International Journal of Engineering & Technology, 7 (4.11) (2018) 174-178

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper



Pose and Illumination Invariance of Attribute Detectors in Person Re-identification

Mohammadali Saghafi[1], Aini Hussain[1]*, Mohamad Hanif Md. Saad[1], Mohd Asyraf Zulkifley[1], Nooritawati Md Tahir[2], Mohd Faisal Ibrahim[1]

[1]Center for Integrated Systems Engineering and Advanced Technologies (INTEGRA), Faculty of Engineering and Built Environment, Universiti Kebangsaan Malaysia, 43600 Bangi, Selangor, Malaysia
[2]Advanced Computing and Communication, Faculty of Electrical Engineering, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
**Corresponding author E-mail: draini@ukm.edu.my*

Abstract

The use of attributes in person re-identification and video surveillance applications has grabbed attentions of many researchers in recent times. Attributes are suitable tools for mid-level representation of a part or a region in an image as it is more similar to human perception as compared to the quantitative nature of the normal visual features description of those parts. Hence, in this paper, the preliminary experimental results to evaluate the robustness of attribute detectors against pose and light variations in contrast to the use of local appearance features is discussed. Results attained proven that the attribute-based detectors are capable to overcome the negative impact of pose and light variation towards person re-identification activities. In addition, the degree of importance of different attributes in re-identification is evaluated and compared with other previous works in this field.

Keywords: *person re-identification; Attribute; metric learning.*

1. Introduction

In general, several challenges related to person re-identification (PRI) includes lighting and view point variations that could cause severe effects on the re-identification output quality. In appearance-based re-identification, the main purpose is representing the subject's image based on the visual features which are robust enough against these variations. Thus, these features must be discriminative enough to handle the close visual differences of the body silhouettes that belong to different people [1]. Previous researches have reported their findings using numerous methods and approaches in evaluating various types of low level features for re-identification [2-5]. For instance, in [2] utilised color encoding technique, whilst in [3] used texture and in [4] employed shape of their database images to represent the visual descriptors based on each stated feature [6]. Although low level features performance is acceptable on images with some variety of illumination and view point, but in the severe cases, low level features are unable to perform well. In such cases, segmented grids are more preferable since this approach could contribute too much better results [4] along with local parts of the silhouettes [2] rather than on individual pixels.

Conversely, in ensuring perfect re-identification rate, the descriptors need to fulfill both descriptive and discriminative features which is possible but challenging if not impossible in some cases. One way to resolve this problem is to perform re-identification based on mid-level representation of the scene rather than low level demonstration. Attributes are the keys that bridges low level features to mid-level understanding from the scenes. These attributes are semantic descriptions of the scenes, which are widely used in object recognition tasks in recent years [7]. The idea of using semantics in re-identification is relatively new and few researches have done that [8, 9] and the whole re-identification process is not being done purely based on attributes in existing approaches. Instead, the attributes-based approach is used to supplement the existing approach [9]. The use of mid-level representation of the scenes brings us one step closer to the way that human really understand what is happening around. It also eliminates the difficulties caused by illumination and pose variations. The latter is the reason that encourages researchers to go beyond the low-level descriptors.

One of the main challenges in implementing this approach is to have attribute detectors with high detection accuracy. The attributes are defined differently according to different applications. In robotic applications, color of a specific object can be the desired attribute while in a face verification application wearing sunglasses is defined as a desired attribute [10]. In PRI, based on the dataset for re-identification purpose, the attributes can be defined vastly from the color attribute for instance clothing or texture like shirts patterns or even carrying objects like backpacks and satchels. In other words, every option that contributed to discrimination can be defined as an attribute [11]. Based on these different options to define attributes for re-identification, it would be of high interest to do PRI purely based on attributes.

2. Attribute Detection

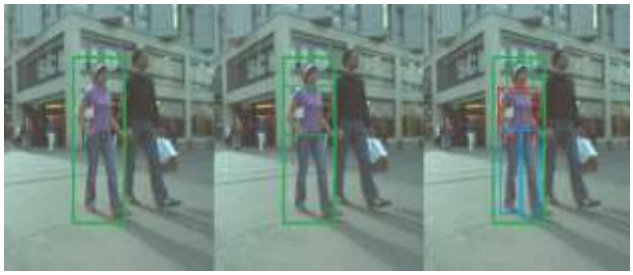
As mentioned before, different attributes can be defined for retrieval, surveillance and re-identification applications. For instance, in some cases, the attributes of the face and head (i.e. sunglasses, bald, hat) are considered due to high quality of the images [10]. While, in some others because of the lack of good quality, it is

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

better to ignore such attributes and focus on other attributes instead. Figure 1 shows some attributes which are suitable to be considered as chosen from the VIPeR dataset [12]. The low-level features that are used to train an attribute detector have significant effect on its accuracy and mostly color, texture, edge and shape are features extracted from the dataset images for training the detectors [7, 9, 13, 14]. The way that the features are extracted is also important as they can be extracted either globally from the whole silhouette or locally from the segmented silhouette [3, 9]. Figure 2 shows different sub-partitions of a silhouette from ETHZ public dataset [15] in which the features can be extracted from.



Fig. 1: The attributes from left to right are: red shirt, backpack, patterned shirt, shorts, stripes, skirt (Images grabbed from VIPeR dataset [12]



A **B** **C**

Fig. 2: Example of different segmentation ways in which features are extracted for attribute detection. A: global silhouette B. head, torso and leg segmentation C. body part-based segmentation (Images grabbed from ETHZ dataset [15]

Upon completion of feature extraction process, the next stage is classification. In [7] used SVM and logistic regression as attributes detectors. Other researchers have investigated the application of other classifier as attribute detectors. The classifiers are trained by training data and the parameters are cross validated to find the best features. The method that is applied to train the classifiers is also crucial to avoid mistakes in conveying the semantic of attribute to the attribute detector. For instance, if a classifier is trained for wheel detector and the training data is only from the wheels of cars and buses. This may cause the detector to detect the wheels in combination of the metals around them and unable to detect the wheel of a wooden carriage during test time. In such cases, feature selection is crucial to ensure matching of the attribute detector and its real concept [7, 14]. Also, the training set must represent different type of sources for a particular attribute. The main problem of pose and illumination differences can be solved in attribute-based methods by using the samples in training set with different illumination and pose conditions, which can be more robust than the descriptors with low level features.

3. Re-Identification by Attributes

Attributes have grabbed attentions in recent years for various applications in computer vision and can be considered new in re-identification application [11, 16]. Hence, to use them in re-identification, the following issues must be addressed.

3.1. Data Insufficiency and Unbalanced Data

Most available datasets for PRI have less than 1000 images and sufficient data are required to train the attribute detectors accurately. In addition, proper feature selection methods [7, 16] can help alleviate this namely by training attribute detectors on a dataset with sufficient samples of that particular attribute and test it on re-identification dataset. Next is to ensure that the number of positive and negative samples must be balanced to avoid over-fitting.

3.2. Classification Metric

The process of attribute detection produces a set of pre-trained attribute detectors. Currently, there is no re-identification method which purely uses attribute for re-identification. In [9] defined attribute distance between probe image and gallery images and used this method in combination with the low-level descriptor distance as shown in (1).

$$d_{W^L,W^A}(I_p, I_g) = \sum_{l \in LL} W_l^L d_l^L(L_l(I_p), L_l(I_g)) + d_{W^A}^A(A(I_p), A(I_g)) \quad (1)$$

The first term of this equation is the distance between the low-level descriptors of probe image (I_P) and gallery image (I_g) and the second term is devoted to the distance of the attributes of probe image and gallery image. As illustrated in this formula, the attribute distance ($d_{W^A}^A$) only plays as a complementary role here.

In this work, one important reason that the attributes were not used purely for re-identification was the accuracy of the detectors. Therefore, if one can train attribute detectors with a very high degree of accuracy, then the re-identification task can be done solely based on attributes. The way the extracted attributes from probe and gallery images are being compared are also important in re-identification process performance. While one can only use a rigid metric in (1) for attribute comparison, the use of learning metrics can make significant improvement to the re-identification process as reported in [17-18].

Suppose, a set of extracted attributes from the probe image is A. Next, to determine a set of the best gallery images which highly matches with the probe attribute set, a prediction function $f_w: A \rightarrow y$ must be defined to find a set of images y^* which maximizes the score over weight vector, w:

$$y^* = \underset{y \in Y}{\max} w^T \psi(A, y) \quad (2)$$

The score function ψ is the output scores of pre-trained attribute detectors that are independently trained for each attribute in the set of probe attributes (A). The goal is to train a model, w, which is able to correctly predict the closest gallery images that have similar attributes A as the probe image which can then be solved as a max-margin problem as in [1, 10, 19].

4. Attribute-based Re-Identification Privileges

As discussed earlier, using attributes in re-identification could resolve the issue of light variations. This is done by giving many different samples with various illumination conditions to train a specific attribute that makes the detector to be robust against lighting variations in comparison with the low-level descriptors. The situation is similar for varying poses of people. Suppose, a person with a backpack must be re-identified with only the front and side or back views as shown in Figure 3. It is obvious that the ordinary low-level descriptors cannot simply work on these two views as the color appearance is different. While low level descriptor may fail to re-identify, in such case, an attribute detector trained to detect backpack will be able to perform re-identification. When training a backpack detector, both images of the subjects carrying backpacks from the front view, back or side view can be used as the front view of the subjects with backpacks can be recognized from the two parallel stripes of the backpacks and therefore these

features can be used to train the detector to detect a backpack in images.



Fig. 3: Persons' front and back (side) view with a backpack. A backpack detector can more easily detect it in two available shots than a low-level color and texture descriptor (Images grabbed from VIPeR dataset [12])

The other advantage of using attributes, in addition to having a semantic representation of the probe and gallery images, is that the dimension has been reduced greatly and this can speed up the procedure and decrease the complexity of the model specifically when using learning metrics as mentioned in the previous section.

5. Results and Discussion

In this section, the performance re-identification based on attributes in comparison to appearance features is investigated. The results attained are compared with a baseline method.

5.1. Preliminary Test on the Effect of Attributes on Re-Identification

It must be noted that the aim is not to reach the re-identification rates which is made by strong appearance-based descriptors but to only evaluate the robustness of the attributes against light and pose variations as compared to simple appearance features. For this reason, a pre-gathered set of 26 video frames from 13 different people crossing a corridor and a room by two cameras (13 videos per camera) is used. The videos are recorded on different times of the day, with different illumination conditions and comprised of different poses of individuals. As depicted in Figure 4, ViBe [20] is used for background subtraction from the video frames followed by segmentation of the foreground silhouette to three parts specifically head, torso and legs as in [3].



Fig. 4: Background subtraction using ViBe [20]

Two simple features of HSV histogram and HOG are used as baseline features to extract the color and texture information from the torso. Further, these extracted features are used to train the attribute detectors. One detector is trained to detect the clothing pattern of the upper body and another four detectors are as color detectors (red, yellow, green and dark) of the torso. The reason that these colors are chosen is due to their high occurrence in the database. Next, linear SVM is used to train the detectors using cross validation method. To train the classifiers, different frames with different lighting conditions are used. This is done by synthesizing the database images with vast illumination changes contrast and intensity as shown in Figure 5.



Fig. 5: The silhouettes' intensities are changed manually for training the robust attribute detectors (here in this work we only used the torsos)

The attribute detectors accuracies are as shown in Table 1. As tabulated in Table 1, patterned torsos contributed to highest recognition rate followed by yellow colour. In the case of the red color, the reason of lower recognition rate is the inefficiency of HSV color space in detecting this color and in the case of dark torsos; it is because of the scarcity of the samples. The low performance of these detectors affected the re-identification rate.

Table 1: Attribute detectors performances

Attribute	Accuracy (%)
Red Torso	59
Yellow	80
Green	76
Dark torso	66
Patterned torso	84

Table 2: Re-identification rates using baseline features versus attribute detectors

Method	Re-Identification Rate		
	Rank1	Rank5	Rank10
Fusion of HSV and HOG	10	14.2	23.1
Attribute detectors	11.5	16.4	26.9

It is worth mentioning that the goal of re-identification based on fusion of HSV and HOG is to evaluate the robustness of an attribute detector against lighting or illumination and pose changes in PRI. Table 2 tabulated the re-identification rate for the first 10 ranks. The performance of pre-trained detectors is higher for all three ranks as compared to raw features due to robustness to varying illuminations and poses.

5.2. Boosting Robustness through Pose and Illumination Using SVM

To evaluate the effect of attribute detectors on light and pose differences, standard public dataset VIPeR is used. The VIPeR consisted of 632 pairs of images that were taken using two cameras with 128*48 resolutions. Having images with drastically different illumination and pedestrian angles made this dataset suitable for evaluating the effect of attribute detectors on improving re-identification rate in the existence of these two phenomena. The attributes that are considered here are those frequently appeared. Note that there are several other attributes that can be considered and trained on VIPeR but for simplicity, five attributes are chosen to evaluate the effectiveness of the developed detectors in this study. The detectors are the same as the one used for our local dataset. To train the attribute detectors, the ensemble of localized features (ELF) proposed by [12] is used and this set of features includes 2784 dimensional vectors of color and texture features. Table 3 enlists these attributes and tabulated the accuracy rate using the proposed detectors in this study. It was found that dark shirt contributed to highest accuracy rate of 88.1% as compared to worst performance with blonde hair as the attribute

Table 3: Performance of selected attribute from VIPeR dataset using the proposed detectors

Attribute	Accuracy (%)
Red shirt	84.0
Blue shirt	71.0
Skirt	75.6
Dark shirt	88.1
Blond hair	68.5

Next, the effect of the attribute detectors on improving re-identification performance is discussed. Out of 316 images from the VIPeR database, 245 pairs with drastically bad conditions of light and pose variations are selected.

Typically, in PRI does not involve a pure binary (0-1) classification problem and since the aim is to find the most relevant test set with query one, applying SVM at the classification stage is proven to be a suitable tool in comparison to other classifiers. Furthermore, the ability of ranking based on the relevancy of attributes between a query image and test images is the significant factor in our approach. As such, to assess the capability of our pose-illumination invariant approach, the ranking SVM approach by [17] for PRI has been utilised. In their method, the aim was to learn a ranking score in which the relevant images have higher score than irrelevant ones. As shown in (3), δ represents the ranking score and is defined as follow:

$$\delta(X_i - X_{i,j}) = W^T|X_i - X_{i,j}| \quad (3)$$

where, X_i is a multi-dimensional feature vector representing the query image and $X_{i,j}$ represents the other images in the dataset. Having a learning problem as in (3), we wish to train W such that:

$$W^T(|X_i - X_{i,j}^+| - |X_i - X_{i,j}^-|) > 0 \quad (4)$$

where, $X_{i,j}^+$ shows the feature vector of relevant images to the query image while $X_{i,j}^-$ shows the irrelevant images. As can be seen in 4, the vectors that represent the images contain low-level features. In our experiment, low-level features are used and the effect of attributes as mid-level features is examined as well. Therefore, a part devoted to attributes and the formula is changed in (3) to (5):

$$\delta(X_i - X_{i,j}) = W^T(|X_i - X_{i,j}| + |A_i - A_{i,j}|) \quad (5)$$

In 5, $A_i$and $A_{i,j}$ are the 5-dimensional attribute vectors that specify the attributes related to the query image and attributes related to the other images, accordingly. By adding this part to equation, the existence of mid-level features in enhancing robustness through pose and light variations is validated if these features contribute to better accuracy rate. First, the detectors are trained using the whole VIPer dataset. Next, 245 pairs of images with bad conditions of light and pose are used as testing images. This procedure is repeated once without applying attribute-related part in 5 (as done by [17]). Results attained showed higher accuracy of attribute-contributed PRI as depicted in Figure 6. As can be seen, the attributes positively have good effect on the result and thus, enhanced the robustness.

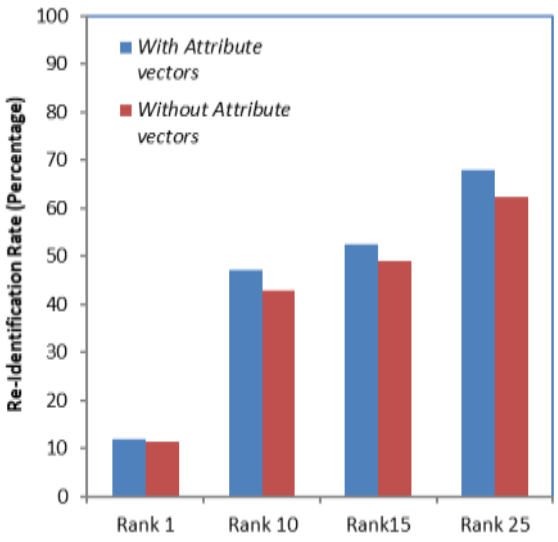


Fig. 6: The effect of attribute vectors on re-identification rates (testing on 245 images of VIPeR)

6. Conclusion

In conclusion, it is proven that attributes with semantic representation of the subjects' images is capable to enhance re-identification rate by making the approach more robust against pose and illumination. However, in order to achieve an optimal correct re-identification rate, some issues still remain to be solved. The issues include having adequate number of training samples and determining significant attribute detectors itself. Future work includes using other standard public datasets and rigorous comparison of the results against state-of-the-art approaches.

Acknowledgement

This research is funded by Ministry of Science, Technology and Innovation (MOSTI) Malaysia (via grant 01-01-02-SF1386) and Universiti Kebangsaan Malaysia (UKM) under the DIP-2015-012 grant.

References

[1] Wang T, Gong S, Zhu X and Wang S, "Person re-identification by discriminative selection in video ranking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38 (2016), 2501-2514.

[2] Satta R, Fumera G and Roli F, "Exploiting dissimilarity representations for re-identification," *Proceedings of the International Workshop on Similarity-Based Pattern Recognition*, (2011), pp. 275-289.

[3] Farenzena M, Bazzani L, Perina A, Murino V and Cristani M, "Person re-identification by symmetry-driven accumulation of local features," *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, (2010), pp. 2360-2367.

[4] Zheng W, Gong Sh and Xiang T, "Associating groups of people," *Proceedings of the British Machine Vision Conference*, (2009), pp. 1.

[5] Poongothai E, Suruliandi, A., "Survey on color, texture and shape features for person re-identification," *Indian Journal of Science and Technology*, 9 (2016), pp. 1-7.

[6] Poongothai, E, Suruliandi A, "color, texture and shape feature analysis for person re-identification technique," *Advances in Vision Computing: An International Journal*, 3 (2016), 17-26

[7] Farhadi A, Endres I, Hoiem D and Forsyth D, "Describing objects by their attributes," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2009), pp. 1778-1785.

[8] Satta R, Pala F, Fumera G and Roli F, "People search with textual queries about clothing appearance attributes," in S. Gong, M. Cristani, S. Yan, & C. Loy (Eds.), *Person Re-Identification. London: Springer*, (2014), pp. 371-389.

[9] Layne R, Hospedales TM and Gong S, "Attributes-based re-identification," in S. Gong, M. Cristani, S. Yan, & C. Loy (Eds.), *Person Re-Identification. London: Springer*, (2014), pp. 93-117.
[10] Siddiquie B, Feris RS and Davis LS, "Image ranking and retrieval based on multi-attribute queries," *Proceedings of the Computer Vision and Pattern Recognition*, (2011), pp. 801-808.
[11] Li A, Liu L, Wang K, Liu S and Yan S, "Clothing attributes assisted person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, 25 (2015), 869-878.
[12] Gray D, Brennan S, Tao H, "Evaluating appearance models for recognition, reacquisition and tracking," *Proceedings of the 10th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, (2007), pp. 1-7.
[13] Jungseock J, Wang S and Zhu SC. "Human attribute recognition by rich appearance dictionary," *Proceedings of the IEEE International Conference on Computer Vision*, (2013), pp. 721-728,
[14] Ferrari V, Zisserman A. "Learning visual attributes," *Proceedings of the Advances in Neural Information Processing Systems*, (2008), pp. 433-440.
[15] Ess A, Leibe B, Gool LV, "Depth and appearance for mobile scene analysis," *Proceedings of the IEEE 11th Int. Conf. on Computer Vision*, (2007), pp. 1–8.
[16] Layne R, Hospedales TM and Gong Sh, "Re-id: Hunting attributes in the wild," *Proceedings of the BMVC*, (2014), pp. 1709—1724
[17] Prosser B, Zheng W, Gong Sh, Xiang T and Mary Q, "Person re-identification by support vector ranking," *Proceedings of the British Machine Vision Conference*, (2010), pp. 1-11.
[18] Varior RR, Wang G, Lu J and Liu T, "Learning invariant color features for person re-identification," *IEEE Transactions on Image Processing*, 25 (2016), 3395-3410.
[19] McFee B, Lanckriet G, "Learning multi-modal similarity," Journal *of Machine Learning Research*, 12 (2011), 491-523.
[20] Barnich O, Van Droogenbroeck M, "ViBe: A universal background subtraction algorithm for video sequences," *IEEE Transactions on Image Processing*, 20 (2011), 1709-1724.