An improved method for Person Re-Identification

Yuan Wang
Information Engineering Institute, Wuhan University of Technology, Hubei, China
yuan_wang@sina.com

Abstract. In this paper, an effective feature representation named color pairs is proposed, which is robust to illumination and viewpoint changes. It was combined with Scale-invariant local ternary pattern (SILTP) texture features used for person re-identification. And we improved the method of Cross-view Quadratic Discriminant Analysis (XQDA) by adding a close function. It improves the performance of discriminant matrix. This method were compared with other hand-crafted methods of feature extraction and measurement matching, then was proved to be advanced.

1. Introduction
Person Re-Identification is a technique that uses computer vision to retrieve images or video sequences across devices. Assuming that an elderly or a child was lost in the railway station, the traditional method is to locate the location and time of the old people or children through the video surveillance system of these venues, so as to find them. Person Re-Identification is the technology used to achieve this purpose, making people free from the heavy task of observation and search. It can be combined with pedestrian detection / pedestrian tracking technology, and can be used in intelligent video surveillance, intelligent security, intelligent criminal investigation and other fields, using multi-camera, cross-view network, combined with pedestrian detection / pedestrian tracking technology [1-3]. Since Gray put forward a database VIPeR [4], which is of great significance to pedestrian recognition in 2007, more and more scholars have begun to pay attention to the research of Person Re-id. After the application of deep learning to Person Re-id in 2014, the technology of Person Re-id has been developed in blowout style, and the performance of network recognition has exceeded the recognition level of human beings [5]. The neural network is used for pedestrian recognition, which requires high equipment, and has the problem of poor performance of database migration, so it does not have the generalization ability. Although the performance of traditional pedestrian recognition method is not as good as that of neural network, but its requirement of equipment is lower. The amount of feature extraction is less, and the training time and matching time are relatively less. It can be applied to practice to some extent. The purpose of this paper is to reduce the dimension of feature extraction and improve the performance of similarity matching. This paper is based on the research of VIPeR dataset [4]. We extract image color pairs and SILTP texture features, and use the improved XQDA metric (CXQDA) to measure the images. This method were compared with other hand-crafted methods of feature extraction and measurement matching, then was proved to be advanced.

2. Related work

2.1 Feature extraction
For feature representation, several effective approaches have been proposed. Shengcai Liao[6] cut the image horizontally, extracted the local feature by sliding block, and compared the local feature one by one. Liang Zheng [7] makes use of the detection algorithm of human body parts (hands, legs, torso, etc.) to divide the images into blocks, and then extract features and measure the blocks. In this paper, the images are cut horizontally, but the feature of the block is not calculated by sliding block step by step. The feature of a whole block is counted directly, then the time of feature statistics is reduced.

Color feature is the most widely used visual feature in image retrieval. The main reason is that the color is often related to the object or scene contained in the image. Its advantage is that it is not affected by the rotation and translation of the image, and can not be affected by the scale change of the image with the help of normalization. EH Land [8] describes color feature classification and color invariant feature extraction based on Retinex algorithm. The results show that HSV color space can satisfy the human visual characteristics and has some advantages in pedestrian recognition. In order to reduce the influence of illumination on images, this paper uses the algorithm mentioned by Retinex theory to extract the original color features of images by removing the change of images feature caused by illumination.

The texture feature is a visual feature that reflects the homogeneous phenomenon in images, which embodies the inherent attribute shared by the surface of the object. GL Yang et al. [9] proposed the use of $SILTP$ to transform texture features and initialize them. The more detailed features extracted, the more the image can be distinguished. The local binary patterns feature can better describe the details of the image. At the same time, we have to take into account the impact of light on the image. $SILTP$ texture feature is used in this paper. This feature improves the texture feature of $LBP$. It can reduce the influence of image noise and change the influence of illumination scale on image texture feature.

2.2 Distance metric
Person Re-id is supervised learning. KQ Weinberger [10] proposed an algorithm for the maximum nearest neighbor classification interval. By seeking a classification hyperplane, the distance between the hyperplane and the nearest point is as great as possible. WS Zheng [11] has proposed a method of comparing probabilistic relative distances, which considers relative constraints when learning distance metric functions. The distance of the same class is as smaller as possible, and the distance of different classes is as larger as possible. Liao Shengcai et al. [6] combine Bayesian face with linear discriminant analysis to measure the Mahalanobis distance between sample pairs so as to match images. The features extracted in this paper are color-pairs and $SILTP$ texture features. In order to avoid the poor measurement caused by inconsistency between different features, We use CXQDA algorithm to measure Mahalanobis distance.

3. Our work
3.1 Feature extraction
3.1.1 Dealing with illumination variations. Because of the different equipment, the environmental changes and the different angles of shooting, the colors of the same person in two images are quite different. In order to reduce the influence of color difference on feature extraction, we used Retinex algorithm [12] to preprocess the image to eliminate the influence of uneven illumination. Color features are extracted from $HSV$ color space model. The hue, saturation and value of $HSV$ color space is normalized, and $H$ is divided into 8 grades, $S$ and $V$ are divided into 6 grades, then 288 colors are produced. The synthetic feature $L$ is calculated as the color feature of the current point.

The effect of $SILTP$ texture on noise and illumination scale is improved to some extent. The $SILTP$ function can be represented as:
I is the value of central pixel, I_k represents the surrounding value, S_τ represents the result of a comparison, τ is the parameter.

3.1.2 Dealing with viewpoint changes. When we extract image features, we always want to extract the invariant features between images as much as possible. In figure 1, the invariant features of the same person in different images are shown: the color pairs (CP), which reflects the spatial structure information of the images to a certain extent. In this paper, when extracting this color pair feature, the synthetic color feature is simulated as "gray value" of the point. We can use the gray level co-occurrence matrix to calculate the number of color pairs of this left to right structure. We counts the number of color pairs of each color, which is on the left.

Figure 1. The invariant color pair in the change of pedestrian posture.

The method of feature extraction in this paper is horizontal cutting. It takes 10 pixels as a layer and 5 pixels as a stride. Color pairs features and SILTP texture features with a radius of 3 and 5 of each layer are counted. In order to obtain low frequency characteristics, this paper carries out two down-sampling. The total statistical characteristics is \((288+81+81)*(24+11+5)=18000\).

3.2 Distance metric

In this paper, Mahalanobis distance is used to measure the similarity between two samples. Gaussian model is used to fit the intra class covariance matrix (inCov) and the external class covariance matrix(exCov). Suppose \(Δ_i = x_i - x_j\) represents the difference between two sample features. Similar sample \(Δ\) accord with \(Ω_i\) Gaussian distribution, different samples \(Δ\) accord with \(Ω_e\) Gaussian distribution, the mean value is 0.

\[
P(Δ | Ω_i) = \frac{1}{(2π)^{d/2} |Σ_i|^{1/2}} e^{-\frac{1}{2} Δ Σ_i^{-1} Δ'}
\]

\[
P(Δ | Ω_e) = \frac{1}{(2π)^{d/2} |Σ_e|^{1/2}} e^{-\frac{1}{2} Δ Σ_e^{-1} Δ'}
\]

\(Σ_i\) is the inCov, \(Σ_e\) is the exCov.

The Mahalanobis distance between the two samples is:

\[
d(x_i - x_j) = (x_i - x_j)' (Σ_i^{-1} - Σ_e^{-1}) (x_i - x_j)
\]

In order to reduce the inCov, increase the exCov and obtain transformation matrix \(W\), we use generalized eigenvalue decomposition to find the optimal solution \(Σ_i W = Σ_e W\). In this process of solving the Mahalanobis distance, The inCov is required to be reversible. To solve the problem of irreversibility of inCov in small samples set, a minimum value \(λ\) is added to the diagonal element of inCov by method of adding disturbance \([13]\), which makes the inCov reversible. In the process of solving exCov, the LDA method retains large values and neglects small values. It is easy to classify classes with large exCov, and it is difficult to distinguish between classes with small exCov. To solve this problem, we add a close function to reduce large exCov, and to increase small exCov. The close
function is:

\[
\text{Close}(\bar{x}_i, \bar{x}) = \frac{\sum_{j=1}^{d} e^{-\frac{||x_i - x_j||^2}{d}}}{d}
\]  

(4)

\(\bar{x}_i\) is the mean of class \(i\), \(\bar{x}\) is the average sample population, \(d\) is the sample dimension. We redefine \(\text{exCov}\) as:

\[
\Sigma_{E} = \frac{1}{N} \sum_{i=1}^{n} n_i \text{Close}(\bar{x}_i, \bar{x}) * (x_i - \bar{x})(x_i - \bar{x})^T
\]  

(5)

\(N\) is the total sample number, \(n_i\) is the number of class \(i\) samples. Then we use the trained transformation matrix \(W\), the \(\text{inCov}\) and the \(\text{exCov}\) to test on the new sample set.

4. Experiments

The experiments are carried out on the dataset VIPeR [4]. Our method is compared with LOMO+XQDA method [6]. Half of 632 pairs of features are selected as training set and the other half as testing set. We repeat the process 10 times, and get average as performance.

![Figure 2. CPSILTP+CXQDA and LOMO+XQDA on dataset VIPeR [4].](image)

As can be seen from figure 2, the performance of this method in dataset VIPeR [4] is slightly better than that in LOMO+XQDA. Therefore, the method is feasible.

| Method            | Rank=1   | Rank=10  | Rank=20  | Reference          |
|-------------------|----------|----------|----------|--------------------|
| CPSILTP+CXQDA     | 40.28    | 81.49    | 92.06    | Proposed           |
| LOMO+XQDA         | 40.00    | 80.51    | 91.08    | CVPR 2015[6]       |
| SCNCD             | 37.80    | 81.20    | 90.40    | ECCV2014[14]       |
| kBiCov            | 31.11    | 70.71    | 82.45    | IVC2014[11]        |
| LADF              | 30.22    | 78.92    | 90.44    | CVPR 2013[15]      |

Table 1 shows the superiority of this method in performance.

5. Summary and future work

In this paper, an efficient and feasible feature extraction method for Person Re-id is proposed, which can be called CPSILTP. The extracted features are robust and can overcome the influence of light and visual angle changes. It shows good performance on dataset VIPeR [4]. Its characteristic dimension is less than LOMO+XQDA. CXQDA distance measurement is also effective, it provides a direction for the improvement of distance measurement algorithm. In this paper, we use the 0° direction statistics in color pairs statistics. We have not statistical analysis of the gray level co-occurrence matrix in the three directions of 45°, 90° and 135°. If these features are added, whether it can improve the performance of feature matching is also worth considering.
References:

[1] LI Y, WU Z, KARANAM S, et al. Real-world re-identification in an airport camera network[C]. International Conference on Distributed Smart Cameras. Venice, Italy. 2014: 35.

[2] GONG S, CRISTANI M, YAN S, et al. Person re-identification [M]. London, UK: Springer, 2014.

[3] CAMPS O, GOU M, HEBBLE T, et al. From the lab to the real world: Re-identification in an airport camera network [J]. IEEE transactions on circuits and systems for video technology, 2016, (99): 540–553.

[4] GRAY D, BRENNAN S and TAO H. Evaluating appearance models for recognition, reacquisition, and tracking[J]. International journal of computer vision, 2007, 89(2): 56–68.

[5] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun. AlignedReID: Surpassing Human-Level Performance in Person Re-Identification. arXiv:1711.08184, 2017.

[6] Shengcai Liao, Yang Hu, Xiangyu Zhu, and Stan Z. Li. Person Re-identification by Local Maximal Occurrence Representation and Metric Learning. IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2015), Boston, USA, June 2015.

[7] Liang Zheng, Yujia Huang, Huchuan Lu and Yi Yang. Pose Invariant Embedding for Deep Person Re-identification. arXiv:1701.07732, 2017.

[8] EH Land. The retinex theory of color vision. Scientific American, 1977, 237(6):108.

[9] GL Yang, D Zhou and JH Zhang. Moving Object Detection Algorithm Using SILTP Texture Information. Computer Science, 2014.

[10] KQ Weinberger and LK Saul. Distance Metric Learning for Large Margin Nearest Neighbor Classification. Journal of Machine Learning Research, 2009, 10(1):207-244.

[11] WS Zheng, S Gong, T Xiang. Person re-identification by probabilistic relative distance comparison. Computer Vision & Pattern Recognition, 2011, 42(7):649-656.

[12] D. J. Jobson, Z.-U. Rahman, and G. A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. Image Processing, IEEE Transactions on, 6(7):965–976, 1997.2.

[13] Hong Z Q and Yang J Y. Optimal discriminant plane for a small number of samples and design method of classifier on the plane[J]. Pattern Recognition,1991.24(4):317-324.

[14] Y. Yang, J. Yang, J. Yan, S. Liao, D. Yi, and S. Z. Li. Salient color names for person re-identification. In Proceedings of the European Conference on Computer Vision, 2014. 5, 6.

[15] Z. Li, S. Chang, F. Liang, T. S. Huang, L. Cao, and J. R. Smith. Learning locally-adaptive decision functions for person verification. In IEEE Conference on Computer Vision and Pattern Recognition, 2013. 1, 2, 5, 6.