Person Re-identification in Surveillance Videos using Multi-part Color Descriptor

P.K. Sathish
Computer Science and Engineering Dept.
Christ University
Bengaluru- 560074

S. Balaji
Centre for Emerging Technologies, Jain University
Bengaluru-562112

ABSTRACT

Person re-identification is gaining significance in today's world. In this paper, we present a novel method for re-identifying persons in surveillance videos. We propose a new color based image descriptor which is used to extract the local interest points by subdividing the pedestrian image into upper and lower parts. The descriptor is experimentally evaluated on two benchmark datasets.

General Terms

Pattern recognition, Video surveillance, computer vision.

Keywords

Person re-identification, color descriptor, part based, weighted color histogram.

1. INTRODUCTION

In recent times, there is steady growth in the use of video surveillance systems. The increase in the use of surveillance cameras in public and private places to track and monitor various human activities has boosted the need for automated person re-identification systems. The person re-identification systems will be useful for tracking and identifying individuals entering in a multi-camera surveillance system. Person re-identification problem is very challenging due to reasons such as change in camera positions, angle of camera placement, lighting conditions, occlusions and so on. To re-identify a person, we cannot depend on the biometric cues such as face or gait because of the relatively low resolution images and variation of the angle in which image is captured. One of the common assumptions in such a scenario is that the appearance of the person being tracked is likely to be same. The information regarding the color of clothes and other local feature descriptors can be used to identify the person.

In this paper, we propose a method of person re-identification by extracting local image descriptors and color features from the upper and lower parts of the body using color Pyramidal Histogram of Visual Words (PHOW) descriptor [1] and combining it with weighted HSV histogram [6]. The image features are combined and matched using k-nearest neighbor search. The overall methodology is shown in Fig 1.

We call our method as Multi-Part Color Descriptor (MPCD) based person re-identification. Our method is tested on VIPeR [8] and ETZH [5] datasets.

The rest of the paper is organized as follows. Section 2 gives an insight on the work done in the related field. Section 3 discusses our method. Section 4 deals with methods used for feature matching and re-identification and Section 5 describes our experiments and results. Section 6 gives concluding remarks.
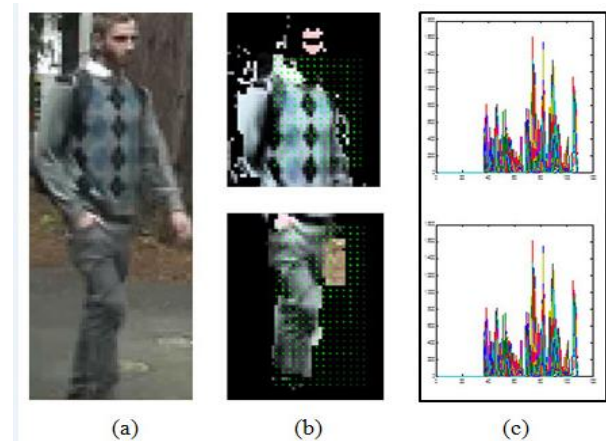

Fig 1. (a) Input Image (b) PHOWC Descriptor of Subdivided Image (c) Weighted Color Histogram of (b)

2. RELATED WORK

Person re-identification is similar to instance recognition problem, where the task is to obtain a ranking score in comparison with the queried person. Appearance based re-identification methods mainly depend on the image descriptors extracted from the interest points detected locally or globally [4]. A spatio-temporal method uses invariant signatures that are obtained by using normalized color and salient edgel histograms is discussed by N. Gheissari et al. [7]. In [5], a costume based identification approach is adopted. G. Jaffre et al. [9], symmetry of the human body is used. The body is divided into regions based on symmetry and descriptors are generated using three separate features such as the color Histograms in HSV color space, Maximally Stable Color (MSCR) to generate blobs of stable color and Recurrent High Structured Patches (RHSP) to extract highly recurrent image patches on the person's appearance. Generation of a human signature by correlation of body parts of the person is discussed in [6]. Fast re-identification is explored in [11] by subdividing the human body into parts. [16], [3], and [13] have discussed advantages of color image descriptors.

3. MULTI PART COLOR DESCRIPTOR

In the pre-processing phase, we obtain the foreground silhouette of the person. This is obtained using STEL model as described in [10]. The meaningful segmentation of parts thus obtained is learned to detect the foreground of the image. This is adapted from [6] to obtain the foreground silhouette images. The foreground image of the person is subdivided into two parts featuring the upper part and lower part. Image feature vectors are computed separately for each subdivision.

We consider the anti-symmetry model proposed in [6] to subdivide the body in parts.

3.1 Pyramidal Histogram of Visual Words Descriptor

PHOW descriptor was introduced in [1]. This is a SIFT based descriptor which is extracted at different scales. We use the color version of this descriptor to obtain the features of the upper part and lower part of the foreground image. In the color version of PHOW, descriptors are extracted from three HSV image channels. K-means algorithm is used to form the visual word [17]. A histogram of the frequency of the visual words is computed. This step is adapted by dividing the image into smaller sub-regions. The frequency histograms which are computed for each sub regions are combined to form the final descriptor.

3.2 Weighted Color Histogram

Histograms exhibit invariance to image transformations such as translation and rotation along the viewing axis and the variation of changes in view angle, scale and occlusion leads to small variations in the histogram [14].

HSV color histogram is used to obtain the chromatic values of the upper and lower part of the body of the person. The weight of pixel is calculated from the vertical axes as in [6]. Pixel values near the vertical axes have more weightage. The pixel weightage is determined by a one dimensional Gaussian kernel.

4. FEATURE MATCHING AND RE-IDENTIFICATION

Different features are combined to form a single signature. This section demonstrates how this is carried out. First, we find the PHOW-color descriptors for both upper part and lower part of the pedestrian.

Let

$$I = \{I_u, I_l\} \quad (1)$$

be a collection of N images of pedestrians. Where I_u, I_l represent the upper and lower body parts, respectively. We find the PHOW-color descriptors for both I_u and I_l. This is denoted as

$$PHOW_C(I_u) = \{i_u^1, \dots \dots i_u^n\} \quad \text{where} \quad i_u^k \text{ are feature}$$
vectors of I_u. $\quad (2)$

$$PHOW_C(I_l) = \{i_l^1, \dots \dots i_l^n\} \quad \text{where } i_l^k \text{ are feature vectors}$$
of I_l $\quad (3)$

A nearest neighbor classifier is used to find the distances between $PHOW_C(I_u^A)$ and $PHOW_C(I_u^B)$ where A and B are the image galleries and B contains the probe set.

Similarly, the distances between $PHOW_C(I_l^A)$ and $PHOW_C(I_l^B)$ are also determined. The combined signature is formed by concatenating the distances.

$$d(I_A, I_B) = \beta_{PHOW_C} \cdot d_{PHOW_C}(PHOW_C(I_l^A), PHOW_C(I_l^B))$$

$$+ \beta_{WH} \cdot d_{WH}(WH(I_A), WH(I_B)) \quad (4)$$

Where d_{PHOW_C} is the distance of the $PHOW_C$ descriptor and dWH is the distance of the weighted histogram and β

normalized weights. The value of β is calculated as mentioned in [2].

5. EXPERIMENTS AND RESULTS

We conducted experiments on two publicly available datasets: VIPeR [8] and ETZH [5]. The PHOW-Color descriptor was implemented using VLFEAT software [15].

VIPeR dataset consists of 632 images of pedestrians which are taken from different non-overlapping cameras. This dataset provides images of pedestrian with different poses, different lighting conditions and varied viewpoints. Each image has a size of 128 x 48. We used images in the folder CAM A as query images and the images in CAM B as gallery images. 200 images were used in this experiment. The CMC and SRR curves are shown in Fig 3. The rank 1 matching is above 20% in MPCD.

ETZH dataset was used in [12] for pedestrian detection. The dataset consists of three sequences of images which are captured from moving cameras. We used the third sequence for our experiments. It consists of images of 28 different persons and a total of 1762 images. The CMC and SRR curves are shown in Fig 4. The figure shows a very good matching score of above 65% for rank 1.
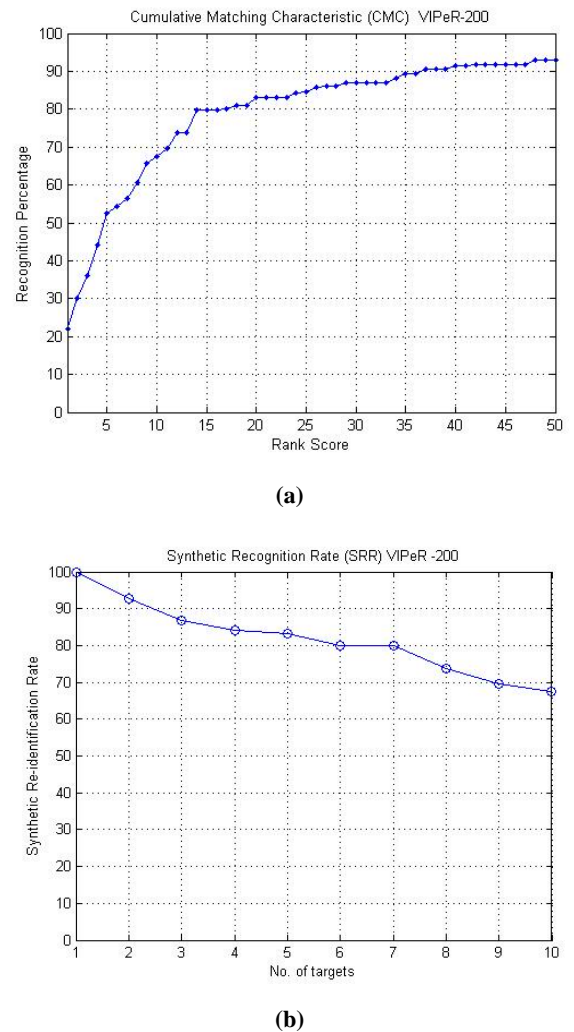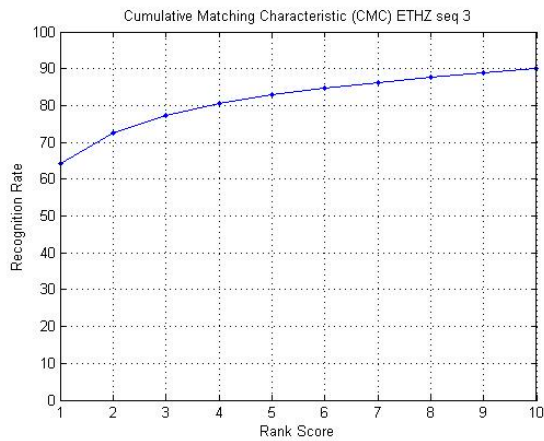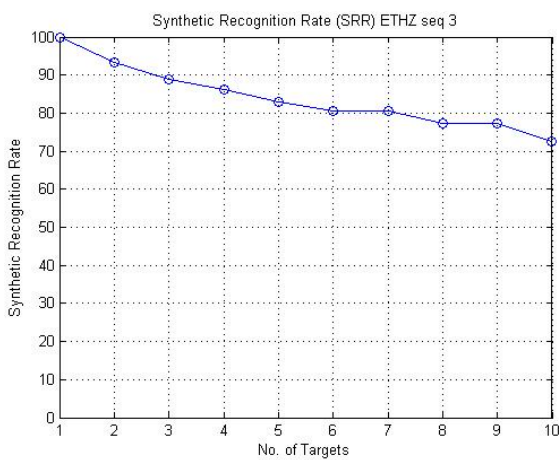

(a)

(b)

Fig 3. Performances on VIPeR dataset. (a) Cumulative Matching Characteristics Curve (b) Synthetic Recognition Rate.

(a)

(b)

Fig 4. Performances on ETHZ dataset. (a) Cumulative Matching Characteristics Curve   (b) Synthetic Recognition Rate

6. CONCLUSIONS

Our paper addresses person re-identification method using appearance based color descriptor. Our method is based on extracting color descriptors from upper and lower parts of the pedestrian images. The challenges of illumination changes and pose invariances are dealt by this method. MPCD shows good recognition rates at lower rank score. Use of different classifying techniques can be explored to improve this work. Further, multiprocessing techniques can be employed to improve the execution speed of the MPCD descriptor.

7. REFERENCES

[1] Bosh, A Zisserman and X. Munoz, 2007, Image classification using random forests and ferns" In Proc. ICCV.

[2] S. Bak, E. Corvee, F. Bremond, M. Thonnat, ,2010, Person re-identification using spatial covariance regions of human body parts," Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Nashville, pp. 435-440.

[3] G. Burghouts, J.-M. Geusebroek, 2009, "Performance evaluation of local color invariants," Computer Vision Image Understanding, pp. 48-62.

[4] G. Doretto, T. Sebastian, P. Tu, J. Rittscher, 2011 "Appearance-based person re-identification in camera networks: problem overview and current approaches", Journal of Ambient Intelligence and Humanized Computing, pp 127-151.

[5] Ess, B-Leibe, L. V. Gool, 2007, "Depth and appearance for mobile scene analysis," In ICCV.

[6] M. Farenzena, L. Bazzani, A. Perina, V. Murino, M. Cristani, 2006, Person re-identification 7. N. Gheissari, T. B. Sebastian, P. H. Tu, J. Rittscher, and R. Hartley, "Person re-identification using spatiotemporal appearance," In CVPR, vol. 2, pp 1528-1535.

[7] N. Gheissari, T. B. Sebastian, P. H. Tu, J. Rittscher, and R. Hartley, 2006, "Person re-identification using spatiotemporal appearance," In CVPR, vol. 2pp 1528-1535.

[8] D. Gray, S. Brennan, and H. Tao, 2007, "Evaluating appearance models for recognition, reacquisition and tracking," In PETS.

[9] G. Jaffre and P. Joly, 2004, "Costume: A new feature for automatic video content indexing, "In proceedings of RIAOpp, 314-325.

[10] N. Jojic, A. Perina, M. Cristani, V. Murino, and B. Frey, 2009, "Stel component analysis: Modeling spatial correlations in image class structure," In CVPR, pp 2044-2051.

[11] Satta R., Fumera G., Roli F., 2012, "Fast person re-dentification based on dissimilarity representations, "Pattern Recognition Letters.

[12] W. Schwartz and L. Davis, 2009, "Learning discriminative appearance-based models using partial least squares," In XXII SIBGRAPI.

[13] H. Stokman, T. Gevers, 2007, "Selection and fusion of color models for image feature dettection,"IEEE Trans. Pattern Anal. Mach. Intel.,pp 371-381

[14] M. J. Swain and D.H. Ballard, 1991, "Color indexing, "International Journal of Computer Vision, pp 11-32.

[15] Vedaldi, B. Fulkerson, 2010, "Vlfeat an open and portable library of computer vision algorithms, "in: Proceedings of the 18th Annual ACM International Conference on Multimedia, Firenze, Italy,pp 1469-1472.

[16] Verma, S. Banerji, C. Liu, 2010, "A new color SIFT descriptor and methods for image category classification", in: Proceedings of the International Congress on Computer Applications and Computational Science.

[17] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, 2010, "Locality-constrained linear coding for image classification", in: Proceedings of the International Congress on Computer Applications and Computational Science.