An Adaptive MCE-KISS Metric Learning Approach For Person Reidentification

Sree Vrinda G M, Dr. P.Jayaprakash

Abstract—Recently, person reidentification is getting an intensive attention in the field of Intelligent Video Surveillance (IVR). The main procedure of reidentification is to match an instance of a person captured from one camera to the instance captured from another camera. Reidentification is considered as a complex task for human tracking. Nowadays, KISS (Keep it Simple and Straightforward) metric learning algorithm and Minimum Classification Error-Based KISS (MCE-KISS) metric learning algorithm has been regarded as top level algorithm for reidentification. KISS metric learning estimated the covariance matrices by maximum likelihood (ML) estimation and the matrices are biased for large training samples. In MCE-KISS metric algorithm, it integrates MCE criterion and smoothing technique to improve the performance of KISS metric learning. Smoothing technique is done for enlarging estimated small eigenvalues while considering small training samples. Here we introduced maximum likelihood function for smoothing technique which will selects small eigenvalues in an adaptive way. Also we proposed a method for selecting appropriate features for reidentification using principal component analysis (PCA).

Index Terms— IVR, person reidentification, MCE, metric learning, LMNN, ITML, maximum likelihood.

1. INTRODUCTION

In recent years, one of the main challenging tasks in the field of intelligent video surveillance (IVR) is person reidentification. Nowadays, more number of non overlapping camera networks has been set up. These network of cameras helps for monitoring pedestrian activities over a large public area such as the parking lot, airport, metro station, etc. This identification is used to acquire an individual's complete movements along that area. An objective of person reidentification is to verify a person has been already captured by another camera networks. In previous years, traditional biometrics such as face [2], [3], iris [4], fingerprint [5], and gait [6], are used for reidentification, but now they are not using for this purpose because images taken from this are variable, low quality and contain motion blur.

- *Sree Vrinda G M is currently pursuing masters degree program in computer science and engineering in Mohandas College of Engineering, Kerala University,Kerala, India, PH-+919526354494. E-mail: vrindaplkv@gmail.com*
- *Dr. P Jayaprash is working as the Head of the Department of computer science and engineering in Mohandas College of Engineering, Kerala University,Kerala, India, PH-+919446701673. E-mail: jpavithran2011@gmail.com*

When a person stays within a camera's view, that particular person's position, background effects and lighting conditions are known to the system. The main problems arise when the person moves out from one camera's view and enters into another camera's view. So the system must know that the person seen in one camera is the same person that is already seen in another camera. If there is any issue with the system regarding matching the instance of person, that issue is known as re-identification problem. Person reidentification faces 3 main problems. At first, the segmented and comparison parts should be determined. Second, invariant signatures should be generated for comparing the corresponding parts. And at last compare the signatures by applying appropriate metric. Those steps are depicted in Figure. 1.

Reidentification problem has two methods for reidentifying, they are Appearance –based methods and gait-based methods. In appearance-based methods it extracts signatures from color, texture and other appearance properties. In gait-based methods, it extract

features from the gait and motion of persons. More importantly, gait based methods are useful because the result obtained through this are not affected by varying lighting conditions between cameras.

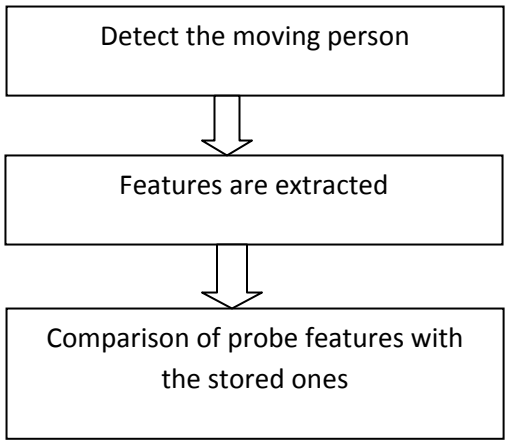

Figure 1. Steps in reidentification

Basically, two important stages which need to be focused for person reidentification are distance learning and visual feature extraction and selection. So many exciting studies have been performed on this area for improving person reidentification. Here, we briefly review some representative works. For visual feature extraction, the use of color features such as HSV and RGB color histograms will effectively save computational cost. In feature extraction, the usage of filters for clearance will improve the performance. Gabor [7] and Schmid [8] filters have been added to extraction. The texture extraction process can use Scale Invariant Feature Extraction (SIFT) [9] or Speeded up Robust Features (SURF) [10]. It is better to use LBP descriptor at last for texture classification to exploit person reidentification. This LBP (Local Binary Pattern) [11] descriptor is commonly used for facial image description and used here for getting local geometric structure of an image.

The stage distance learning can significantly improve the performance of retrieval applications. In this paper, robust distance learning is applied to update the retrieval precision. Traditionally several approaches are used for image retrieval applications but still they shouldn't produce any good result for person reidentification. Among other approaches, KISS metric learning is efficient and effective. In KISS metric learning, the results are estimated by evaluating covariance which will produce more accurate results robustly.

In this paper, we introduce the minimum classification error (MCE) [1] based KISS metric distance. The eigenvalues of the covariance matrices are biased in KISS metric algorithm. Because of biased matrices, reidentification will lead in poor performance. The covariance matrices of KISS are estimated using maximum likelihood (ML) estimation. But with increasing number of training samples, MCE criterion is more preferable than classical ML estimation. The MCE criterion technique does not work well with small eigenvalues in the covariance matrices. Therefore, the smoothing technique is required to improve the small eigenvalues of a covariance matrix. Here, we introduced maximum likelihood function in a logarithmic way for smoothing technique. Maximum likelihood function selects a parameter from a set of parameters based on probability. In MCE-KISS metric algorithm the features extracted from the images are completely used for the reidentification purposes. In the proposed technique, we introduced Principal Component Analysis (PCA) for feature selection.

IJSER

The main steps for MCE-KISS-based person reidentification can be summarized by the following steps:

1) Images are partitioned into a regular grid of size 8 x 4, and the color features and texture features are extracted using PCA.

2) Concatenating the feature descriptors together

3) Training MCE-KISS with smoothing technique; and

4) Finally finding the matching rank.

The main contributions of Adaptive MCE-KISS are;

- The proposed MCE-KISS algorithm integrates MCE criterion and smoothing technique to improve the performance of KISS metric learning.
- KISS metric integrates with maximum likelihood function for smoothing technique.
- The proposed technique used PCA for selecting appropriate features from the images.
- The newly proposed MCE-KISS exploits a learning procedure to adjust the parameters of Gaussian density model effectively.

2. LITERATURE REVIEW

In Section I, we briefly reviewed the techniques used in person reidentification. Li *et al*. [12] proposed a new approach called latent SVM technique. This technique uses clothing attributes as real value variables for comparing the person pairs. Geng et al. [13] introduced region based feature selection and feature fusion method for reidentification. In feature fusion, the method represents different features for each region which increase the computational complexity. Ma et al. [14] formulate multitask distance metric learning problem in camera networks. The proposed method is named as Multi task Maximally Collapsing Metric Learning (MtMCML) which is built on multiple Mahalanobis distance metric.

Recently, Zhao et al. [15] proposed unsupervised salience learning for reidentification. Unsupervised salience learning gives better feedback for pedestrian matching because human salience is used here as a descriptor. Adaptive Ranking Support Vector Machine (AdaRSVM) is used for person reidentification without person labels and it is introduced by J. Ma et al. [16]. But it only considers single source domain cameras. Metric learning methods are used for several person reidentification algorithms. Li et al. [17] introduced Common Near Neighbor Analysis (CNNA) using metric learning method. In addition to metric learning schemes, Tao et al. [1] added MCE (Minimum Classification Error) criterion to improve the performance of reidentification. He named his technique as MCE based KISS metric learning.

3. PROPOSED SYSTEM

3.1 KISS Metric Learning

KISS metric learning is proposed recently for the best retrieval performance in real world applications such as person reidentification, face recognition etc. It is based on the assumption that pair wise differences are Gaussian distributed.

Consider a set of image pair samples for person reidentification problem and extract feature descriptors from these. It is known that both texture features and color histograms are useful for person reidentification. Principal

IJSER

Component Analysis is used for feature selection. PCA has the potential to select a number of important individuals from all the feature components. When evaluating the significance of the feature components, the proposed method takes a number of eigenvectors into account. To find the principal component, it's needed to take maximum variance with these feature data points. After extraction, all the feature descriptors are concatenated together and these feature vector pairs are split into two sets as test and train sets using random permutation.

Consider S_a and S_b represent the samples of feature vector pair. Here two hypotheses are mentioned as H_1 and H_2, where H_1 can assume that the feature vector pair is dissimilar, i.e., S_a and S_b are sampled from different people, and the hypothesis H_2 can assume that the feature vector pair is similar, i.e., S_a and S_b are sampled from same person.

$$\delta(S_a, S_b) = log\left(\frac{p(S_a, S_b|H_1)}{p(S_a, S_b|H_2)}\right) \quad (1)$$

Equation (1) defines the logarithm of ratio between the two hypotheses. For metric learning, a small δ (S_a , S_b) indicates the two samples represent same person, while large δ (S_a , S_b) indicates the two samples represent different people. Define X_{ab} as an indicative variable of S_a and S_b : X_{ab} =1 if S_a and S_b are the same person, otherwise X_{ab} =0.

Let P_1 denote the number of similar feature vector pairs and P_0 denotes the number of dissimilar feature vector pairs. The covariance matrices are estimated as:

$$\sum_0 = \frac{1}{P_0} \sum_{X_{ab}=0} (S_a - S_b)(S_a - S_b)^T \quad (2)$$

$$\sum_1 = \frac{1}{P_1} \sum_{X_{ab}=1} (S_a - S_b)(S_a - S_b)^T \quad (3)$$

Equation shows that the eigenvalues of $\sum_0$ and $\sum_1$ are positive.

Finally the KISS metric matrix M is calculated by, $\sum_1^{-1} - \sum_0^{-1}$

3.2 Adaptive MCE - KISS Metric Learning

KISS has largely improved the accuracy of person reidentification, but still there is a lot to improve its efficiency and stability. The result of matching can be improved by improving the accuracy of covariance matrices. Sometimes small eigenvalues must appear in covariance matrices and this will cause estimate errors. To avoid this errors, smoothing technique and MCE criterion are introduced to improve the accuracy of estimate of covariance matrices in KISS. By enlarging the estimated small eigenvalues of a covariance matrix, the smoothing technique can compensate for the decrease in performance which arose from the estimate errors of small eigenvalues.

The covariance matrix $\sum_a$ is first diagnalized and can be written as

$$\sum_a = \varphi_a \wedge_a \varphi_a^T \quad (4)$$

where $\wedge_a = diag[\lambda_{a1}, \lambda_{a2}, \ldots, \lambda_{an}]$ with λ_{ab} being an eigenvalue of $\sum_a$, $\phi_a = [\phi_{a1}, \phi_{a2}, \ldots, \phi_{an}]$ with ϕ_{ab} being an eigenvector of $\sum_a$.

Recently, Tao et al. [1] proposed MCE-KISS by applying smoothing technique and in that the small eigenvalues of covariance matrix are replaced with a constant value. This constant is set to the value of the average of all the small eigenvalues. Here we used smoothing technique with maximum likelihood function. Likelihood function estimates a parameter in an adaptive way from a set of statistics based on probability. In maximum likelihood function, it selects the

set of values of the parameters that maximizes the likelihood function. Most importantly, the usage of logarithm achieves the maximum value from the likelihood function.

According to MCE metric learning, we need to minimize the empirical loss by updating the parameters via gradient descent method. We can compute the empirical loss by using (5)

$$E = \frac{1}{N} \sum_{n=1}^{N} l_{ci}(X_i) \qquad (5)$$

where 'ci' is the class information, N is the maximum number of training samples. The loss of misclassification can be estimated by using the following equation.

$$l_{ci}(x) = \frac{1}{1+e^{-\xi M_{ci}(x)}} \qquad (6)$$

where ξ is a trade-off parameter and is selected in the range of (0, +∞]. Here we have the evaluation of misclassification of a sample x belonging to class 'T'

$$M_T(x) = max_T\ \delta(x, x_T) - min_r\ \delta(x, x_r) \qquad (7)$$

where x_T is a sample of the class T, and x_r is the closest interclass sample. First element in (7) represents the distance between x and the farthest intraclass sample and second element represents the distance between x and the closest interclass sample.

The parameters in MCE-KISS metric algorithm include the eigenvalues, eigenvectors and the updated constant values through smoothing technique i.e., λ_{1n}, λ_{0n} , β_0, β_1, ϕ_{1n} and ϕ_{0n}. Before updating we need to make sure that eigenvalues are positive, so we define

$$\begin{cases} \lambda_{an} = e^{\sigma_{an}} \\ \beta_a = e^{\tau_a} \end{cases} \qquad (8)$$

$$\begin{cases} \sigma_{an} = \ln \lambda_{an} \\ \tau_a = \ln \beta_a \end{cases} \qquad (9)$$

The parameters of covariance matrices $\sum_0^{-1}$ *and* $\sum_1^{-1}$ are optimized using (10), (11) and (12).

$$\begin{cases} \dfrac{\partial \delta(X, X_b)}{\partial \sigma_{1n}} = -e^{-\sigma_{1n}} [\, \varphi_{1n}^T (X - X)]^2 \\ \dfrac{\partial \delta(X, X_b)}{\partial \sigma_{0n}} = -e^{-\sigma_{0n}} [\, \varphi_{0n}^T (X - X_b)]^2 \end{cases} \qquad (10)$$

$$\begin{cases} \dfrac{\partial \delta(X, X_b)}{\partial \tau_1} = -e^{-\tau_1} [\, ||X - X_b||^2 - \sum_{n=1}^{C} [\, \varphi_{1n}^T (X - X_b)]^2] \\ \dfrac{\partial \delta(X, X_b)}{\partial \tau_0} = -e^{-\tau_0} [\, ||X - X_b||^2 - \sum_{n=1}^{C} [\, \varphi_{0n}^T (X - X_b)]^2] \end{cases} \qquad (11)$$

$$\begin{cases} \dfrac{\partial \delta(X, X_b)}{\partial \varphi_{1nl}} = 2(e^{-\sigma_{1n}} - e^{-\tau_1}) [\, \varphi_{1n}^T (X - X_b)] (X - X_b)_l \\ \dfrac{\partial \delta(X, X_b)}{\partial \varphi_{0nl}} = 2(e^{-\sigma_{0n}} - e^{-\tau_0}) [\, \varphi_{0n}^T (X - X_b)] (X - X_b)_l \end{cases} \qquad (12)$$

Based on the above discussions, we can compute the distance metric by $\sum_1^{-1} - \sum_0^{-1}$

4. RESULT ANALYSIS

For implementing the proposed algorithm, here we used VIPeR [18] dataset which contains 1264 images that are taken from two camera networks. That is 632 images from camera A and 632 images from camera B. Then the images are normalized to a standard size of 128 x 48. In our experiments, half of the samples were selected to form test set, while the rest were used for model training. During training, we used intraperson image pairs as similar pairs, and generated interperson image pairs (by randomly selecting two images from different subjects) as dissimilar pairs. The image pairs are used to estimate $\sum_0^{-1}$ and $\sum_1^{-1}$. During testing, the tests set were divided into two parts, i.e., a gallery set and a probe set. We randomly chose

IJSER

one sample of each subject to comprise the gallery. The rest were used for the probe set. Person reidentification [19] aims to identify a person's photo in the probe set by comparing it with images of several individuals stored in the gallery set.

All the images are then partitioned into a regular grid with 8 pixel spacing in the horizontal direction, and 4 pixel spacing in the vertical direction. From the grid, the LBP descriptor [20], HSV histogram, and RGB histogram were extracted from overlapping blocks of size 8 x 8. The HSV and RGB histograms encoded the different color distribution information in the HSV and RGB color spaces, respectively. The texture distribution information was modeled effectively by LBP descriptor. All the feature descriptors were concatenated together. After extraction PCA is conducted to select the most important features from the whole set of features. Through this here we took 134 features into account.

4.1 KISS Metric Learning

In KISS metric learning approach, first of all two feature vector pairs are selected which represents the image pair samples. Based on the sample image pairs the covariance matrices are evaluated. These covariance matrices are used to evaluate KISS matrix; $\sum_1^{-1} - \sum_0^{-1}$

The cumulative match characteristic curves (CMC) are used to evaluate the performance of algorithms. Because of the complexity of the reidentification problem, the top n-ranked matching rate was considered. The following figure shows the CMC curve for KISS metric learning algorithm. In this figure, x-coordinate represents the rank score and the y-coordinate represents the matching rate.

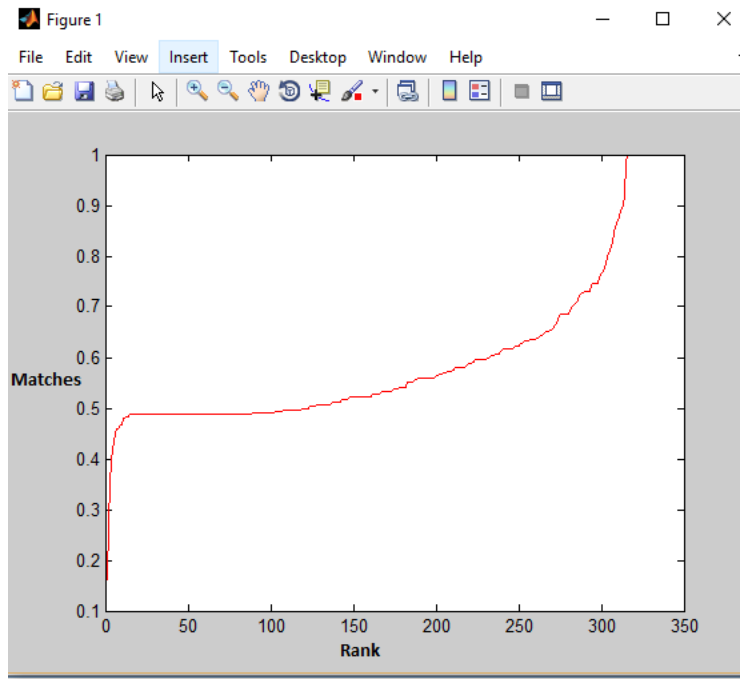

Figure 2. CMC for KISS matrix

4.2 Adaptive MCE - KISS Metric Learning

The covariance matrices of KISS are obtained by maximum likelihood (ML) estimation. With increasing the number of training samples, discriminative learning based on MCE is more reliable than classical ML estimation. The introduction of MCE criterion to the training procedure does not work well to estimate the small eigenvalues of the covariance matrices. Therefore, the smoothing technique is required to improve the estimate of the small eigenvalues of a covariance matrix. Smoothing technique is done using Maximum Likelihood function. It selects the set of values of the parameters that maximizes the likelihood function. Here log likelihood function used, so

that it achieves maximum value from the selected small eigenvalues.

IJSER

	1	2	3	4	5	6	7	8	9	10
1	1.8376 + 0.0...	1.0916 + 0.0...	-0.3451 + 3....	1.7608 + 3.1...	-0.0829 + 0....	2.6914 + 0.0...	0.7163 + 0.0...	-0.6024 + 3	4.6257 + 0.0..	.9518 + 3.1..

Figure 3. Selected small eigenvalues

The algorithm selects small eigenvalues in an adaptive way which is shown in the following Figure 3. In this figure it is noticeable that a particular eigenvalue is marked. From the set of small eigenvalues, maximum likelihood function selects this particular value for replacing with the other small eigenvalues. After smoothing technique, optimize the parameters of covariance matrices and finally the Adaptive MCE-KISS matrix is calculated.

The CMC curve of Adaptive MCE-KISS also shows the performance rate. In this figure, x-coordinate represents the rank score and the y-coordinate represents the matching rate. The figure represents the performance comparison using CMC curve. Here, we compare the proposed adaptive MCE-KISS algorithm with the existing MCE-KISS algorithm and KISS metric learning algorithm. From this we can understand that KISS metric learning algorithm performs poorly for person reidentification than MCEKISS and adaptive MCE-KISS algorithms. Adaptive MCE-KISS algorithm helps to improve the matching rate with increasing number of image samples than the previous MCE-KISS algorithm.

Based on the performance of the algorithm, here we also compared top matching rates. The comparison is done on the same dataset with the popular algorithms such as Adaptive MCE-KISS, MCE-KISS algorithm and KISS metric learning algorithm. Table I reports the performance of all the algorithms in the scope of first 200 ranks. In most of the cases Adaptive MCE-KISS performs best in terms of rank score.

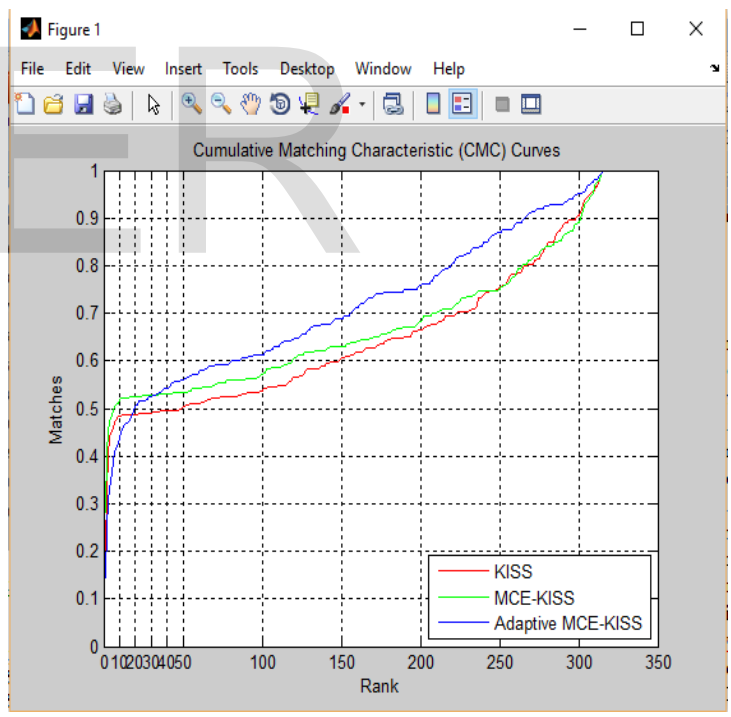

Figure 4. CMC curve for Adaptive MCE-KISS matrix

Rank	10	50	100	200	300
KISS	0.44	0.5	0.54	0.64	0.9
MCE-KISS	**0.51**	0.54	0.56	0.68	0.9
Adaptive MCE-KISS	0.42	**0.56**	**0.62**	**0.76**	**0.96**

Table I. Top matching rates

IJSER

Considering the computational complexity, KISS metric learning algorithm shows the less training time than the other two algorithms. The following figure shows the computational complexity of the algorithms.

```
KISS :
Elapsed time is 3.254601 seconds.
MCE-KISS :
Elapsed time is 11.948347 seconds.
Adaptive MCE-KISS :
Elapsed time is 13.223791 seconds.
```

Figure 5. Computational complexity

5. CONCLUSION

In recent years, distance metric algorithms are developed for effective person reidentification. ITML [21] and LMNN [22] algorithms have been developed but still these are suitable only for reidentifying limited training samples. Here we introduce MCE-KISS metric algorithm which works well in large training samples. In KISS algorithm, the estimated covariance matrices are biased because of small number of training samples. The proposed technique exploits the smoothing technique in maximum likelihood function to enlarge small eigenvalues in the estimated covariance matrix. In addition to that Principal Component Analysis (PCA) is conducted to select the most important features from the list of extracted features. Hence, the proposed algorithm significantly improves the performance of MCE-KISS for person reidentification.

REFERENCES

[1] D. Tao, L. Jin, Y. Wang, and X. Li, "Person reidentification by minimum classification error-based KISS metric learning," *IEEE Trans. Cybern.*, 2014, 10.1109/TCYB.2014.2323992.

[2] V. Chatzis, A. G. Bors, and I. Pitas, "Multimodal decision-level fusion for person authentication," *IEEE Trans. Syst., Man, Cybern. A, Syst., Humans*, vol. 29, no. 6, pp. 674–680, Nov. 1999.

[3] F. Dornaika and A. Bosaghzadeh, "Exponential local discriminant embedding and its application to face recognition," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 921–934, Jun. 2013.

[4] R. M. Da Costa and A. Gonzaga, "Dynamic features for iris recognition," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 4, pp. 1072–1082, Aug. 2012.

[5] R. Cappelli, M. Ferrara, and D. Maio, "A fast and accurate palmprint recognition system based on minutiae," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 42, no. 3, pp. 956–962, Jun. 2012.

[6] I. Venkat and P. De Wilde, "Robust gait recognition by learning and exploiting sub-gait characteristics," *Int. J. Comput. Vis.*, vol. 91, no. 1,pp.7–23,2011.

[7] I. Fogel and D. Sagi, "Gabor filters as texture discriminator," *Biol.Cybern.*,vol.61,no.2,pp.103–113,1989.

[8] C. Schmid, "Constructing models for content-based image retrieval," in *Proc. IEEE CVPR*, 2001, pp. 39–45.

[9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[10] X. Liu *et al.*, "Attribute-restricted latent topic model for person reidentification,"*Pattern Recognit.*, vol. 45, no. 12, pp. 4204–4213, 2012.

[11] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[12] Annan Li, Luoqi Liu, Kang Wang, Si Liu and Shuicheng Yan, "Clothing Attributes Assisted Person Re-identification," IEEE Transactions on ircuits and Systems for Video Technology, Vol:25, Issue:5, May 2015

[13] Yanbing Geng, Hai-Miao Hu, Jin Zheng and Bo Li. "A person re-identification algorithm by using region-based feature selection and feature fusion," IEEE International Conference on Image Processing *(ICIP)*, 2013.

[14] L. Ma, X. Yang, and D. Tao, "Person re-identification over camera networks using multi-task distance metric learning," *IEEE Trans. Image Process, vol. 23, no. 8, pp. 3656-3670, Aug. 2014.*

[15] R. Zhao, W. Ouyang, and X. Wang, "Unsupervised salience learning for person re-identification," in *IEEE Conf. Computer Vision and Pattern Recognition, CVPR, 2013, pp. 4321-4328.*

[16] Andy J. Ma, Jiawei Li, Pong C. Yuen and Ping Li, "Cross Domain Person Reidentification Using Domain Adaptation Ranking SVMs," IEEE Transactions on Image Processing, vol. 24, no. 5, May 2015.

[17] LI, W., WU, Y., MUKUNOKI, M., AND MINOH, M. 2012. Common-near-neighbor analysis for person reidentification. In Internationl Conference on Image Processing, 1621-1624.

IJSER

[18] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking,"in Proc. 10th PETS, 2007.
[19] W. R. Schwartz and L. S. Davis, "Learning discriminative appearancebased models using partial least squares,"in Proc. SIBGRAPI, Rio de Janiero, Brazil, 2009, pp. 322-329.
[20] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns,"IEEE Trans. Pattern Anal. Mach. Intell., vol. 24, no. 7, pp. 971-987, Jul. 2002.
[21] J. V. Davis, B. Kulis, P. Jain, S. Sra, and I. S. Dhillon, "Informationtheoretic metric learning,"in Proc. ICML, Corvallis, OR, USA, 2007, pp. 209-216.
[22] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification,"J. Mach. Learn. Res., vol. 10, pp. 207-244, Feb. 2009.

IJSER