

Color Based Pre-rank Categorization for Person Re-identification

Jamal Hussain Shah, Zonghai Chen, Saeed ur Rehman, Mudassar Raza, Lin Mingqiang

Department of Automation

University of Science and Technology of China

Hefei, Anhui

jamalhussainshah@gmail.com, chenzh@ustc.edu.cn, msaeed1981@yahoo.com, mudassarkazmi@yahoo.com,

kdlmq@mail.ustc.edu.cn

Abstract—The most addressed issue in person re-identification systems is to manage the performance under large gallery set and invariant appearance. In this study, considering these issues we proposed a color based pre-rank categorization (color basket) for person re-identification. In addition, coarse saliency dense color features were extracted to generate signature. For this purpose, we remove background and apply scale-invariant feature transform (SIFT) to get convex hull boundary to find the coarse saliency region. Furthermore, an image is divided into three horizontal strips and the dense color features are obtained. For training, incremental LDA (linear discriminant analysis) is utilized and extended with mRMR (minimum Redundancy Maximum Relevance) to handle the high computational cost. The proposed method is evaluated on three publically available datasets, namely i-LIDS, VIPeR and GRID. Moreover, a cumulative matching characteristic (CMC) curve is generated. The curve shows the propose strategy is good in dealing with the above-mentioned problems.

Keywords— *pre-rank categorization; dense color features; LDA; mRMR*

I. INTRODUCTION

Person re-identification has recently drawn great interest in computer vision applications. Person re-identification is a task to match and recognize persons across non-overlapping cameras. The key factors that degrade the performance of a person re-identification system are pose variation, viewpoint, poor resolution, occlusion and a large dataset. In real-time detection and re-identification, the most important factor involves maintaining the system's performance. The focus of this study is on sustaining the performance of a person re-identification system with a large gallery set.

Several researchers have studied the person re-identification problem. Re-identification methods are classified into two groups, namely descriptive (feature-based) [1], and discriminative (distance metric learning) [2]. Based on these categories, literature on person re-identification is presented as follows.

Feature-based techniques have robust features that can distinguish images in terms of viewpoint orientation, illumination and lighting conditions. In recent years, several approaches of generating descriptors have been introduced. Appearance-based approaches frequently used for image retrieval or person re-identification are based on the color, texture and shape of the object [3]. The color histogram-based

feature [4] is the most popular and frequently used person re-identification. Morioka et al. [5] used color histograms to extract features for different views and extended them with PCA to combine them for person re-identification in a multi-camera environment. Javed et al. [6] also utilized color histogram features for the same purpose. Many authors introduced saliency based and graph based feature descriptor to handle these problems in person re-identification [7]. Disadvantages of descriptive methods relate to high-dimensional data and the intra-class feature variation problem. Discriminative methods have widely been employed to overcome these limitations.

Other re-identification methods are based on distance metric learning, which involves measuring the distance or similarity of pairs of extracted features and requires training. Based on these categories, many strategies have been introduced to solve it, such as K-Nearest Neighbor [8], LMNN[9], logistic discriminant metric learning (LDML) [10], information theoretic metric learning (ITML) , L1-Norm and L2-Norm[11]. These methods compare two non-overlapping cameras images and seek a shortest distance for an output image. The differences between the descriptive and distance metric methods are: 1) metric learning methods perform a direct comparison instead of data training, 2) comparatively discriminative methods mostly use Euclidian distance for decision-making, whereas distance metric learning methods use probability distance or non-Euclidean distance.

The majority of human re-identification works presented are based on appearance, as discussed above. Color features are always dominant in object recognition. The main concern in color-based human re-identification is with retrieving accurate objects from large gallery sets. Thus, it is necessary to use an accurate distance between the probe and gallery set. In this context, this article introduces a method of maintaining the performance of a person re-identification system with a large gallery set. The principal contributions of the proposed work include:

- Color-based learning is introduced for pre-rank color categorization.
- Visual salience is detected to remove the background using the SIFT-based salience technique to obtain foreground images and partition them into three horizontal strips.

The authors would like to thank National Natural Science Foundation of PR China (61375079), and the Chinese Academy of Science-The World Academy of Sciences (CAS-TWAS) President's Fellowship.

- Dense color features are generated corresponding to the partition strips.
- 4. Incremental LDA (iLDA) learning is used to deal with the overhead of a large gallery set for fast processing. To deal with the high computational cost of i-LDA, the minimum Redundancy Maximum Relevance mRMR technique is employed.
- A similarity measurement technique is used with the gallery set to check the robustness of the system.

In subsequent sections, the proposed framework will be discussed in detail.

II. SYSTEM DESCRIPTION

An overview of the propose model is presented in Fig. 1. Regarding the person re-identification subject, a pre re-rank color categorization basket is introduced to handle a large gallery set. The appearance features are extracted from the coarse saliency region. Before this, the image is divided into three horizontal strips is trained with iLDA and extended with mRMR. A detailed description is given in upcoming sections.

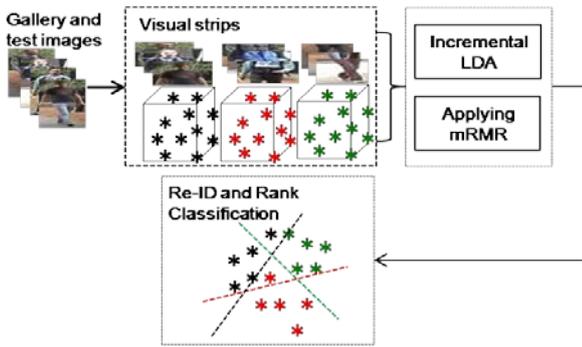


Fig. 1. A flow diagram of the system

A. Human tracking and detection

The histogram of oriented gradients (HOG) is a renowned object classification method that is extensively used for human detection [12]. In this study, the HOG method is also used for this purpose. Gradient computation is the main step in HOG, which can be positive or negative and is calculated by 1-D discrete derivation masks in the X-axis $[-1 \ 0 \ 1]$ or Y-axis $[-1 \ 0 \ 1]^T$ directions at pixel (x, y) . Gradient orientation is done with the following equation:

$$\theta = \tan^{-1}(p_y/p_x) \quad (1)$$

where θ is the orientation on point (x, y) . Rectangular (R-HOG) are adopted to fabricate cell and block descriptors. Next, a normalization process is applied to eliminate redundant features. SVM is a prominent method that is frequently used for accurate classification. Here, SVM is extended for classifying positive (human) body parts and negative (non-human) body parts.

B. Color-based pre-ranking

Classification aims to categorize objects based on robust features. Accurate classification usually depends on two things: 1) discriminative features and 2) learning. The challenge in person re-identification is how to organize large, continuous training data. Researchers have used different strategies to overcome this challenge, such as the data forgotten [13]. In the current study, the color-based pre-rank optimization learning method is proposed, which comprises two stages: 1) color basket creation and 2) color matching. In this case, a color basket is initially created and only six colors are used to represent all 255 remaining colors, namely red, green, blue, yellow, black, and white. When a new image arrives, the training algorithm obtains its color and places it in the nearest neighbor's color basket. Thus, the number of objects in each basket is continuously incrementing. The advantages of the proposed pre-rank categorization learning method over other methods include: 1) it minimizes the search time and accurately classifies unseen data and 2) it offers an optimal solution and handles huge databases easily.

C. Signature generation and training

In this section, feature extraction and signature generation for human re-identification are discussed. As mentioned in the previous section, prior to creating the color basket, a dense color feature is created from the coarse saliency region. To determine the coarse saliency region, the image background is first eliminated and the SIFT method is applied to detect the key points. The convex hull is computed by using the detected key points and the boundary for the coarse saliency region is marked.

The next step is to segment the output image into three horizontal equal strips and perform dense color on each strip to extract diverse feature vector. It can be shown as:

$$VC_t = \left\{ \left\{ s_1^u, s_1^m, s_1^l \right\}, \left\{ s_2^u, s_2^m, s_2^l \right\}, \dots, \left\{ s_n^u, s_n^m, s_n^l \right\} \right\} \quad (2)$$

where, u, m and l are known as upper, middle and lower strips respectively and s is represent as strip. In addition, individual strip such as $\left\{ s_n^u, s_n^m, s_n^l \right\}$ is known as single node and have same label ID as of top, middle and lower strips.

Data labeling and classification is an important step in human re-identification. Here, incremental LDA learning is utilized for training purposes. To deal with the high computational cost, the minimum Redundancy Maximum Relevance mRMR technique is extended [14].

Let considered h as a single node (upper, middle or lower strip) and features are represented as σ . The correlation ξ between nodes is calculated by the formula:

$$\xi(x, y) = \iint p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \quad (3)$$

where $p(x)$ and $p(y)$ are shows that the probability distribution functions to occur within the specific visual cue

(single strip is known as visual cue) while $p(x, y)$ is the combined probability of occurrence of two mutually correlated visual cues.

According to the definition of mRMR, the maximum relevance between two classes can be defined as:

$$\max D(\sigma, h), D = \frac{1}{|\sigma|} \sum_{x_i \in \sigma} \xi(x_i, h) \quad (4)$$

The redundant features within a class are found and minimized using:

$$\min R(\sigma, c), R = \frac{1}{|\sigma|} \sum_{x_i, x_j \in \sigma} \xi(x_i, x_j) \quad (5)$$

Where, $\xi(x_i, x_j)$ is the mutual information between features x_i and x_j . By subtracting Eq. (4) from (5), we get the following mRMR equation:

$$\max \Phi(D, R), \Phi = D - R \quad (6)$$

This process is applied to each newly-acquired test image and distance d is calculated and compared with the set of gallery images as:

$$d_i^s = \{d(VC_i(x), VC(x)), s = u, m, l\} \quad (7)$$

In our case, the distances where two out of three strips have the same ID are first sorted and their ranks are computed accordingly. The shortest distance is calculated as:

$$o = \arg \min_i \{d(f_i(x), f(x))\} | d(f_i(x), f(x)) \in \Delta_t \quad (8)$$

where t and Δ_t are the threshold and calculated distance below the threshold respectively. Proposed method is summarized in the Algorithm 1.

Algorithm 1: person re-identification

Input: Probe image I

Output: A correctly classified in dataset VC

1. Perform color based pre-rank and select into respective color basket.
 2. Remove the background and mark SIFT points.
 3. Mark the convex hull boundary and find the coarse saliency region.
 4. Compute overall incremental LDA and mRMR using Eq. (6).
 5. Apply distance function using Eq. (7).
 6. Sort out those distances whose two out of three strips have same ID and compute rank.
-
-

III. EXPERIMENTS AND RESULTS

The proposed method was evaluated using two publically available datasets, namely VIPeR and GRID. VIPeR is the most challenging and widely used dataset in person re-identification. It is captured from an outdoor environment and contains 632 images with two images per person [15]. The DRID dataset was captured by 8 disjoint cameras at a busy underground station. It has 250 images with occlusion and

illumination variations [16]. For a fair evaluation of the proposed technique on all three datasets, all images were partitioned to form the gallery and probe set. In addition, different P values were selected to compute the cumulative match characteristic (CMC) curve. Different P value such for VIPeR $p=100$ and $p=316$ and for GRID $p=125$ were chosen to evaluate the performance. Each test was repeated 20 times and obtained top-30 ranking of each subjects reported as the final result.

The proposed method is compared with existing person re-identification methods. We also use similar protocol as suggested for i-LIDS and VIPeR and DRID datasets by them. Cumulative match characteristic (CMC) curve is calculated to compare the performance of proposed algorithm. The comparison results of person re-identification are shown in Fig.2, Fig. 3, Fig. 4. Curve shows the average results of the proposed method.

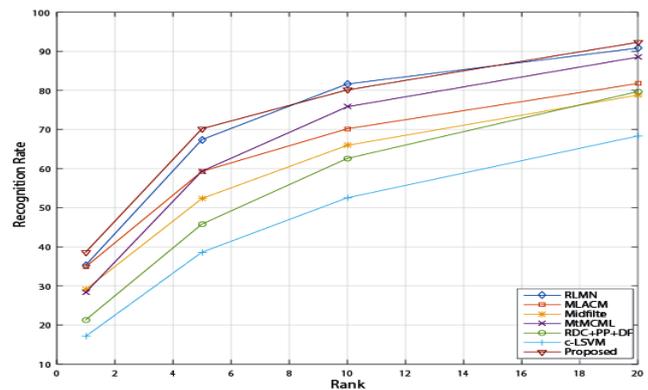


Fig. 2. Comparison of the proposed method with existing state-of-the-art methods (RLMN[2], MLACM[17], Midfilter [18], MtMCML[19], PDC+PP+DF[20], c-LSVM[21]) for the VIPeR dataset. The top 20 ranking positions are depicted at $p=316$.

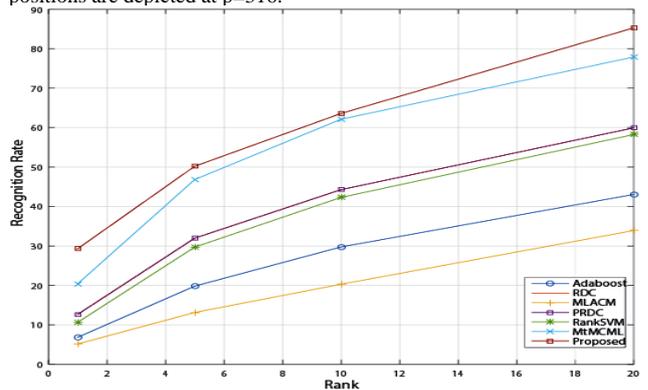


Fig. 3. Comparison of the proposed method with existing state-of-the-art methods (AdaBoost [22], RDC[23], MLACM[17], PRDC[23], RankSVM [23], MtMCML[19]) for the VIPeR dataset. The top 20 ranking positions are depicted at $p=432$.

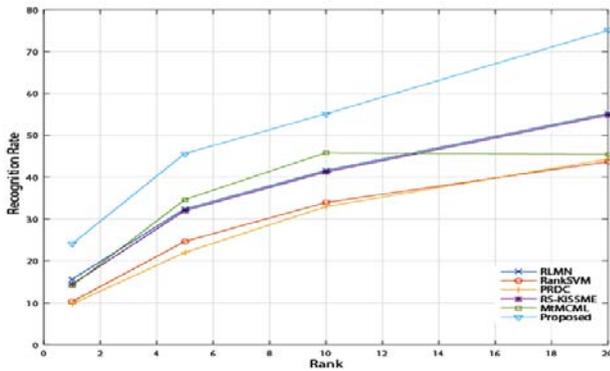


Fig. 4. Comparison of the proposed method with existing state-of-the-art methods (RLMN[2], RankSVM [23], RS-KISSME[24], MtMCMC[19]) for the GRID dataset. The top 20 ranking positions are depicted at $p=125$.

In the subsequent experiment, the proposed feature extraction method was compared with the pre-rank optimization strategy and without. Table 1 compares the average results of the proposed method with and without pre-rank optimization.

TABLE I. PROPOSED METHOD COMPARISON WITH AND WITHOUT PRE-RANK OPTIMIZATION.

Method	Without pre-rank optimization				With pre-rank optimization			
	Rank 1	Rank 5	Rank 10	Rank 20	Rank 1	Rank 5	Rank 10	Rank 20
VIPeR								
$P=316$	34.8	58.1	78.3	88.5	35.6	63.2	74.5	91.3
$P=432$	25.0	42.3	59.6	72.1	30.3	51.2	63.6	86.2
GRID								
$P=125$	21.0	36.5	56.5	66.5	25.0	45.6	55.1	75.5

IV. CONCLUSION

In this study, the person re-identification problem in a multi-camera environment was solved. The main contribution of the proposed state-of-the-art technique is in handling re-identification in two situations, namely with invariant appearance and continuous data. The SIFT-based coarse salience region was used to detect and obtain the dense color features that are accurate in invariant appearance. In addition, each object was divided into three horizontal strips to make searching faster. Continuous data is a major problem that arises in person re-identification. Hence, pre-rank categorization was utilized to minimize the data overhead. The comparison with other state-of-the-art methods showed that the proposed technique has superior performance in all conditions.

Acknowledgment

The authors would like to thank National Natural Science Foundation of PR China (61375079), and the Chinese Academy of Science-The World Academy of Sciences (CAS-TWAS) President's Fellowship.

References

[1] B. Ma, Y. Su, and F. Jurie, "Discriminative Image Descriptors for Person Re-identification," in *Person Re-identification*, ed: Springer, 2014, pp. 23-42.

[2] V. E. Liong, J. Lu, and Y. Ge, "Regularized local metric learning for person re-identification," *Pattern Recognition Letters*, 2015.

[3] A. Bedagkar-Gala and S. K. Shah, "A survey of approaches and trends in person re-identification," *Image and vision computing*, vol. 32, pp. 270-286, 2014.

[4] N. K. E. Abbadi and E. H. A. Saadi, "Improvement of Automatic Hemorrhages Detection Methods Using Shapes Recognition," *arXiv preprint arXiv:1310.5999*, 2013.

[5] K. Morioka, X. Mao, and H. Hashimoto, "Global color model based object matching in the multi-camera environment," in *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, 2006, pp. 2644-2649.

[6] O. Javed, Z. Rasheed, K. Shafique, and M. Shah, "Tracking across multiple cameras with disjoint views," ed: Google Patents, 2008.

[7] Y. Geng, H.-M. Hu, G. Zeng, and J. Zheng, "A person re-identification algorithm by exploiting region-based feature salience," *Journal of Visual Communication and Image Representation*, vol. 29, pp. 89-102, 2015.

[8] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, 2013, pp. 3586-3593.

[9] M. Hirzer, P. M. Roth, M. Köstinger, and H. Bischof, "Relaxed pairwise learned metric for person re-identification," in *Computer Vision-ECCV 2012*, ed: Springer, 2012, pp. 780-793.

[10] L. An, M. Kafai, S. Yang, and B. Bhanu, "Reference-based person re-identification," in *Advanced Video and Signal Based Surveillance (AVSS), 2013 10th IEEE International Conference on*, 2013, pp. 244-249.

[11] R. Mazzone, S. F. Tahir, and A. Cavallaro, "Person re-identification in crowd," *Pattern Recognition Letters*, vol. 33, pp. 1828-1837, 2012.

[12] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, 2005, pp. 886-893.

[13] V. S. Kenk, S. Kovačič, M. Kristan, M. Hajdinjak, and J. Perš, "Visual Re-Identification Across Large, Distributed Camera Networks," *Image and vision computing*, 2014.

[14] M. Mandal and A. Mukhopadhyay, "An improved minimum redundancy maximum relevance approach for feature selection in gene expression data," *Procedia Technology*, vol. 10, pp. 20-27, 2013.

[15] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Proc. IEEE International Workshop on Performance Evaluation for Tracking and Surveillance (PETS)*, 2007.

[16] C. Liu, S. Gong, and C. C. Loy, "On-the-fly feature importance mining for person re-identification," *Pattern Recognition*, vol. 47, pp. 1602-1615, 2014.

[17] S.-C. Shi, C.-C. Guo, J.-H. Lai, S.-Z. Chen, and X.-J. Hu, "Person Re-identification with Multi-Level Adaptive Correspondence Models," *Neurocomputing*, 2015.

[18] R. Zhao, W. Ouyang, and X. Wang, "Learning mid-level filters for person re-identification," in *Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on*, 2014, pp. 144-151.

[19] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *Image Processing, IEEE Transactions on*, vol. 23, pp. 3656-3670, 2014.

[20] Z. Wu, Y. Li, and R. J. Radke, "Viewpoint invariant human re-identification in camera networks using pose priors and subject-discriminative features," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 37, pp. 1095-1108, 2015.

[21] A. Li, L. Liu, K. Wang, S. Liu, and S. Yan, "Clothing Attributes Assisted Person Re-identification," 2014.

[22] S. G. W. Zheng, T. Xiang, "Re-Identification by Relative Distance Comparison," *IEEE Trans. Pattern Anal. Mach. Intel.*, vol. 35, pp. 653-668, 2013.

[23] W.-S. Zheng, S. Gong, and T. Xiang, "Person re-identification by probabilistic relative distance comparison," in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, 2011, pp. 649-656.

[24] D. Tao, L. Jin, Y. Wang, Y. Yuan, and X. Li, "Person re-identification by regularized smoothing kiss metric learning," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 23, pp. 1675-1685, 2013.