Research Journal of Applied Sciences, Engineering and Technology 15(1): 7-15, 2018
DOI:10.19026/rjaset.15.5286
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: June 29, 2017 Accepted: October 10, 2017 Published: January 15, 2018

Research Article

Suitable Clustering for Multi-shot Person Re-identification

[1, 2]Yousra Hadj Hassen, [1, 2]Kais Loukil, [1, 2]Tarek Ouni and [1, 2]Mohamed Jallouli
[1]Computer and Embedded Systems Laboratory,
[2]National School of Engineers of Sfax, Street of Soukra km 4.5, Sfax 3038, Tunisia

Abstract: This study address the problem of selecting the most informative images for multi-shot person re-identification approaches. Actually, clustering algorithms have been one of the most proposed solutions. The objective of this study is to propose a suitable clustering method for most discriminative images selection in person re-identification. Clustering methods aim to divide huge data amount into groups depending on their characteristics, so that further data processing can be easier and more manageable. Actually, clustering is concerned in multiple fields such as document treatments, video summery and image processing. Person re-identification, a new area of research deeply investigated by the vision community, would be an interest application of the clustering algorithms. Person re-identification can be defined as the process of finding the identity of an unknown person who has already been observed in a camera view. Recently, re-identifying people over large public cameras networks has become a crucial task of great importance to ensure public security. Person Re-identification approaches can be either single shot, using one image to model a person's appearance, or multiple shot, using many images to identify a person. Actually, the real person re-identification framework is a multi-shot scenario. However, redundant images remain a challenging problem because execution time and memory consumption are significantly affected. In this study, an extensive comparison of clustering algorithms of state of art associated to a person re-identification framework is studied. Specifically, we evaluate the impact of clustering algorithms in re-identification rates. A standard re-identification framework is detailed and a synthesis of state of art methods comparison is conducted, using personal images and two standard datasets PRID_2011 and iLIDS-VID. Performing results are achieved in both person re-id rates (67.1% and 59.2%) and memory gain (92.7% and 97.8%) for Prid_2011 and iLIDS-VID datasets respectively.

Keywords: Camera network, clustering, identity selection, multi-shot, person re-identification, redundancy

INTRODUCTION

Most public places are monitored by large cameras networks with non-overlapping fields of views, especially in the last few years, to tackle terrorist acts. The ultimate goal of any surveillance system is not to track and reacquire targets anymore, but to understand the scene and to determine whether a given person of interest has already been observed over a network of cameras. As the person appearance varies greatly due to variations in lighting conditions, orientations and poses, there is a need of recognizing a person rather than tracing his trajectory. This is what we call "person re-identification". Recently, the computer vision has largely investigated in person re-id field (Karanam *et al*., 2016; Zheng *et al*., 2015). The overall scheme of a standard person re-identification (re-id) framework is detailed in Fig. 1.

The case of a small network formed of 2 cameras is given in Fig. 1. Depending on the camera view, the re-identification process can be divided into 2 principal steps; on one hand, in camera 1, a person is detected, modeled and tracked.

Of course, multiple detection algorithms have been proposed and great results have been achieved (Geronimo *et al*., 2010). The algorithm used in this study is the Histogram of Oriented Gradients (HOG) detector since it shows substantial gains over intensity based features detectors. Alternatively, the detected person is modeled using different features forming a discriminative descriptor. Person modeling is a dynamic field of research and plenty descriptors have been proposed (Heider *et al*., 2011). Actually, a descriptor is a set of spatial or textural features defined to identify visual person's appearance. In this study, the covariance descriptor is used (Hirzer *et al*., 2011). The detected person, defined by his model vector {f}, is tracked, over frames, in camera 1. On the other hand, the presence of the yet detected person in a different camera view (camera 2) needs to re-identify him; that is

Corresponding Author: Yousra Hadj Hassen, Computer and Embedded Systems Laboratory, National School of Engineers of Sfax, Street of Soukra km 4.5, Sfax 3038, Tunisia, Tel.: (+216) 24 327 904

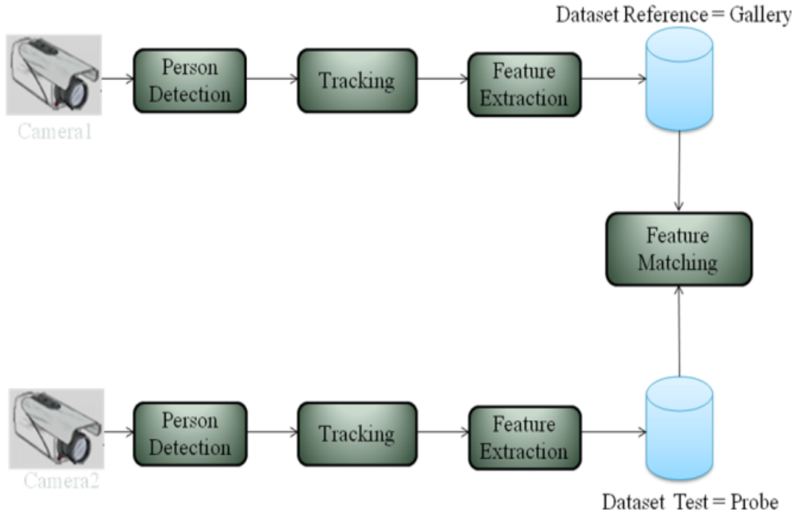

Fig. 1: Standard person re-identification framework

(a) (b) (c) (d) (e)

Fig. 2: Re-identification constraints; (a): miss-detection; (b): illumination variations; (c): viewpoint variations; (d): pose variations; and (e): occlusions

to match his identity across the two cameras views despite the changes that may occur to his appearance.

As mentioned, real re-id scenarios are multi-shot based process. So, multiple images of the same person are used to match his identity across different views. However, million images are captured for each target per day in uncontrolled conditions showing hard re-id constraints such as miss- detection, illumination variations, pose and viewpoints changes and occlusions presented in Fig. 2. Moreover, person's tracking sequences (tracks) used to be monotone and redundant. Therefore, the selection of the images that would be used to model a person's appearance is an inherently challenging task. That's why, clustering algorithms have been proposed in person re-id process.

The contribution of this clustering operation is two-fold:

- It captures only the relevant information
- It keeps low the computational cost of the matching process, where the clustering results are used.

In this study, a synthesis of clustering algorithms, frequently used in favor of re-identification, is detailed and deep evaluations are conducted to test the impact of the clustering algorithm in person re-id field.

LITERATURE REVIEW

Two different literatures are studied below. First, multi-shot person re-id related works are enumerated. Then, most proposed clustering algorithms for re-id goals are described.

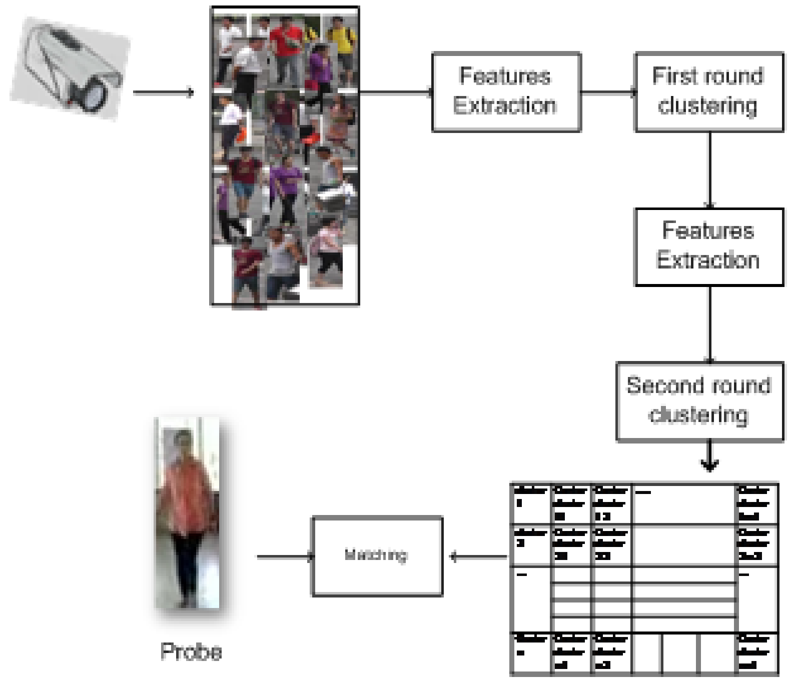

Fig. 3: Embedding clustering module in person re-id framework

Multi-shot person re-identification methods: Simply, person re-id is a similarity measure between different image descriptions. It is almost apparent that person re-id contributions rely on either image descriptions or distance measure.

Image descriptions based approaches differ by the visual features used to efficiently describe person's appearance by overcoming re-id constraints such as; view points and illumination variations, person appearance changes, occlusions and scale zooms yet shown in Fig. 3. Actually, the most commonly used features are color and texture. Gray and Tao (2008) use 8 color channels (RGB, HS and YC_bC_r) and 21 texture filters on the luminance channel and they succeed to overcome illumination variations problems. To overcome occlusions, person's appearance is modeled by a set of region covariance descriptors in Hirzer *et al*. (2011). To handle view variations, the symmetric property of the human body is exploited in Bazzani *et al*. (2013) and the SDALF (Symmetry Driven Accumulation of Local Features) is proposed. LAB color histogram and SIFT features are extracted by Li *et al*. (2013), Chen *et al*. (2016) and Zhao *et al*. (2013a, 2013b, 2014). Das *et al*. (2014) apply HSV histograms on the head, torso and legs from the silhouette. In Liu *et al*. (2014), the HSV histogram, gradient histogram and the LBP histogram are extracted for each local patch. Including the color and SILTP histograms in Liao *et al*. (2015), Liao *et al*. (2015) propose the Local Maximal Occurrence (LOMO) descriptor. A quad-tree feature is introduced in Ayedi *et al*. (2012) and the multi-scale covariance descriptor (MS-Cov) is proposed to tackle scale zooming and occlusions in person re-id.

Person re-id methods based on distance metrics learning have been largely studied (Yang and Jin, 2006). KISSME, one of the most popular metric learning, formulates the similarity measure as a likelihood ratio test and the difference space is assumed to be a Gaussian distribution. In addition to the Mahalanobis distance, Chen *et al*. (2015) add a bilinear similarity to model cross-path similarities. Yang *et al*. (2016) treat the similarities and the differences between image pairs in metric learning showing that the covariance matrices of different pairs can be distinguished from those of the similar pairs. Support vector machine or boosting are also used as learning tools for re-id such as the structural SVM proposed by Liu *et al*. (2015) based on different color descriptors and the Adaboost of Gray and Tao (2008) that defines a single similarity function combining different simple features.

For all the re-id methods mentioned above, the redundancy of images models is a common issue for both descriptor and metric learning based approaches. One of the proposed solutions is the use of PCA (Principal Component Analysis) in order to eliminate dimension correlations (Karanam *et al*., 2016; Zheng *et al*., 2015). Alternatively, clustering methods would be an efficient way to eliminate redundancy and to guaranty memory and execution time reduction for real person re-id systems.

Clustering algorithms: The word "Data clustering" refers to the process of partitioning a set of data into a set of meaningful sub classes called clusters (Aravind *et al*., 2010). Data clustering has immense number of applications in every field of life. Depending on the spread of the data, different clustering algorithms have been proposed. Namely, K-Means (Kanungo *et al*., 2002), K-Medoids (Park and Jun, 2009), the EM algorithm (Fraley and Raftery, 1998), different types of linkage methods, the mean-shift (Derpanis, 2005; Georgescu *et al*., 2003) algorithms that minimize some graph-cut criteria, Knn (Wong and Lane, 1981), fuzzy and grid based clustering (Lin *et al*., 2008; Chiang and Hao, 2003), etc. To the best of our knowledge, k-means, knn and mean shift algorithms are the most used clustering methods in person's tracking module (Fig. 1) embedded in the re-id framework (Yilmaz *et al*., 2006). Therefore, these algorithms are described below then, they are evaluated in terms of person re-id accuracy.

The k-means algorithm (Kanungo *et al*., 2002) is an expectation maximization technique iteratively alternating membership calculation and centroids adjustment. The heuristic initial step, for the proposed approach, allows the k-means algorithm to start from reasonable initial values, thus the local optima prove to be adequate solution in general. However, this algorithm remains notoriously sensitive to the initial choice of parameters as the number of clusters 'k' must be fixed in advance while in person re-identification neither the number of targets nor the number of person's appearance states cannot be estimated.

The k nearest neighbor (knn) clustering (Wong and Lane, 1981) is a non-parametric method. It consists in finding a cluster that a tested sample belongs to, among the k closest training samples in the feature space. A sample is assigned to the most common cluster among its k nearest neighbors cluster, by a majority vote of its neighbors. Knn is among the simplest clustering algorithm but it needs a tedious learning phase.

The mean shift algorithm (Derpanis, 2005), a simple but powerful tool of clustering literature, consists in shifting a point to the local center of mass around this point. It is an iterative gradient ascent method for finding local density maxima. It does not require prior knowledge of the number of clusters and does not constrain the shape of the clusters too. The algorithm begins by placing a window around each point in the feature space and each window moves in the direction of a specified mean shift vector iteratively until convergence.

Person re-id methods would give promising results associated to a clustering algorithm in order to enhance the matching results and eliminate redundant and monotonous data collected over tracking process. The proposed clustering algorithm embedded to re-id framework is described in the next section.

PROPOSED APPROACH

The study has been started since January 2014 in the Computer and Embedded Systems Laboratory, National School of Engineers of Sfax, Tunisia in the context of a project that aims to create a demonstrator for a camera IP for person re-identification.

The person re-id based clustering method is defined in Fig. 3. A two round clustering process is proposed. The mean shift clustering is performed on two stages. On one hand, to take into account inter-persons' variations in order to avoid erroneous re-id between similar people, the mean shift clustering is applied to the huge amount of images collected over cameras network. On the other hand, to consider the intra-person variations and to avoid failed re-id, the mean shift is implemented in a second stage for the yet clustered images. Therefore, the selected images would be discriminative in terms of variations between different persons and informative in terms of variable appearances for each clustered person's sequence. Consequently, an unknown person is re-identified by comparing the latter to most informative data collected over cameras network avoiding useless comparisons to noisy and redundant samples removed thanks to the clustering task.

The overall re-id approach presented by Fig. 3 is formed of three main steps; first, the appearance is modeled by extracting the visual features, then, the first round clustering module is developed to distinguish different persons and avoid inter persons re-id errors. After that, features are again extracted from different clusters and a second round clustering is introduced to consider intra person variations. Finally, the matching step finds the most similar identity to the unknown probe by comparing the latter to all selected discriminative frames presented by the table in Fig. 3. Actually, the two rounds clustering leads to n different clusters presenting n different persons. For each cluster, m_i clusters centers are selected. Consequently, both inter and intra persons variations are treated and re-id is granted in crowd places (big number of persons) and in real hard conditions (dynamic person's appearance variations due to illumination variations, pose variations…). The different modules are detailed below.

Feature extraction: The gallery set is modeled using the covariance descriptor (Hirzer *et al*., 2011). A (1x9) feature vector, defined in (1), is computed for each image and combined in a covariance matrix for each pixel localized at (x, y):

$$f = [\, x\ y\ Y\ C_b C_r\ mag\ grad\ I_x I_y \,] \quad (1)$$

where,

- Y = The luminance component
- C_r = The red chrominance component
- C_b = The blue chrominance component
- mag = The magnitude
- $grad$ = The gradient using the norm of the first order derivatives in x and in y respectively I_x and I_y The performance of the use of this combination of features is experimentally proved in Agarwal *et al*. (2004).

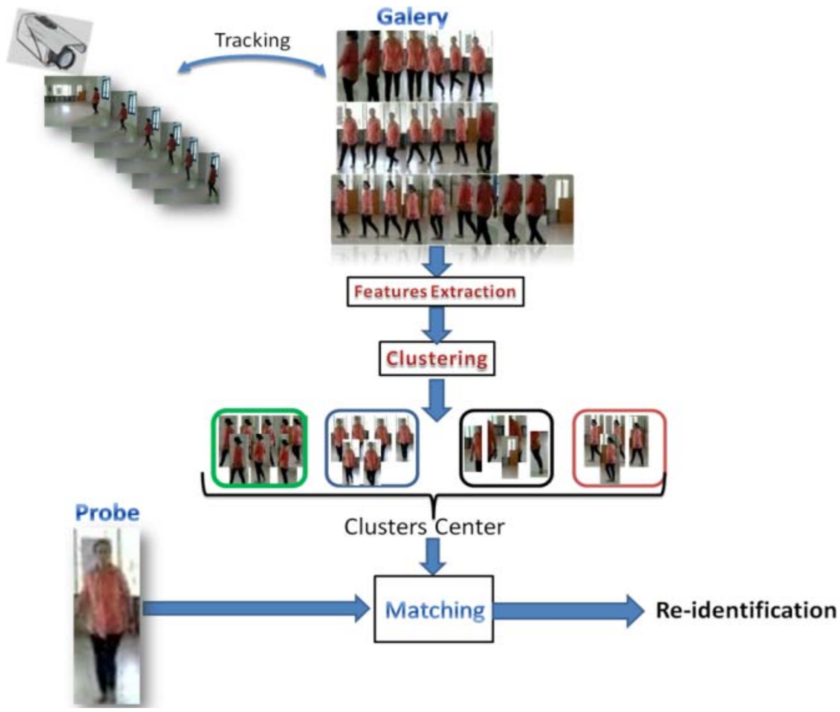

Fig. 4: Second round clustering for person re-identification

Clustering: For a gallery set formed of noisy, erroneous and redundant images that would be useless for re-id, to select and keep only informative frames seems to be a crucial task that must preprocess the matching phase in re-id framework. Different clustering algorithms are implemented and the impact on person re-id is highlighted and evaluated. The two levels clustering proposed for re-id is based on mean-shift algorithm.

The proposed mean shift clustering algorithm performs efficiently for person re-id because it does not require prior knowledge of the number of clusters and does not constrain the shape of the clusters too. The algorithm begins by placing a window around each point in the feature space (Chen *et al*., 2016). Each window moves iteratively in the direction of the mean shift vector, which is computed as follows:

$$Y_{t+1} = \frac{1}{|\theta_k|}\sum_{x\in\theta_\lambda}(x - y_t) \tag{2}$$

where, y_t is the window center at iteration t and θ_λ is the set of points in the hyper-sphere window of radius λ. It is also possible to use a kernel function to weight points according to how far they are from the window center. The windows eventually converge towards local density maxima yielding the cluster centroid. The points that converge to the same local maxima naturally fall into the same cluster. As such, the mean shift clustering algorithm avoids the issue of knowing the number of clusters at the price of introducing another bandwidth parameter λ. This parameter, however, is intuitive and easy to tune regarding all possible inputs (Chen *et al*., 2016). In this case, the radius of the hyper-sphere window λ is fixed as the average of experimental values tested for ten different scenarios for each image sequence. As just mentioned, the first round clustering aims to distinguish between different persons.

Figure 4 describes the second stage clustering. For the case of one target, the set of the huge amount of images captured over tracking are gathered in a 'gallery set'. The probe formed of an unknown person is matched to one of the element of the gallery. Thanks to the clustering, the gallery set of the treated target is refined. Consequently, comparison will be easier, faster and more efficient because a test person's image is matched to clusters the treated target is refined. Consequently, comparison will be easier, faster and more efficient because a test person's image is matched to clusters centers and not to the whole gallery images.

Covariance matrices belong to the group of symmetric positive definite (SPD) matrices, which can be interpreted as points on Riemannian manifolds. As such, the underlying distance and similarity functions might not be accurately defined in Euclidean spaces (Hassen *et al*., 2015). So, a logarithm definition of a covariance matrix c is used (3):

$$\log(c) = \sum_{i=1}^{\infty}\frac{(-1)^{i+1}}{i}(c - I_n) \tag{3}$$

where, I_n is an n*n identity matrix.

Table 1: Precision and recall for k-means, knn and mean-shift clustering for PRID_2011 and iLIDS-VID

	PRID_2011		iLIDS-VID	
Datasets	Precision	Recall	Precision	Recall
K-means	80%	23%	62%	30%
Knn	92.47%	100%	72%	25%
Mean shift	95%	63%	90%	55%

Table 2: Re-identification Rates for Prid_2011 and Ilids-VID datasets

Datasets	Prid_2011	iLIDS-VID
K-means	55.2%	74.5%
Knn	69.7%	82.3%
Mean shift	92.7%	86%

Table 3: Re-identification Rates and memory gain with and without Mean-shift clustering

	Prid_2011		iLIDS-VID	
Datasets	With Mean-sfift	Without Mean-sfift	With Mean-sfift	Without Mean-sfift
Re-id rates	67.1%	84.3%	59.2%	76.9%
Memory gain	92.7%	0%	97.8%	0%

Therefore, frames are decomposed on patches or search window which the size depends on considering multi-scales or not. Covariance matrices of each patch are computed. The best matched patch of the target considered is picked by its minimum logarithm Euclidian distance compared to the target covariance matrix. This distance is computed by (4) between two covariance matrices A and B of two different frames respectively:

$$\|\log A - \log B\| \quad (4)$$

The major contribution of this study is to cluster twice the gallery set elements in order to summarize person's appearance variations through concise and few numbers of frames. The efficient clustering gives a set of representative images for each person eliminating the huge amount of frames resulting from tracking algorithm while keeping a unique and discriminative identity for each person. Only the centers of the clusters are kept in the gallery set. Hence, the memory consumption would be obviously reduced. To that end, a comparative study is conducted to highlight the performing impact of clustering algorithms on multi-shot person re-id.

RESULTS AND DISCUSSION

Datasets: For experiments, we use multi-shot standard datasets PRID_2011 (Bak *et al*., 2010) and iLIDS-VID (Wang *et al*., 2014). These two datasets are very challenging due to clothing similarities among people, lighting and view point variations across camera views, cluttered background and occlusions.

PRID2011 dataset: PRID2011 dataset is formed of images of 200 and 749 people captured by two cameras A and B respectively. Each person has 5 to 675 images available. This dataset is hard because it presents real images with noisy background and illumination variations.

iLIDS-VID dataset: It presents different frames of 300 people captured by two non-overlapped cameras in an airport arrival hall. It is a challenging dataset due to the huge amount of images per person with clothing similarities and both partial and total occlusions.

Test design: The results of re-id rates and memory consumption are computed in Table 1 to 3 for respectively Prid_2011 and iLIDS-VID datasets for the three basic clustering algorithms k-means (Kanungo *et al*., 2002), knn (Wong and Lane, 1981) and mean-shift (Derpanis, 2005). Of course, the person re-identification rates are not outperforming effective re-id literature results even with the embedded clustering but they remain close to the competing rates given by the most robust multi-shot person re-id methods. Nevertheless, the memory gain is significant thanks to the notable reduction of the data for matching. Thus, real world re-id scenarios could be efficiently treated.
As argued in Ayedi *et al*. (2012), the suitable measure for evaluating clustering accuracy on a given test set is a recall-precision metric. Recall and precision are defined as follows:

$$Recall\ (\tau) = \frac{TP(\tau)}{TP(\tau)+FN(\tau)} \quad (5)$$

$$Precision(\tau) = \frac{TP(\tau)}{TP(\tau)+FP(\tau)} \quad (6)$$

TP, FP and FN denote the sets of true positives, false positives and false negatives, respectively. There is a trade-off between precision and recall. Greater precision decreases recall and greater recall leads to decreased precision.

Table 2 presents the re-identification rates achieved by the proposed re-id framework for the three basic

Fig. 5: Examples of clusters center used for re-identification of 100 personal images

clustering algorithm cases using both challenging datasets PRID_2011 and iLIDS-VID that are divided into random sets of probe and gallery to evaluate different re-id scenarios. The different parameters for the tested clustering algorithms (k, λ, θ_{λ}…) are fixed through empirical tests.

Some samples of personal tracking sequences used to evaluate discriminative identities provided by mean-shift clustering for three tracked targets are presented in Fig. 5.

The memory gain in terms of the percentage of the used images in the galley independently of the computer memory and the re-id rates have been computed and presented in Table 3 in both cases of presence and absence of the mean-shift clustering.

The above clustering evaluation proves the outperformance for mean-shift clustering compared to k-means and knn. The mean-shift clustering does not require a previous knowledge of the number of clusters which allows a best decomposition of appearances based on similarities and dissimilarities. Actually, to efficiently re-identify a person we should have a concise and brief description of his appearance. Consequently, to cluster a person's images without a constraining number of clusters leads to keep discriminative images to model a person for re-id.

First, the proposed clustering algorithm is tested in terms of its ability to cluster images. Table 1 shows that mean-shift achieves performing results of recall and precision in both multi-shot datasets. This can be explained by the unfixed number of clusters of persons unlike K-means. While knn is steel providing competing recall and precision results because of its non-parametric aspect but for person re-id this clustering algorithm will be avoided due to some reasons that will be explained below.

Then, Table 2 summarizes the re-id rates of different randomly selected sequences of both datasets for the three tested clustering methods. The obtained results prove significant outperformance of mean-shift compared to k-means and knn. These results confirm the findings presented in Table 1 and the avoidance of the knn algorithm. Figure 5 shows the clusters centers of an example of 100 personal tested images forming the discriminative identity used in re-id.

Finally, the major performance of the double round clustering proposed in the favor of person re-id is presented through Table 3. Although, the re-id rates without using clustering are steel the most performing, the proposed method's results remain close and promising. While the consumed memory in the gallery is widely performing for the case of the clustering based re-id proposed in this study. In fact, without embedding clustering, a large memory is needed to save the huge amount of images collected over large cameras networks and their dense computed features. That is a 0% memory gain in terms of used dataset because all the available data must be saved and trained i.e., 100% of the dataset is used. However, with embedding the mean-shift clustering task, the memory gain achieved is 92.7% and 97.8% for prid-2011 and iLIDS-VID datasets respectively. This effective gain presents a major contribution for the multi-shot person re-id procedure. These results enhance temporal re-id and real time re-id may be achieved thanks to the clustering based re-id with some improvements in descriptor and matching modules of the multi-shot re-id.

The obtained results are compared to some previous multi-shot person re-id works. Table 4 presents person recognition rates at rank 1 for both prid-2011 and iLIDS-VID datasets.

Table 4: Comparison of the proposed method with state-of-the-art using recognition rate (%) at rank 1

Method	Prid-2011	iLIDS-VID
SDALF (Bazzani *et al*., 2013)	5.2%	5.1%
DVR (Wang *et al*., 2014)	28.9%	23.3%
DSVR (Wang *et al*., 2016)	40.0%	39.5%
Ours	60.9%	46.0%

For fair comparison with state of the art methods, using experimental set up as in Wang *et al*. (2014), we use only partial dataset in evaluation. We compare performance of our approach with SDALF (Bazzani *et al*., 2013), supervised model learning approaches as Discriminative Video Ranking (Wang *et al*., 2014) (DVR) and Discriminative Selection in Video Ranking (Wang *et al*., 2016) (DSVR). The top recognition rates ranks of these approaches and our method are given in Table 4 for both Prid2011 and iLIDS-VID datasets. As shown in Table 1, the proposed method outperforms significantly the recent proposed state of the art methods. The person re-id based clustering proposed method is outperforming previous re-id works for both tested datasets. These results can be explained by the discriminative samples used for person identification in the gallery set thanks to the clustering step embedded in the re-id framework.

CONCLUSION

Person re-identification, has become an inherently task for extensive interest in the modern scientific community. In this study, an evaluation of different literature clustering is monitored to select the most discriminative images avoiding noisy, redundant and monotonous images in order to achieve both memory and execution time gain. The re-identification rates are computed using frequent clustering algorithms and the impact of these algorithms is highlighted. The paper highlights the automatic treatment of the huge data amount collected over large cameras networks while guarantying performing re-identification accuracy. Even though, the embedding of clustering leads to a successful multi-shot person re-identification system. The proposed approach will be deeply studied to reach more performing real-world results and further evaluations will be delivered in future works.

REFERENCES

Agarwal, S., A. Awan and D. Roth, 2004. Learning to detect objects in images via a sparse, part-based representation. IEEE T. Pattern Anal., 26(11): 1475-1490.

Aravind, H., C., Rajgopal and K.P. Soman, 2010. A simple approach to clustering in excel. Int. J. Comput. Appl., 11(7): 19-25.

Ayedi, W., H. Snoussi and M. Abid, 2012. A fast multi-scale covariance descriptor for object re-identification. Pattern Recogn. Lett., 33(14): 1902-1907.

Bak, S., E. Corvee, F. Bremond and M. Thonnat, 2010. Person re-identification using spatial covariance regions of human body parts. Proceeding of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. IEEE Press, Boston, pp: 435-440.

Bazzani, L., M. Cristani and V. Murino, 2013. Symmetry-driven accumulation of local features for human characterization and re-identification. Comput. Vis. Image Und., 117(2): 130-144.

Chen, D., Z. Yuan, G. Hua, N. Zheng and J. Wang, 2015. Similarity learning on an explicit polynomial kernel feature map for person re-identification. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. Boston, MA, USA, pp: 1565-1573.

Chen, D., Z. Yuan, B. Chen and N. Zheng, 2016. Similarity learning with spatial constraints for person re-identification. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, Las Vegas, pp: 1268-1277.

Chiang, J.H. and P.Y. Hao, 2003. A new kernel-based fuzzy clustering approach: Support vector clustering with cell growing. IEEE T. Fuzzy Syst., 11(4): 518-527.

Das, A., A. Chakraborty and A.K. Roy-Chowdhury, 2014. Consistent re-identification in a camera network. In: Fleet, D., T. Pajdla, B. Schiele and T. Tuytelaars (Eds.), Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, Springer, Cham, 8690: 330-345.

Derpanis, K.G., 2005. Mean Shift Clustering. Lecture Notes.

Fraley, C. and A.E. Raftery, 1998. How many clusters? Which clustering method? Answers via model-based cluster analysis. Comput. J., 41(8): 578-588.

Georgescu, B., I. Shimshoni and P. Meer, 2003. Mean shift based clustering in high dimensions: A texture classification example. Proceeding of the 9th IEEE International Conference on Computer Vision (ICCV). Nice, France, 3: 456.

Geronimo, D., A.M. Lopez, A.D. Sappa and T. Graf, 2010. Survey of pedestrian detection for advanced driver assistance systems. IEEE T. Pattern Anal., 32(7): 1239-1258.

Gray, D. and H. Tao, 2008. Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: Forsyth, D., P. Torr and A. Zisserman (Eds.), Computer Vision – ECCV 2008. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 5302: 262-275.

Hassen, Y.H., T. Ouni, W. Ayedi and M. Jallouli, 2015. Mono-camera person tracking based on template matching and covariance descriptor. Proceeding of the IEEE International Conference on Computer Vision and Image Analysis Applications (ICCVIA). Sousse, Tunisia, pp: 1-4.

Heider, P., A. Pierre-Pierre, R. Li and C. Grimm, 2011. Local shape descriptors, A survey and evaluation. In: Laga, H., T. Schreck, A. Ferreira, A. Godil, I. Pratikakis and R.C. Veltkamp (Eds.), Proceeding of the 4th Eurographics Workshop on 3D Object Retrieval (3DOR). Eurographics Association, Llandudno, UK, April 10, 2011, pp: 49-56.

Hirzer, M., C. Beleznai, P.M. Roth and H. Bischof, 2011. Person Re-identification by Descriptive and Discriminative Classification. In: Heyden, A. and F. Kahl (Eds.), Image Analysis. SCIA 2011. Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, 6688: 91-102.

Kanungo, T., D.M. Mount, N.S. Netanyahu, C.D. Piatko, R. Silverman and A.Y. Wu, 2002. An efficient k-means clustering algorithm: Analysis and implementation. IEEE T. Pattern Anal., 24(7): 881-892.

Karanam, S., M. Gou, Z. Wu, A. Rates-Borras, O.I. Camps and R.J. Radke, 2016. A comprehensive evaluation and benchmark for person re-identification: Features, metrics, and datasets. Comput. Vis. Pattern Recogn., pp: 1-19.

Li, Z., S. Chang, F. Liang, T.S. Huang, L. Cao and J.R. Smith, 2013. Learning locally-adaptive decision functions for person verification. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, Portland, pp: 3610-3617.

Liao, S., Y. Hu, X. Zhu and S.Z. Li, 2015. Person re-identification by local maximal occurrence representation and metric learning. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, Boston, pp: 2197-2206.

Lin, N.P., C.I. Chang, H.E. Chueh, H.J. Chen and W.H. Hao, 2008. A deflected grid-based algorithm for clustering analysis. WSEAS T. Comput., 7(4): 125-132.

Liu, X., M. Song, D. Tao, X. Zhou, C. Chen and J. Bu, 2014. Semi-supervised coupled dictionary learning for person re-identification. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition. IEEE Press, Columbus, pp: 3550-3557.

Liu, X., H. Wang, Y. Wu, J. Yang and M.H. Yang, 2015. An ensemble color model for human re-identification. Proceeding of the IEEE Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA, pp: 868-875.

Park, H.S. and C.H. Jun, 2009. A simple and fast algorithm for K-medoids clustering. Expert Syst. Appl., 36(2/2): 3336-3341.

Wang, T., S. Gong, X. Zhu and S. Wang, 2014. Person re-identification by video ranking. Proceeding of the European Conference on Computer Vision. Zurich, Switzerland.

Wang, T., S. Gong, X. Zhu and S. Wang, 2016. Person re-identification by discriminative selection in video ranking. IEEE T. Pattern Anal., 38(12): 2501-2514.

Wong, M.A. and T. Lane, 1981. A kth nearest neighbour clustering procedure. In: Eddy, W.F. (Ed.), Computer Science and Statistics: Proceeding of the 13th Symposium on the Interface. Springer, NY, US, pp: 308-311.

Yang, L. and R. Jin, 2006. Distance metric learning: A comprehensive survey. Michigan State Univ., 2: 51.

Yang, Y., S. Liao, Z. Lei and S.Z. Li, 2016. Large scale similarity learning using similar pairs for person verification. Proceeding of the 13th AAAI Conference on Artificial Intelligence.

Yilmaz, A., O. Javed and M. Shah, 2006. Object tracking: A survey. ACM Comput. Surv., 38(4).

Zhao, R., W. Ouyang and X. Wang, 2013a. Unsupervised salience learning for person re-identification. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Press, Portland, USA, pp: 3586-3593.

Zhao, R., W. Ouyang and X. Wang, 2013b. Person re-identification by salience matching. Proceeding of the IEEE International Conference on Computer Vision (ICCV). IEEE Press, Sydney, pp: 2528-2535.

Zhao, R., W. Ouyang and X. Wang, 2014. Learning mid-level filters for person re-identification. Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Press, Columbus, USA, pp: 144-151.

Zheng, L., Y. Yang and A.G. Hauptmann, 2015. Person re-identification: Past, present and future. J. Latex Class Files, 14(8): 1-20.