International Journal of Research in Engineering, Science and Management
Volume-3, Issue-3, March-2020
www.ijresm.com | ISSN (Online): 2581-5792

A Literature Survey on Person Re-Identification

Vyshna Radh[1], Smitha Suresh[2]
[1]*M.Tech. Student, Department of Computer Science and Engineering, Sree Narayana Gurukulam College of Engineering, Ernakulam, India*
[2]*Professor, Department of Computer Science and Engineering, Sree Narayana Gurukulam College of Engineering, Ernakulam, India*

***Abstract*: Person re-identification is an important topic in surveillance. Re-identification is the process by which anonymous personal information is matched with its true owner. Person re-identification is the process in which pedestrians can be identified by the images that have been captured by different cameras. Existing person re-identification problem focuses on matching persons captured between different true-color and grayscale videos. Grayscale video consists of single channel with 256 levels. True-color video consists of three channels with 256*256*256 levels. Grayscale images have two colors, generally black and white. True-color is used to display RGB images. In this paper, we have done the literature survey about person re-identification. And based on the literature survey, we have done a comparative analysis.**

***Keywords*: Person re-identification.**

1. Introduction

Surveillance is the monitoring of behavior or activities of a person or an object. It can be done by using electronic equipment such as closed-circuit television (CCTV). Person re-identification is the process in which pedestrians can be identified by the images that have been captured by different cameras. Re-identification is an important for surveillance system.

The existing person re-id methods can be divided into two categories: methods based on feature learning and methods based on distance learning [1]. This paper mainly focuses on matching persons captured by true color and grayscale videos. The grayscale images differ from color images is that less information needs to be provided for each pixel i.e. each pixel in grayscale image is store with 8 bits. True-color image consists of three channels: R (Red), G (Green), and B (Blue). Grayscale image consist of one channel: Shades of gray. And grayscale images have two colors: black and white.

This paper is organized as follows. In section 2, literature survey is described. Section 3 presents comparative analysis of person re-identification. Finally, conclusion is delineated in section 4.

2. Literature Survey

A. True-color and grayscale video person re-identification

Fei Ma et al. [1] proposed a semi-coupled dictionary pair learning (SDPL) approach. This is the first method for matching person captured between true color and grayscale video. For re-identification method, we want two sets that are probe set and gallery set. First of all, we want to calculate the distance between probe set and gallery set. After that sort the distance in ascending order. The shortest distance is the best matching pair. Here they used a new dataset called CGVID (color to grayscale video person re-identification). They used two cameras for capturing the videos of the pedestrians. Here camera A is captured by true-color video and camera B is captured by grayscale video. First we apply feature set extraction. Here features used are corners etc. Then we apply asymmetric within-video projection matrices. Asymmetric within-video projection matrices are used to reduce the variations that are caused due to the occlusions. Then SDPL updates dictionary learning and semi-coupled mapping matrices. The semi-coupled mapping matrices will reduce the gap between features of true-color and grayscale videos.

B. Salient color names for person re-identification

Y. Yang et al. [2] proposed a salient color names based color descriptor (SCNCD) to describe colors. Person re-identification can be classified into two stages: feature representation and person matching [2]. This paper mainly focuses on feature representation. Color and texture are the two features used here. In image, the RGB value is mapped to the color name. The mapping function used here is probability distribution. Higher probability will be assigned to the color name which is nearer to the color [2]. The advantage of SCNCD is good robustness against photometric invariance. Photometric invariance refers to detecting the true shape and color of an object without considering the light sources. And disadvantage is processing speed is comparatively less.

C. Adaptive graph representation learning for video person re-id

Yiming Wu et al. [3] proposed an adaptive graph representation learning for video person Re-ID, which enables the contextual interactions between the relevant regional features. The advantage for this method is it is consistent among different temporal resolution for same identities. The disadvantage of this method is that it exploits pose alignment connection and feature affinity connection.

D.Person re-identification via ranking aggregation

Mang Ye et al. [4] proposed a ranking aggregation algorithm. Existing methods focuses only on the similarities between probe image and gallery image. And it doesn't consider the dissimilarities between them. But this method focuses on both similarities and dissimilarities. The main idea behind this method is that it contains an original ranking list. The original ranking list consists of probe image, strongly similar galleries, neutral galleries, strongly dissimilar galleries. If the probe images matches with the strongly similar galleries, then it will be pulled. If the probe images matches with the strongly dissimilar galleries, then it will be pushed.

E. Unified sparse subspace learning via self-contained regression

Shuangyan Yi et al. [5] proposed a joint sparse pixel weighted PCA method. The advantage of this method is it enhances robustness. And the disadvantage is that it is not good for classification due to redundant features.

F. Learning deep context-aware features for person re-identification

Dangwei Li et al. [6] proposed a multi-scale context aware network (MSCAN) to learn powerful features over full body and body parts. The advantage of this method is that it learns the powerful features over full body and body parts. The disadvantage of this method is that we cannot directly train this model on small datasets because it would be easily over fitting and insufficient to learn such a large capacity network on small datasets.

G.Recurrent convolutional network for video person re-identification

Niall McLaughlin et al. [7] proposed recurrent neural network architecture for video-based person re-id. The advantage of this approach is that it makes use of color and optical flow information in order to compute appearance and motion information. The disadvantage of the approach is that it is difficult to create an accurate appearance model.

H.Person re-id by descriptive and discriminative classification

Martin Hirzer et al. [8] proposed a method to combine both the descriptive and discriminative models. Here two cameras are used to detect the person from different views. The image of selected person is referred as probe image. And the image searched is referred as gallery images. The descriptive model is applied to the probe image to get the initial ranking. Then the first 50 images is send to a human operator. The human operator will find whether the matching has found or not. If the matching is not found, then we apply discriminative model to get refined ranking.

I. Person re-id by local maximal occurrence representation and metric learning

Shengcai Liao et al. [9] proposed a retinex algorithm to pre-process person images. The advantage of using retinex algorithm is that it improves brightness, contrast, sharpness of an image through dynamic range compression. The disadvantage of using retinex algorithm is that it either provides dynamic range compression or tonal rendition not both and it also suffers color distortion.

J. Fisher discrimination dictionary learning for sparse representation

Meng Yang et al. [10] proposed a fisher discrimination dictionary learning (FDDL) approach to sparse representation based image classification. FDDL method is applied to face, digit and gender recognition to evaluate its performance [10]. The advantage of this method is that it improves pattern classification. And the disadvantage of this method is that it is time consuming.

K.Person re-identification by video ranking

Taiqing Wang et al. [11] proposed a model for selecting and matching discriminative video fragments from a pair of image sequences. This model is based on flow energy profile and histogram of oriented gradient (HOG). HOG is used for object detection. HOG is used to extract the features from image.

In this model, there will be an image sequence captured by the camera. Then the image sequence will be broken down and generates a candidate fragment pools by flow energy profile (FEP). We can observe the two legs of a walking person. Each will be having different movement. This is called FEP. Then we will group the image into positive instance and negative instance. According to that we will select and rank the most discriminative fragments. The advantage of this method is that it is capable of capturing more accurate space-time features.

L. Scalable person re-identification: A Benchmark

Liang Zheng et al. [12] proposed a Bag of Words (BOW) descriptor to bridge the gap between person re-identification and image search. BOW produces relatively low accuracy. For feature extraction, BOW model is used. Here Market-1501 dataset is contributed. It consists of high quality images. It consists of 1501 images collected by 6 cameras. Market-1501 uses deformable part model (DPM) as pedestrian detector. In DPM, we can model the human face as two eyes, a mouth and a nose.

M. RGB-infrared cross-modality person re-identification

Ancong Wu et al. [13] addressed the RGB-IR cross-modality Re-identification problem and contribute a dataset named SYSU-MM01, including RGB and IR images. This is the only method for matching RGB and IR images. The disadvantage of this method is that it is a challenging task due to large cross modality variations between RGB and IR images.

N.MARS: A video benchmark for large scale person re-identification

Liang Zheng et al. [14] introduced a video re-identification dataset called Motion Analysis and Re-identification Set (MARS). It is the video extension of the Market- 1501 dataset.

3. Comparative Analysis

Topic	Method used	Advantages	Disadvantages
True-color and grayscale video person re-identification	Semi-coupled dictionary pair learning (SDPL)	Saves storage space.	Only focuses on matching person between true-color and grayscale videos.
Salient color name for Person re-identification	Salient color Name based Color Descriptor (SCND)	Good robustness against photometric invariance	Processing speed is comparatively less.
Adaptive graph representation learning for video person re-id	Adaptive graph learning scheme	Consistency among different temporal resolution for same identities	Exploits pose alignment connection and feature affinity connection
Person re-identification via ranking aggregation	Ranking aggregation algorithm	Considers both similarities and dissimilarities in person re identification	Not an effective method
Unified sparse subspace learning via self-contained regression	Joint sparse pixel weighted PCA	Enhances robustness	Not good for classification due to redundant features
Learning deep context- Aware features for person re-identification	Multi-Scale Context Aware Network (MSCAN)	Learns powerful features over full body and body parts	Cannot directly train this model on small dataset
Recurrent convolutional network for video based person re-identification	Recurrent neural network architecture	Makes use of color and optical flow information	Difficulty of creating an accurate appearance model
Person re-identification by descriptive and discriminative classification	Combination of both models	Higher accuracy	Limited amount of dataset
Person re-id by local maximal occurrence representation and metric learning	Retinex algorithm	Improves brightness, contrast, sharpness of an image through dynamic range compression	It either provide dynamic range compression or tonal rendition not both and also suffer color distortion
Fisher discrimination dictionary learning for sparse representation	Fisher discrimination based DL (FDDL)	Improves pattern classification performance	Time consuming
Person re-identification by video ranking	DVR framework and iLIDS-VID dataset	Capable of capturing more accurate space-time information	Difficulty to implement
Scalable person re-identification: a benchmark	Bag of words (BOW) descriptor	Bridge the gap between person re-identification and image search	Relatively low accuracy
RGB- infrared cross modality person re-identification	Multiple modality re-id dataset named SYSU-MM01	Only method for matching RGB and IR images	Challenging task due to large cross modality variations between RGB and IR images
MARS: A video benchmark for large scale person re-identification	Motion analysis and re-identification set	Largest video re-id dataset	Designed for the matching between true-color videos

MARS is the largest video re-id dataset. The pedestrians are captured by using six cameras. It consists of high quality images. For detecting the pedestrians deformable part model is used. GMMCP tracker is used to fill the missing result. The disadvantage is that it is designed for matching between true-color videos.

4. Conclusion

Person re-identification is used to identify and match the pedestrians captured between true-color and grayscale videos. The advantage of using grayscale videos is that it saves storage space and less information is needed to store a pixel. To study more on person re-identification, a literature survey was done. From the survey, it is clear that each method has limitations and still needs to work on person re-identification

References

[1] F. Ma, X.Y. Jing, X. Zhu, Z. Tang and Z. Peng, "True-Color and Grayscale Video Person Re-identification," IEEE 2019.
[2] Yang, J. Yan, S. Liao, D. Yi and S. Z. Li, "Salient Color Names for Person Re-identification," European Conference on Computer Vision (ECCV), Switzerland, pp. 536-551, 2014.
[3] Y. Wu, O. E. F. Bourahla, X. Li, F. Wu, and Q. Tian, "Adaptive Graph Representation Learning for Video Person Re-identification," IEEE 2018.
[4] M. Ye, C. Liang, Y. Yu, Z. Wang, Q. Leng, C. Xiao, J. Chen and R. Hu, "Person Re-identification via Ranking Aggregation of Similarity Pulling and Dissimilarity Pushing," IEEE 2016.
[5] S. Yi, Z.Y. He, Y.M. Cheung and W.S. Chen, "Unified Sparse Subspace Learning via Self-Contained Regression," IEEE 2018.
[6] D. Li, X. Chen, Z. Zhang and K. Huang, "Learning Deep Context Aware Features over Body and Latent Parts for Person Re-identification," IEEE 2017.
[7] N. McLaughlin, J.M. Rincon and P. C. Miller, "Recurrent Convolutional Network for Video-Based Person Re-identification," IEEE 2016.
[8] H. Martin, B. Csaba and Bischof, "Person re-identification by descriptive and discriminative classification," Scandinavian Conference on Image processing, 2011.
[9] S. Liao, Y. Hu, X. Zhu and S. Z. Li, "Person re-identification by Local Maximal Occurrence representation and metric learning," IEEE 2015.
[10] M. Yang, L. Zhang, X. Feng and D. Zhang, "Fisher Discrimination Dictionary Learning for sparse representation," IEEE 2011.
[11] T. Wang, S. Gong, X. Zhu and S. Wang, "Person Re-identification by Video Ranking," European Conference on Computer Vision (ECCV), Switzerland, 2014.
[12] L. Zheng, L. Shen, L. Tian, S.n Wang, J. Wang and Q. Tian, "Scalable Person Re-identification: A Benchmark," IEEE, US, 2015.
[13] A. Wu, W.-S. Zheng, H. Yu, S. Gong and J. Lai, "RGB-Infrared Cross-Modality Person Re-identification," IEEE 2017.
[14] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang and Q. Tian, "MARS: A Video Benchmark for Large-Scale Person Re-identification," European Conference on Computer Vision (ECCV), Switzerland, 2016.