A PERSON RE-IDENTIFICATION ALGORITHM BY USING REGION-BASED FEATURE SELECTION AND FEATURE FUSION

Yanbing Geng[1,2]*, Hai-Miao Hu*[1]*, Jin Zheng*[1]*, Bo Li*[1]

1. Beijing Key Laboratory of Digital Media, School of Computer Science and Engineering, Beihang University, Beijing 100191, China
2. College of Electronic and Computer Science and Technology, North University of China, Taiyuan, Shanxi 030051, China

ABSTRACT

In outdoor surveillance, person appearances captured by different cameras have obvious variations due to different poses and viewpoints, which affect the accuracy of person re-identification. In this paper, a person re-identification algorithm by using region-based feature selection and future fusion is proposed to divide one body into the upper region and the lower region. According to their different characteristics, each region adopts different kinds of features, which can efficiently reduce the negative impact from different poses and viewpoints. Moreover, since different features of one region may have different intrinsic meanings, during the feature fusion, different features of one region are separately represented instead of being comprehensively processed. The proposed feature fusion can make full use of the salience of different features. The experimental results demonstrate that the proposed algorithm improves the accuracy of person re-identification compared with the state of the art.

Index Terms—*person re-identification, outdoor surveillance, region-based, feature selection, feature fusion*

1. INTRODUCTION

Person re-identification is the task of establishing correspondences between observations of the same person in different videos and it has been widely applied in video surveillance. Due to the complex background and the low quality of objects in outdoor surveillance, biometric features, such as face and gait, are not easily extracted. And they are not reliable for person re-identification, either. Therefore, the appearance-based person re-identification is widely used to achieve an accurate recognition ratio [1-4]. Omar [1] considers the spatial information of each person. It adopts the Mean-Shift segmentation method to divide a body into several sub-regions, and processes a cross-match between the sub-regions of voter and the sub-regions of candidate. However, this method ignores the structure information of person bodies, which easily leads to the mismatch between torsos and legs. In order to deal with the above problem, the structure information is utilized as the spatial constraint to divide a body into different regions (e.g., head, torso and legs) [2, 3]. Different regions are implicitly isolated and the sub-regions located in different regions are matched individually, which can efficiently reduce the mismatch between different regions. Moreover, its accuracy of identification is improved by employing multi-feature fusion methods [1, 2, 3].

However, the above methods adopt the same features to describe different regions and it is assumed that these features are optimal for all regions. Since the person appearances captured by different cameras have obvious variations of pose and viewpoint, different features should be adopted for each region according to their respective characteristics. For example, since leg postures can changed easily, texture features (e.g., Local Binary Patterns, LBP) are suitable to describe leg regions, while oriented gradient features (e.g., Histograms of Oriented Gradients, HOG) are suitable for representing torso regions since torso regions are relatively stable during normal walking.

On the other hand, since there are various intrinsic meanings for different features, literatures [1, 2, 3] fail to differentiate the salience among different features during the process of feature fusion, which cannot make full use of the contribution of each feature. The above two observations will be further discussed in details through the experiments in next section.

In order to improve the accuracy of person re-identification, this paper proposes a region-based feature selection and feature fusion algorithm aiming at outdoor surveillance. According to their respective characteristics, different features are adopted for different regions. Furthermore, during the process of feature fusion, different features for each region are separately represented, which can make full use of the salience of different features. The experimental results demonstrate that the proposed

This work was partially supported by the National Science Fund for Distinguished Young Scholars (No.61125206) and the 973 Program (No. 2010CB327900).

Corresponding author: Hai-Miao Hu, frank0139@163.com

algorithm significantly improves the accuracy of person re-identification compared with the similar algorithms.

The remainder of this paper is organized as follows. Section 2 will gives some observations. Section 3 describes the proposed method in details. Section 4 evaluates the proposed method by comparing with previous algorithms. Finally, the paper is concluded in Section 5.

2. OBSERVATIONS

This paper has mainly two contributions based on the two observations.

A. Feature Selection

Different regions have different characteristics. It is obvious from Fig. 1 that the lower region varies seriously whereas the upper region maintains relatively stable in general walking behavior. It is necessary for different regions to adopt different features [5]. This conclusion can be proved through the following experiment.

In this experiment, for the persons of ETZH SQE 2 database [8], a body is divided into the upper region and the lower region by using adaptive body segmentation [2], each region is further divided into several sub-regions with Mean-shift [2]. Fig. 1 shows the separated regions and sub-regions. The HSV, HOG and combined features are selected to describe sub-regions of upper region and the lower region, respectively. Table 1 reflects the rank 1 correct matches. From Table 1, it can be observed that the color feature is valid for two regions, yet, the edge feature (e.g., HOG) is more effective for the upper region than for the lower region, and the texture feature (e.g., LBP) is more suitable for the lower region than for the upper region.

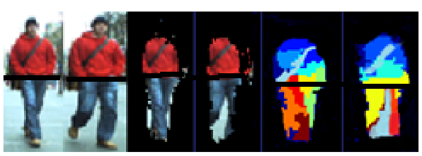


Fig. 1. Appearance discrepancy between the upper region and lower region of the same person.

Table 1. The rank 1 correct matches when using different feature on different regions.

SEQ.2	**Upper Region**	**Lower Region**
HSV	***75.4%***	60.2%
HOG	***65.7%***	46.6%
LBP	42.3%	***55.7%***

The experimental result demonstrates the above observation，adopting different features aiming at different regions can improve the accuracy of person re-identification.

B. Feature fusion

For each region, the intrinsic meanings of different features are different [6], which can be observed from Fig. 2, wherein ***sr*** represents one of the sub-regions of legs. ***A*** represents the hue values (HSV channels) of one sub-region, ***B*** is the normalized LBP histogram of the same sub-region. Comparing with ***A***, the data of ***B*** is sparse. Eq. (1) is the feature fusion method in [2], wherein the combined HSV and LBP descriptors of the sub-regions in the lower region are directly inputted Principal Components Analysis (PCA) for dimensionality reduction and noise data remove. However, during this process, as one of the combined feature, the LBP feature is possible to be removed as noise due to its sparsity, which can be proved through the following experiment.

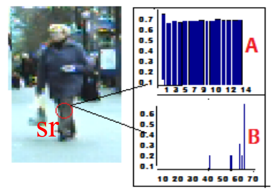


Fig. 2. Different feature descriptor of the sub-region.

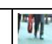
In this experiment, 6 persons are sampled from ETHZ SEQ. 3 database [8], and acquire the ranks of correct matches for the lower region by using different features. The ranks of correct matches are adopted to evaluate the efficiency of different features. The feature fusion is described in Eq.(1).

$$fs(R) = PCA\left(\bigcup\nolimits_{k=1}^{m} fd(SR_k)\right) \qquad (1)$$

where $fd(.)$ refers to the selected single feature descriptor (e.g., HSV, HOG or LBP), SR_k refers to the k_{th} sub-region in the region R, $fs(.)$ refers to apply $PCA(.)$ on combined features directly.

Rank of correct matches are shown in Table 2, the results are better by using LBP descriptor other than by using HSV, however, by adopting the HSV and LBP features which are fused through Eq. (1), the results are similar to the results by using only HSV, which indicates that the salience of LBP is not considered.

Table 2. The rank of correct matches when using different features.

Lower Region						
HSV	1	2	3	1	8	6
LBP	1	1	2	1	1	7
HSV+LBP(Eq.1)	1	2	4	1	5	5

Therefore, based on the above observation, different features for each region are separately represented to make full use of the salience of different features.

3. THE PROPOSED METHOD

According to the above observations, this paper proposes a person re-identification algorithm by using region-based feature selection and feature fusion.

A. The proposed region-based feature selection method

Several experiments are carried to evaluate the different characteristics of different regions. We randomly sampled 12 persons from each sequel of ETHZ three sequels database, and we select 6 images for each person, 1 for testing data and 5 for training data, where each person is divided into the upper region and lower region, and each region is segmented into several sub-regions.

The upper regions have plenty of color information. And the color feature (e.g., HSV) is an important feature for the upper regions. However, single feature is far from enough and other kind of feature should be taken into consideration. Since the upper regions are relatively stable and have slight posture change, the HOG is suitable to describe the edge information.

The above conclusion is demonstrated by the following experiment. Rank 1 correct match is shown in Table 3, wherein HSV, HOG and LBP are separately adopted. It can be observed that, for the upper regions, the average recognition ratios of adopting HSV and HOG descriptor are 84% and 71%, respectively, they are obvious higher than 57% of the average recognition ratio of adopting LBP feature, the HSV and HOG features are better than LBP feature. Therefore, the upper regions adopt both the HSV and the HOG features in the proposed algorithm.

Table 3. The rank 1 correct matches when using different feature on upper region.

Upper region	HSV	HOG	LBP
SEQ.1	87%	71%	63%
SEQ.2	75%	65%	42%
SEQ.3	91%	78%	67%
Avg.	***84%***	***71%***	***57%***

For the lower regions, HSV is also adopted to describe the color information. Moreover, since the lower regions may have obvious posture change and the edge of the lower regions has obviously variations, the HOG feature is not suitable for the lower region. On the other hand, LBP feature can adapt to the posture change and it is suitable to describe the texture information of the lower regions.

The above conclusion is demonstrated by the following experiment. The experimental result is shown in Table 4. For the lower region, it is obvious that the HSV and LBP features are better than HOG feature.

Therefore, the lower regions adopt both the HSV and the LBP features in the proposed algorithm. The proposed feature fusion will be introduced in the next section.

Table 4. The rank 1 correct matches when using different feature on lower region.

Lower region	HSV	HOG	LBP
SEQ 1	78%	50%	69%
SEQ 2	60%	46%	55%
SEQ 3	88%	53%	75%
Avg.	***75%***	***50%***	***66%***

B. The proposed feature fusion method

In order to differentiate the salience among different features, *PCA* is applied on each feature descriptor for each region during the feature fusion, then weighted combined is adopted, as shown in Eq.(2).

$$fs(R)=\begin{cases}\bigcup(\alpha\cdot PCA(\bigcup_{k=1}^{m} HSV(SR_k)),\ \beta\cdot PCA(\bigcup_{k=1}^{m} HOG(SR_k)),\ SR_k\in U \\ \bigcup(\alpha\cdot PCA(\bigcup_{j=1}^{n} HSV(SR_j)),\ \beta\cdot PCA(\bigcup_{j=1}^{n} LBP(SR_j)),\ SR_j\in L\end{cases} \qquad (2)$$

where α or β is the normalized weight, U and SR_k refer to the upper region and its m sub-regions, L and SR_j refer to the lower region and its n sub-regions.

Finally, the above-mentioned methods are evaluated by using Earth Mover Distance (EMD) [2]. The distance between correspondent regions is computed.

$$d(R_{va},R_{cb})=\frac{\sum_{i=1}^{m}\sum_{j=1}^{n}c_{ij}f_{ij}}{\sum_{i=1}^{m}\sum_{j=1}^{n}f_{ij}},R\in(U,L) \qquad (3)$$

where f_{ij} refers to the number of pixel matched from the *i*th sub-region of each region voter *va* to the *j*th region of candidate *cb*. Corresponding c_{ij} refers to the ground distance between two sub-regions [2].

The whole body EMD dissimilarity value is calculated according to the EMD values weighted sum of the upper region and lower one, as shown in Eq.(4):

$$d(va,cb)=\alpha\cdot d(U_{va},U_{cb})+\beta\cdot d(L_{va},L_{cb}) \qquad (4)$$

where *va* and *cb* represent the voter and candidate image, $d(U_{va}, U_{cb})$ and $d\ (L_{va}, L_{cb})$ are achieved by using Eq.(4), and are the normalized weight, where $\alpha+\beta=1$.

4. EXPERIMENTAL RESULTS

In this section, the performance of the proposed algorithm is evaluated based on three sequences of (ETHZ) dataset [8]. This dataset contains three image series: SEQ. 1 contains 4857 images for 83 persons, SEQ. 2 contains 1936 images for 35 persons, and SEQ. 3 contains 1762 images for 28 persons. Wherein the image sizes are between 30*80 and 181*402 pixel, and different images of one person have obvious posture changes and serious occlusions.

In this experiment, 6 different images of the same person are random selected, 5 for training data and 1 for testing data, the average performance over 10 trials is employed as the identification performance. For retaining the proportion unchanged of the image size, all images are normalized to 79 pixels wide. The performance is evaluated through the Cumulative Matching Characteristic (CMC) [7], which represents the probability of finding the correct match in the top n candidates. The proposed algorithm is compared with the algorithm proposed in [2], which adopt the color value of HSV color space and the HOG feature descriptor.

Firstly, the matching performance for each region is evaluated by comparing with the algorithm proposed in [2], which select the same features for both the upper region and the lower region. For upper region, the rank 1 correct

Table 5. The rank 1 correct matches for upper region and lower region

Upper region	HSV+HOG [2]	HSV+HOG (our method)	Lower region	HSV+HOG [2]	HSV+LBP （our method）
SEQ. 1	82.5%	90.60%	SEQ. 1	76.2%	81.90%
SEQ. 2	79.7%	83.40%	SEQ. 2	67.3%	74.80%
SEQ. 3	90.8%	96.70%	SEQ. 3	86.9%	92.10%
Avg.	***84.3%***	***90.2%***	***Avg.***	***76.8%***	***82.9%***
Avg. Gain		***5.90%***	***Avg. Gain***		***6.13%***

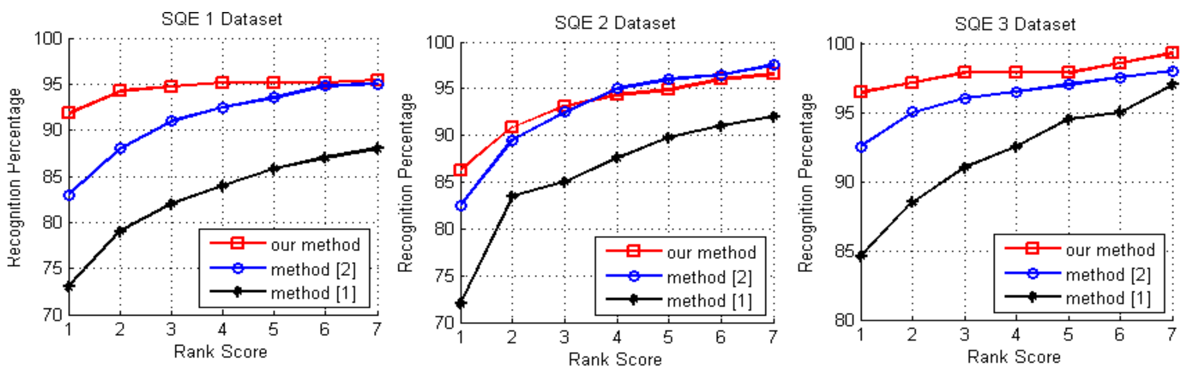


Fig.4. Recognition rate on ETHZ database

matches is shown in the left side of Table 5. Obviously, the proposed algorithm can improve the average matching performance by 5.9%. The similar conclusion can be observed for the lower region, as shown in the right side of Table 5, the average matching performance of the proposed algorithm is improved by 6.13%.

Furthermore, the whole body matching performance is compared with the algorithm [1, 2]. The CMC curves are shown in Fig.4. For SEQ.1, the rank 1 correct matches of the proposed algorithm is improved by 9% when compared with [2]. For SEQ.3, the rank 1 correct matches of the proposed algorithm can achieve 96.7% and it is improved by4.8%. For SEQ.2, since the images are more blurred than the other database, the rank 1 correct matches is 87.6% and also improved by 4.6% when compared with [2]. The experiment results demonstrate that our propsed algorithm is effective in improving the recognition performance campared with the state of the art.

Note that there will be a slight increase of computational complexity during the features extraction and fusion for the upper and lower region, but this increase is negligible. The weights of different regions in Eq.(4) can be adaptively adjusted for different persons which can further improve the performance of the proposed algorithm. This is one of our future works.

5. CONCLUSIONS

This paper proposes a person re-identification algorithm by using region-based feature selection and future fusion. Different features are adopted for different regions according to their characteristics. In addition, the different features of each region are separately represented during the feature fusion, which can make full use of the salience of different features. The experimental results demonstrate that the proposed algorithm achieves the average rank 1 correct matches to 90%.

In order to further improve the accuracy of person re-identification, one of our future works is to refine the proposed feature fusion through adaptively adjusting the weights of different regions according to their different characteristics.

REFERENCES

[1] O.Oreifej, R.Mehran, and M.Shah, “Human identity recognition in aerial images,” IEEE Conference on Computer Vision and Pattern Recognition, pp. 709–716. 2010.

[2] L Ma, X Yang, Y Xu, and J Zhu, ”Human identification using body prior and generalized EMD,” IEEE International Conference on Image Processing, pp.1441-1444, 2011.

[3] Farenzena.M, Bazzani.L, Perina.A, Murino.V, and Cristani.M, “Person re-identification by symmetry-driven accumulation of local features,” IEEE Conference on Computer Vision and Pattern Recognition, 2010

[4] R.Satta, G.Fumera, and F.Roli, “Fast person re-identification based on dissimilarity representations,” Pattern Recognition Letters, pp.1838-1848, 2012.

[5] L.Bazzani, M.Cristani, and V.Murino, “Symmetry-driven accumulation of local features for human characterization and re-identification,” Computer Vision and Image Understanding, PP.130-144, 2013.

[6] B.Fernando, E.Fromont, and D.Muselet, ”Discriminative feature fusion for image classification,” IEEE Conference on Computer Vision and Pattern Recognition, PP.3434-3441, 2012.

[7] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling," IEEE International Conference on Computer Vision, pages 1–8, 2007.

[8] A. Ess, B. Leibe, and L.Van Gool, “Depth and appearance for mobile scene analysis,” IEEE International Conference on Computer Vision, 2007.