	Overall[L]	91.37%	73.26%	**93.05%**

the same experimental settings.

The average results over ten runs are shown in table 2, it should be pointed out that [9] has implemented 3 MCD descriptors and the results below are taken from the best one for each color. Our method has improved the classification accuracy a lot.

Our proposed joint representation is proved to be more discriminative than using either one alone. Moreover, compared with reported method, our whole approach improves the performance a lot, which demonstrates the effectiveness of both our processing pipeline and attribute representation.

4 Conclusion

In this paper we have presented a new processing pipeline and a novel hybrid approach to clothing color attribute extraction in pedestrian images based on our novel joint representation which combines low-level color features with mid-level color descriptors in order to overcome the drawbacks of using either color histograms or color name descriptor alone. Experimental results on our own newly built dataset and VIPeR dataset well demonstrate that our approach and representation efficiently improve the performance of clothing color classification, which is of great importance in pedestrian analysis. The proposed approach and the discriminative clothing color attribute representation provide great potential for future further researches and applications in more realistic open-world scenarios.

References

[1] M. Yang and K. Yu: "Real-time clothing recognition in surveillance videos," *In Proc. ICIP*, 2011.

[2] Van de Weijer, J., Schmid, C.: "Applying Color Names to Image Description," *In Proc. ICIP*, 2007.

[3] Van de Weijer, J., Schmid, C., Verbeek, J., Larlus, D.: "Learning Color Names for Real-World Applications," *IEEE Trans. On Image Processing*, vol.18, no.7, pp.1512–1523, 2009.

[4] P. Luo, X. Wang, and X. Tang: "Pedestrian Parsing via Deep Decompositional Neural Network," *In Proc. ICCV*, 2013.

[5] Satta, R.: "Appearance Descriptors for Person Re-identification: a Comprehensive Review," *In Proc. CoRR*, 2013.

[6] Yang, Y. Yan, J. Yan, J. Liao, S. Yi, D. Li, S. Z. Li.: "Salient Color Names for Person Re-identification," *In Proc. ECCV*, 2014.

[7] J. Zhu, S. Liao, Z. Lei, D. Yi, and S. Z. Li.: "Pedestrian Attribute Classification in Surveillance: Database and Evaluation," *In Proc. ICCV Workshop*, 2013.

[8] Farenzena, M., Bazzani, L., Perina, A., Murino, V., Cristani, M.: "Person Re-identification by Symmetry-driven Accumulation of Local Features," *In Proc. CVPR*, 2010.

[9] Satta, R., Pala, F., Fumera, G., Roli, F.: "People Search with Textual Queries about Clothing Appearance Attributes," *In Person Re-Identification: Advances in Computer Vision and Pattern Recognition*, pp.371–389, 2014.

[10] R. Layne, T. M. Hospedales, S. Gong, and Q. Mary: "Person Re-identification by Attributes," *In Proc. BMVC*, 2012.

[11] S. Khan, F. Anwer, R. Muhammad, J. van de Weijer, A. Joost, M. Vanrell, and A. Lopez: "Color attributes for object detection," *In Proc. CVPR*, 2012.

[12] Wang. XW, Zhang. T, Tretter. DR, Lin, Q Wang: "Personal Clothing Retrieval on Photo Collections by Color and Attributes," *IEEE Trans. On Multimedia*, vol.15, no.8, pp.2035-2045, 2013.

[13] H. Chen, A. Gallagher, and B. Girod: "Describing Clothing by Semantic Attributes," *In Proc. ECCV*, 2012.