

# Multiple-shot People Re-identify based on Feature Selection with Sparsity

Dongping Zhang, Yanjie Li, Jiao Xu and Ye Shen

College of Information Engineering, China Jiliang University, Hangzhou 310018,  
China  
[silenttree\\_zju@cjlu.edu.cn](mailto:silenttree_zju@cjlu.edu.cn)

## Abstract

*In a video surveillance network, it is always required to track and recognize people when they move through the environment. This paper presents a novel re-identification method for multiple-people using feature selection with sparsity. By using the multiple-shot approach, each of appearance models is created in this method. The human body is divided into five parts from which the features of color, height, gradient were extracted respectively. Our appearance model is represented by linear regression method. Experimental results show that our appearance model is robust and attain a high precision rate and processing performance.*

**Keywords:** *people re-identify, multiple-shot, sparsity, feature selection*

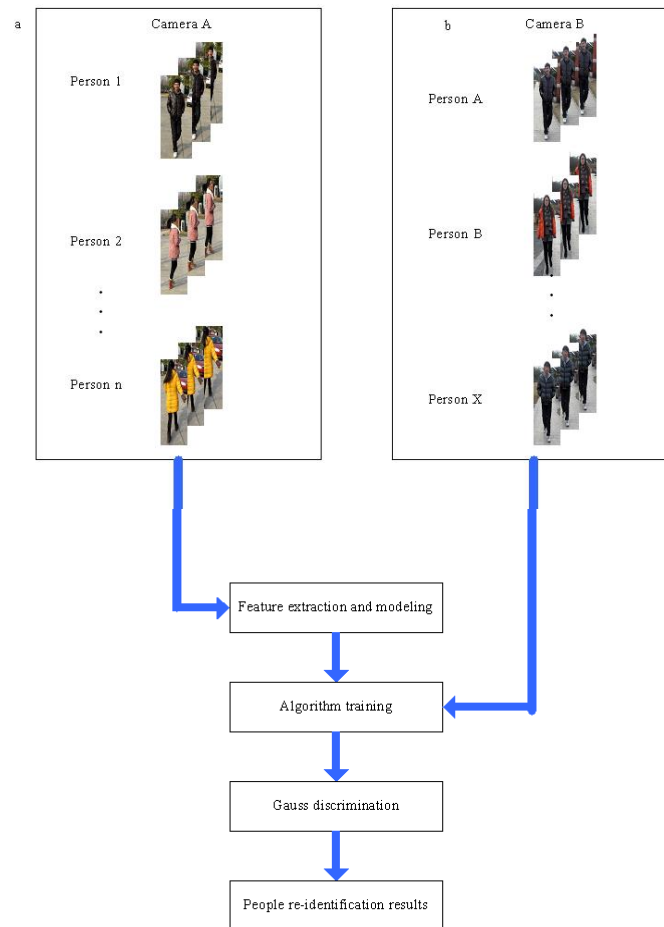
## 1. Introduction

The intelligent video surveillance system, which demands video capture hardware and computer vision technologies to detect and track moving objects, is essential to public safety [1-2]. An intelligent video surveillance system can save a considerable amount of work to mitigate the damage, due to the rapid response and 24 hours tireless monitoring. Nevertheless, detecting suspicious moving objects, *e.g.*, wandering person or a specific object, among overlapping or non-overlapping cameras is still a challenge since the appearance of people gained from different cameras is generally diverse.

Therefore, the identification ability plays an important role in avoiding security vulnerabilities in many surveillance systems, and therefore it strengthens the surveillance applications.

People re-identification involves identify different individuals in the entire camera network. It is a challenging task, because there are a lot of different factors, such as posture of body or camera and ambient lighting conditions [18]. The people re-identification can be divided into two method: single-shot and multiple-shot method. The single-shot method is modeled only using a single image of each individual, such as [3-6], and [10]. The information of person's appearances will be incomplete if we just use the single-shot from the intelligent monitoring system for people re-identification. The multiple-shot is modeled using multiple frames of each person, such as [7-9]. In general, people re-recognition research has focused on pioneering appearance models that are taken into account viewpoint, wearing, pose and illumination variations. Traditionally, some models can identify only one person used a single probe assuming that the probe has been in the gallery.

In this paper, we discuss multiple person re-identification across cameras in the wide area surveillance and treat matching as an assignment problem based on Linear regression model [17]. The model is to solve multiple-people re-identification under multi-camera. We illustrate the analyzing process of our method in Figure 1.



**Figure 1. The Process of our Method: (a) is the Training Data with Three Camera, (b) is the Testing Data with One Camera**

## 2. Related Works

Majority of previous researches in person re-identification are matching task based on improved model features. The feature of color, shape and texture are selected by the researcher generally. There are spatial and temporal distribution used for matching task. Bedagkar-Gala *et al.* 2012 proposed spatio-temporal appearance model which was built by using color and facial features. Texture and color features are combined for appearance model in [9]. Feature extraction based on color and edge features are adopted in [12]. In order to obtain the contour appearance, researchers use edge features generally. The shape and appearance of the semantic model are presented in [13]. Using local and global spatial relationships to describe a person is presented by Zheng *et al.* 2009, [6]. The researchers propose feature extraction based on local interest points used in [14]. Shape model can be obtained to represent basic information of people proposed by Jungling *et al.*, 2011. Bedagkar-Gala *et al.*, 2011 proposed a part-based spatio-temporal model which learns a person appearance model. Bedagkar-Gala *et al.*, 2012 proposed a model of human appearance added the facial features. Farenzena *et al.*, 2010 divided people into five parts, then the corresponding match each local body part.

In this paper, we propose the method that divided human body into different parts to establish appearance model for person re-identification. We prefer more distinctive feature of colors and textures after the person's physical characteristics were extracted. On this basis, we choose height and gradient as the feature of appearance, and use the sparsity matching method for feature selection.

### 3. The Proposed Method

A Surveillance video is composed of a multi-frame image of each person. In order to achieve the purpose of re-identification, linear regression method is exploited to match each part of the body. Video preprocessing is necessary to detect the pedestrians, followed by further detailed analysis of the pedestrians.

#### 3.1. Pedestrian Detection

Pedestrian detection is an important branch in the field of object detection based on image/video data. In this paper, pedestrian detection is regarded as a classification problem with a large number of training samples and a classification model for pedestrian detection has been learned. Taking a video as an example, the results for testing is shown in Figure 2.

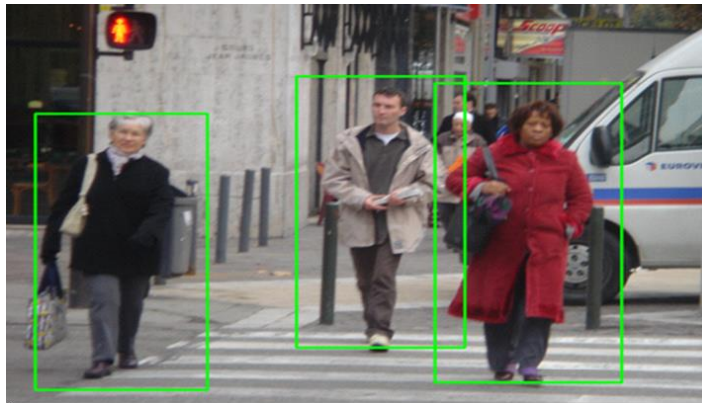


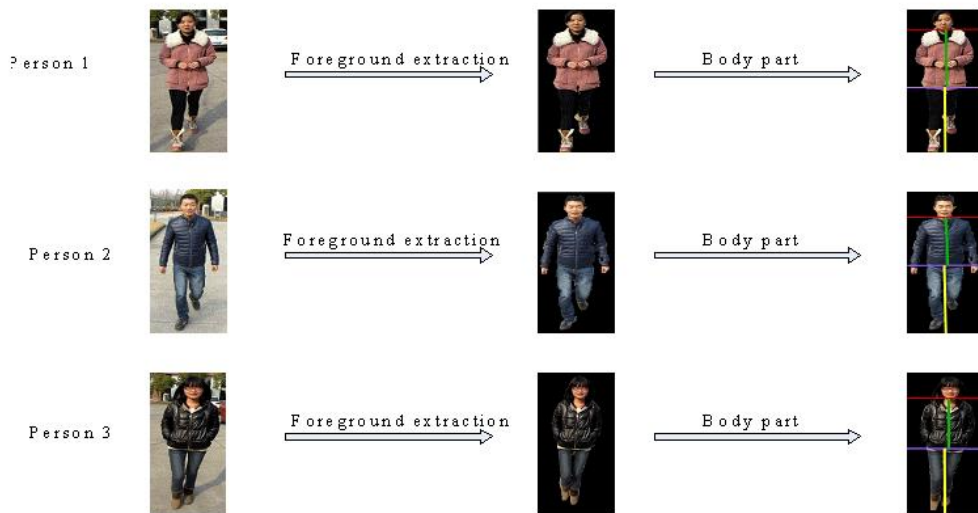
Figure 2. The Result of the Pedestrian Detection for a Video

#### 3.2. Foreground Extraction and Body Part

Background is crucial to object identification and tracking. A background subtraction strategy is enough to receive the silhouette in the multiple-shot. Farenzena *et al.*, 2010 proposed the method to get the information from local body parts. Features match each sub-section of the body which extracted from the foreground image. Taking three persons as examples, the result of foreground extraction and body part is showed in Figure 3.

#### 3.3. Part Appearance Feature Extraction

This paper exploits more expressive and powerful features for each body part. Color is the most intuitive features to target identification. Seen as the most significant feature, texture is an important part of human observation of things. Height has also important significance in the field of human re-identification.



**Figure 3. The Result of Foreground Extraction and Body Part after Pedestrian Detection**

**3.3.1. Color Correlogram:** The traditional color histogram depicts only a proportion of the number of pixels for a certain color and the total number of pixels. Color histogram is a global statistical relationship, however color correlogram is a spatial relationship between the color change and distance. Not only the color correlogram contains image color statistics also includes spatial relationships. Using HSV color space is an effective way to characterize the local body color information.

**3.3.2. Texture Extraction:** This paper uses statistical methods to calculate the texture, the choice of texture features are: GLCM. There are two pixels with a gray scale statistical by GLCM, which maintain a certain distance, respectively.

**3.3.3. Height:** Spatial reference point measurement method uses a set of spatial coordinates and pixel coordinates corresponding to establish two collinear equations The image plane projection matrix calculation determines the target height using the projection matrix on the ground position of object.

### 3.4. Multiple Person Re-identification by Matching

The manifold regularization method is utilized for learning problem. Linear regression model were constructed as:

$$Y = X^T W + 1_n b^T \quad (1)$$

where  $X$  is the features of person appearance represented as  $X = [x_1, x_2, \dots, x_n] \in R^{d \times n}$  and  $Y$  is the ID of each person represented as  $Y = [y_1, y_2, \dots, y_n]^T \in R^{n \times c}$ .  $d$  is the dimension of the feature,  $n$  is the number of the image, and  $c$  is the number of training person.  $Y_{ij}$  denotes the  $j$ th data of  $y_i$  and  $Y_{ij} = 1$  if  $x_i$  in the  $j$ th person.  $L_{2,1}$ -norm based algorithms can be represent as the following objective function:

$$\min_W \text{loss}(W) + \gamma \|W\|_{2,1} \quad (2)$$

Where  $W$  is a projection matrix used for feature selection and  $loss(W)$  is the loss function.  $\gamma$  is a regularization parameter. The definition of  $\|W\|_{2,1}$  is

$$\|W\|_{2,1} = \sum_{i=1}^d \sqrt{\sum_{j=1}^c w_{ij}^2} \quad (3)$$

Least-squares method is a mathematical optimization technique. It matches the data by minimizing the sum of squared errors to find the best function. Using the least square method can easily solve for the unknown data, and make the square error between the actual data and the obtained data minimized. Our algorithm can be denoted as

$$\min_{W,b} \sum \left\| x_i^T W + 1_n b^T - y_i \right\|_2^2 \quad (4)$$

The objective function is optimized by leveraging the Manifold Regularization built upon the graph Laplacian to the loss function in (1), we obtain

$$\arg \min_{W,b} \text{Tr}(W^T X L X^T W) + \mu \left\| X_i^T W + 1_n b^T - Y_i \right\|_F^2 + \gamma \|W\|_{2,1} \quad (5)$$

We expect all training data to contribute to the optimization of  $W$ . Assume that a predicted ID matrix as  $F = [f_1, f_2, \dots, f_n]^T \in R^{n \times c}$ . Following the methodology in Nie *et al.*, 2010 [19], our objective function can be transformed as:

$$\arg \min \text{Tr}(F^T L F) + \text{Tr}((F - Y)^T U (F - Y)) + \gamma \|W\|_{2,1} + \mu \left\| X^T W + 1_n b^T - F \right\|_F^2 \quad (6)$$

#### 4. Experiments and Results

The dataset is obtained from two places. This paper collected 159 individuals of the video, and multiple-shot is used for each person. In this experiment, N images are chosen from each person. Features as the input of our algorithm are extracted from parts of the body for each image after the foreground extracted. Our algorithms were designed to test under different circumstances. Finally, our method was compared with PBSTC [17], SDALF [16], and their recognition rate results are shown in Table 1.

**Table 1. Comparison of Re-identification Accuracy for Different N by our Method, PBSTC, and SDALF**

	Our method	PBSTC	SDALF
Re-identification Accuracy(N=2)	0.7846	0.6654	0.6323
Re-identification Accuracy(N=5)	0.8941	0.7147	0.6452
Re-identification Accuracy(N=10)	0.9175	0.7612	0.6829
Re-identification Accuracy(N=15)	0.9398	0.7811	0.6864
Re-identification Accuracy(N=20)	0.9579	0.7885	0.6891

## 5. Conclusion

In this paper, the proposed appearance model can effectively solve the problem that a person can not be recognized in multi-camera of the different look. Our method using human appearance feature selected for person re-identification is theoretical. Multiple-shot approach was chosen to increase the practicality and accuracy. Finally, our approach and the other two methods for comparison. Experimental results show that our method has better performance on human re-identification, which made great contributions in the National Public Security.

## Acknowledgements

This work was supported by Zhejiang Provincial Science & Technology Research Program (Grant No. 2013C33046, 2014C33075) and Zhejiang Provincial NSF (Grant No.Y1110506, LY13H180011).

## References

- [1] W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares [C]", Computer Graphics and Image Processing (SIBGRAPI), 2009 XXII Brazilian Symposium on. IEEE, (2009), pp. 322-329.
- [2] A. Yilmaz, O. Javed and M. Shah, "Object tracking: A survey [J]", ACM Computing Surveys (CSUR), vol. 38, no. 4, (2006), p. 13.
- [3] B. Prosser, W. S. Zheng and S. Gong, "Person Re-Identification by Support Vector Ranking [C]", BMVC, vol. 2, no. 5, (2010), p. 6.
- [4] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features [M]", Computer Vision–ECCV 2008. Springer Berlin Heidelberg, (2008), pp. 262-275.
- [5] S. Bak, E. Corvee and F. Brémont, "Person re-identification using spatial covariance regions of human body parts [C]", Advanced Video and Signal Based Surveillance (AVSS), 2010 Seventh IEEE International Conference on. IEEE, (2010), pp. 435-440.
- [6] W. Zheng, S. Gong and T. Xiang, "Associating groups of people", in: British Machine Vision Conference, (2009).
- [7] N. D. Bird, O. Masoud and N. P. Papanikolopoulos, "Detection of loitering individuals in public transportation areas [J]", Intelligent Transportation Systems, IEEE Transactions on, vol. 6, no. 2, (2005), pp. 167-177.
- [8] O. Hamdoun, F. Moutarde and B. Stanculescu, "Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences [C]", /Distributed Smart Cameras, 2008. ICDS 2008. Second ACM/IEEE International Conference on. IEEE, (2008), pp. 1-6.
- [9] M. Farenzena, L. Bazzani and A. Perina, "Person re-identification by symmetry-driven accumulation of local features [C]", Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, (2010), pp. 2360-2367.
- [10] M. Hirzer, P. M. Roth and H. Bischof, "Person Re-Identification by Efficient Impostor-based Metric Learning [C]", Advanced Video and Signal-Based Surveillance (AVSS), 2012 IEEE Ninth International Conference on. IEEE, (2012), pp. 203-208.
- [11] A. Bedagkar-Gala and S. K. Shah, "Part-based spatio-temporal model for multi-person re-identification [J]", Pattern Recognition Letters, vol. 33, no. 14, (2012), pp. 1908-1915.
- [12] J. Kang, I. Cohen and G. Medioni, "Object reacquisition using invariant appearance model [C]", Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on. IEEE, vol. 4, (2004), pp. 759-762.
- [13] X. Wang, G. Doretto and T. Sebastian, "Shape and appearance context modeling [C]", Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on. IEEE, (2007), pp. 1-8.
- [14] C. Arth, C. Leistner and H. Bischof, "Object reacquisition and tracking in large-scale smart camera networks [C]", Distributed Smart Cameras, 2007. ICDS'07, First ACM/IEEE International Conference on. IEEE, (2007), pp. 156-163.
- [15] K. Jungling, C. Bodensteiner and M. Arens, "Person re-identification in multi-camera networks [C]", Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on. IEEE, (2011), pp. 55-61.
- [16] M. Farenzena, L. Bazzani and A. Perina, "Person re-identification by symmetry-driven accumulation of local features [C]", Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on. IEEE, (2010), pp. 2360-2367.

- [17] A. Bedagkar-Gala and S. K. Shah, "Multiple person re-identification using part based spatio-temporal color appearance model [C]", Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on. IEEE, (2011), pp. 1721-1728.
- [18] A. Bialkowski, S. Denman and P. Lucey, "A database for person re-identification in multi-camera surveillance networks [C]", Proceedings of the 2012 International Conference on Digital Image Computing Techniques and Applications (DICTA 12). IEEE, (2012), pp. 1-8.

## Authors



**Dongping Zhang**, he was born in 1970. He received the Ph.D. in Information & Communication Engineering from Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2006. Since 2006, He is an associate professor at College of Information Engineering, China Jiliang University in Hangzhou. His research interests include image processing and pattern recognition, computer vision and videos.



**Yanjie Li**, she received the bachelor degree in College of Information Engineering from China Jiliang University, China, in 2014. She is pursuing a master's degree from China Jiliang University, China, at present. Her research interests include machine learning, pattern recognition, data mining, and image processing



**Jiao Xu**, she graduated at the China Jiliang University in Hangzhou in 2012. Currently she is a MSc. student at the College of Information Engineering of China Jiliang University. Her research interests are oriented to image processing and image compression for security and multimedia applied imaging systems.



**Shen Ye**, he received the Ph.D. degree in Biomedical Engineering from Zhejiang University, China in 2014. He is an associate professor in the department of signal and information processing from China Jiliang University, China at present. His research interests are in machine learning, data mining, computer-aided diagnosis.

