CONTEXT-BASED FEATURE-LEVEL FUSION METHOD FOR PERSON RE-IDENTIFICATION

M. L. Sworna Kokila[a], Dr. V. Gomathi[b]
[a]*Ponjesly college of engineering, Nagercoil.*
[b]*National Engineering College, Kovilpatti*
[a]*kokilaolive@gmail.com*

Abstract: This work proposes an improved technique for improving the person re-identification task by considering the multi-view based context-aware feature sets to match the probe person image among the gallery of image sets, with or without the identity labels. The hybrid feature-level fusion method was employed by extracting the context-based global color histograms and dense SIFT features. The Multi-SVM models for labelling the person's identity related to the global and local features with context-aware kernel distance metric wascarried out in this work and proficient results were produced, when compared with other feature extraction techniques. The colour and dense SIFT based feature extraction technique was found to be proficient in the re identification and yields an efficiency of 91.2% for viper data set and 84.9% for Grid data set, when coupled with the multiclass SVM.

Keywords: Global Color Histogram, Dense SIFT, Multi Class SVM.

1. Introduction

For any multi-camera networked area under visual surveillance, there is a great scope that the vicinity of the same person will be available with different camera views, orientation, size and either full or partial occlusions based on the context of the scenes. These video footages are providing richer scope for the person re-identification task to be accomplished with robustness and accuracy of identifying the repetitive visual tracks of the person of interest. In the past decade, the person re-identification research work has become popular and relies on the multiple tracks of a person appearing across the non-overlapping camera setups that are distributed at different physical locations.

Conventionally, the objective of person re-identification is to re-identify the same person captured by one camera in yet another camera that was located at a new location. It is treated as the process of identifying the distinct person from the video captured by the surveillance camera. This approach has greater usefulness in the suspecting / loitering areas such as airports, metro stations, railway stations, etc. based on their context-based recordings. In a simplest way, theperson re-identification task was carried out using themodel that can record the macroscopic characteristics [1]. The extractions of macroscopic features are difficult due to the low frame-rates and resolutions of multiple surveillance cameras. Recent advancements in camera with RGB-3D depth sensorsis possible to extract the person's soft-biometric features such as shape orientation profiling, height, and the lengths of limbs.Though there are several researchers who have contributed in the field of person re-identification, there still exist several research issues such as, handling of ***appearance variations*** in terms of arbitrary poses, changing illumination, uniforms, and occlusion across the cameraviews.

The gallery formation for person re-identification task normally contains the images of labelled persons from one particular camera view meanwhile the probes are the same person's coverage from another camera view. For the recognition of a given probe froma large gallery, the key concept is to first extract robust features for both probe and gallery images. The features are then compared to determine the percentage of matching. This kind of matching is called ***appearance-based*** and it makes use of visual cues only. For the same person, features having low intra class variation are extracted and for the different persons, features having high inter-class variation are extracted.

A conventional person re-identification system usually takes a probe image as input, which has the appearance of a person as a whole, and generates output in the form of similarity score based ranking between the input and among all gallery images. Sometimes, the output will also be a best matched person's label from the classifiers. The contributions related to person re-identification task have been widely studied in recent years. Research initiatives focused mainly on either extracting improved features for labelling or finding the distance metric for improved similarity measures [1] or a combination of both. In addition, there exists a pressing need for identifying the suitable and discriminative features that are invariant to environmental conditions like illuminations, background, occlusions and camera view-point changes bring major role in improved performance [2].The various feature extraction techniques are there in literature like histogram of color images,local binary patterns, Gabor features,features based on dominant colors and that relies on spatial neighbourhood connectivity. The combination of spatial and colouring information that relies on localized features coupled with classifiers are also widely used for person reidentification[3,4]. Some of the works relies on the features from the body and extracts in keeping with-part color details as well as colour displacement in the whole body to deal with pose variation. Kviatkovsky et al [5] put forward illumination invariant characteristic illustration primarily based at the log chromaticity(log) color area and demonstrate that shade as a single cue was found to be efficeint in figuring out persons under greatly various imaging situations.

The wavelet transform along with the graph theory approach was used for the face recognition from large data base. The bunch graph technique was employed and was validated on public database [6]. The support vector machine classifier along with the feature extraction by local binary pattern was used for the gender classification [7]. The elastic graph matching technique was employed for the face recognition and gender classification [8]. The SIFT features was found to generate efficient results for face authentication in [9], 3 different matching strategies are employed. In [10], a detailed analysis has been performed in the object recognition using SURF features. The feature extraction by SURF was found to be proficient in the face recognition, when compared with the SIFT features. The proposed approach in [11] generates efficient results for different face poses, when compared with the classical approaches. A novel face recognition technique that replaces the gabor features by HOG features was proposed in [12].

The elastic bunch graph matching was employed for the face identification. The SIFT feature extraction along with the classification by fast nearest-neighbour algorithm and hough transform was employed in [13]. The final validation was done by least-squares solution. The classical SIFT feature was modified by the incorporation of PCA termed as PCA-SIFT was found to be proficient in improving the efficiency of image retrieval application [14]. The SIFT flow algorithm was proposed in [15], robust results were produced for image identification, when compared with the classical SIFT approach. In [16], a detailed analysis has been performed in the performance validation of feature extraction techniques.In order to improve the performance, multiple features are combined and that will generate good accuracy in the person reidentifcation task. By the coupling of multiple features, the dominant feature that is used to generate accurate classification can be analysed. The hybrid fusion of global and local features is described in [17] for person identification using deep learning architecture. A graph based feature fusion scheme was proposed in [18] for person re identification. An improved Siamese network with hard sample was proposed in [19] for person reidentifcation, a fusion module comprising of local and global features was employed in this work. The details about the proposed methodology were presented in section 3. It mainly focuses on building a Multi-SVM model for labelling the person's identity by extracting their global and local features with context-aware kernel distance metric.

First, the brief reviews about the existing related works are reported in section 3.2. Then, section 3.3 describe the proposed methodology and section 3.4 introduces the experimental datasets and section 3.5 analyses the performance evaluation of the proposed work and finally conclusions re derived in section 3.6.

2. Materials and Methods

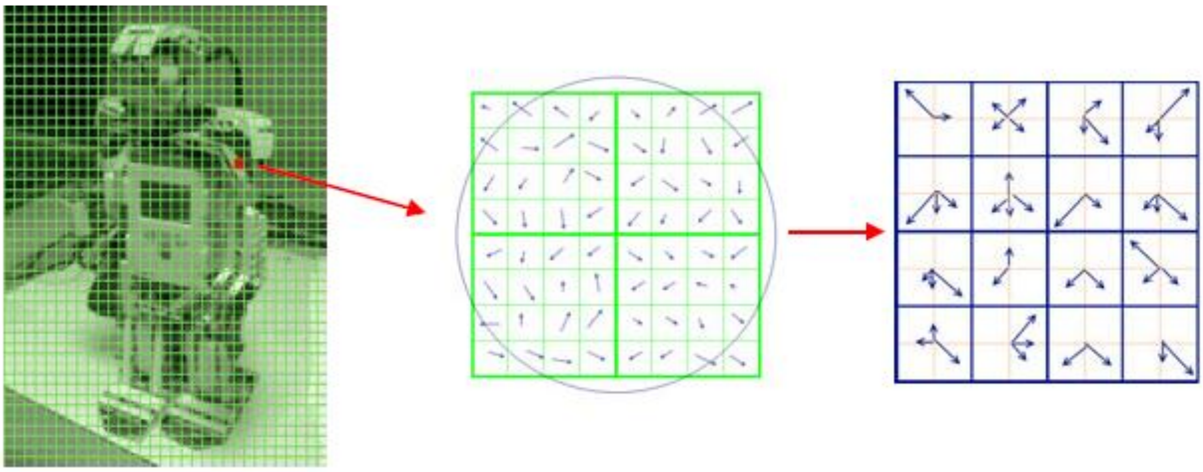
This work handles the ***appearance variations*** during person re-identification by considering the VIPER data set that inherits the arbitrary poses, changing illumination, uniforms, and occlusion across the camera views.The major phases of the proposed work are, (i) Data sets for probe and gallery; (ii) feature descriptors for person object; (iii) Multi-SVM design; (iv) Experimental analysis of performance. As reported in related work, after careful analysis of feature contributions, in this work both weighted global colorhistogram (w-GCH) and Dense Scale Invariant Feature Transform (D-SIFT) are identified as the feature descriptors**.**

2.1 Colour Histogram

The proposed method used Global Color Histogram (GCH) features toidentify similarity in the images. First we discretize color space into n colors by using 8*8*8 = 512 color instead of 256*256*256= 16777216 and then bin was created for each color and store number of pixels for each color in histogram's bin.GCH represents one whole image with a single colorhistogram.But this histogram is not unique and also not robust. Because the color distribution of two different images are similar means histogram also similar.To increase uniqueness and robustness color histogram features are fused with DSIFT.

2.2 Dense SIFT

The input image is divided into sub locks or kernel of size 8×8, each of the sub block is further divided into 4×4 blocks, depicted in figure 1. The final block comprises of 16 elements and the gradients are computed in the respective directions [20,21]. The quantization is performed with respect to the gradient values in the 8 directions and it consist of 4x4x8=128 dimensional SIFT features.



(a) (b) (c)

Figure 1: (a) Sample image, (b) Gradient estimation, (c) Cell descriptor

The 4 × 4 descriptor with 8 orientations, comprising of feature vectors with 128 dimensions is used here. The dimensionality of the descriptor is high;however, the performance is better than lower dimensional descriptor with low computational complexity.

2.3 Distance Measure and ReIdentification Score

This work proposes fusion of GCH and dense SIFT features.The motivation for this work is to exploit the fusion of global and local appearance information involved in a probe image to improve its retrieval accuracy in the gallery. Let the probe image p and the gallery image set as G = {g_1, g_2, gN} with N training samples. Form these N training samples, GCH and D-SIFT features are extracted and concatenated, which bring up the dimensional feature space as 256 and 100 respectively.

F = [GCH_{gi}, D-$SIFT_{gi}$]

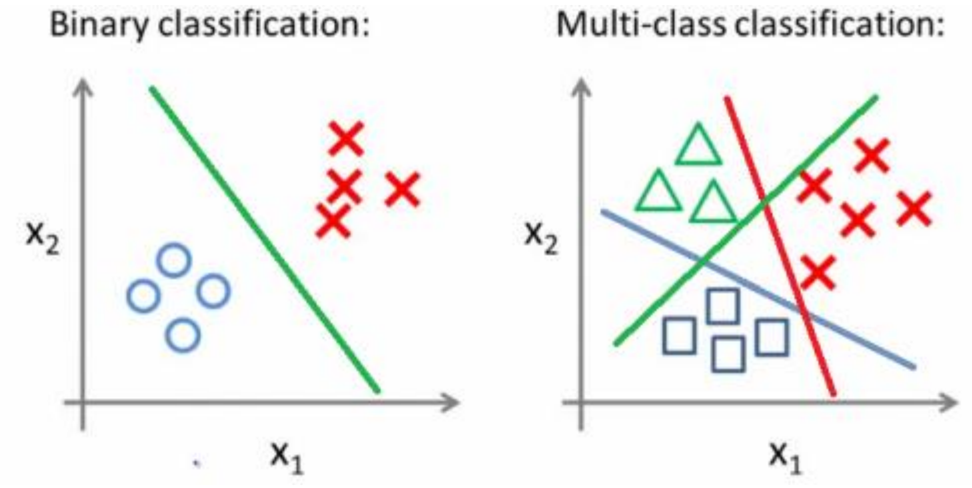
To measure the similarity metric between the probe and gallery images, the multi-SVM classifier model is proposed in this work. Because, this multi-SVM model is being build over the kernel distance metric as the basis for discriminating the feature space. The features extracted from the N training samples along with their person-labels from P_1 to P_m, where m is the number of subjects considered during training.

Multi-SVM can optimize the learning patterns of linear SVMs. For the training set (x1,y1) ... (xn,yn) with labels yi in [1..k], the solution of the optimization problem is estimated. A retrieval score was estimatedin accordance with the following evaluation criterion: the testing stage determines the efficiency of the process and the number of mismatches are determined.

2.4 Multi-SVM

The multi-svm classifier learns from the training set, the margin parameters w_i and ξ_i for each person as they contribute by discriminating the labelling task during re-id ranking. The multiclass SVM is depicted in figure 2. The constraint to be satisfied for Multi-SVM based ranking approach is as given below:

$$\forall i \forall y \neq y_i \; w^T \Phi(\boldsymbol{x}_i, y_i) - w^T \Phi(\boldsymbol{x}_i, y) \geq 1 - \xi_i \qquad (1)$$



(a) (b)

Figure2:(a) Binary classifier, (b) Multi-class classifier

At the time of querying with the probe image, the proposed approach will return the ranked label of the person re-identity based on the following equation,

$$\text{y} = \arg\max_{y'} \boldsymbol{w}^T \Phi(\boldsymbol{x}, y') \qquad (2)$$

The model selected for the person re-identification problem here is multi-SVM with kernel of Radial Basis Function.The objective function of SVM can be written as

$$J(\alpha) = \frac{1}{m}\sum_{i=1}^{m} L\left(\sum_{j=1}^{m} \alpha_j K(\boldsymbol{x}_i, \boldsymbol{x}_j), y\right) \quad (3)$$

In the above equation, $K(\boldsymbol{x_i}, \boldsymbol{x}_j)$ is the Radial Basis Function (RBF) kernel which is representing the discriminating the similarity score for person re-identification as a kernel based distance metric during multi-SVM learning. Thus, the context-based fused features are mapped with this RBF kernel as stated below:

$$K(\boldsymbol{x_i}, \boldsymbol{x}_j) = \exp\left(-\frac{1}{2\sigma^2}\left\|\boldsymbol{x_i} - \boldsymbol{x}_j\right\|^2\right) \quad (4)$$

This RBF kernel based multi-SVM can build the margin for each person object a separate hyper-plane. During testing phase, the similar scored features will be fall in that corresponding hyper-space. Hence, the proposed multi-SVM has been proved mathematically a suitable model to outperform this re-identification task.

3. Experimental Results and Discussion

The public datasets, the VIPeR and GRID are adopted to evaluate our approach. The VIPeR dataset is a widely used data set for many applications and it’s a challenging data set too, since it comprises of images acquired from the cameras in the outdoor environment. The images in various views and poses with some of the images in low resolution are there demonstrating a real time scenario like poor background illumination. The features of data set are represented in figure 3.

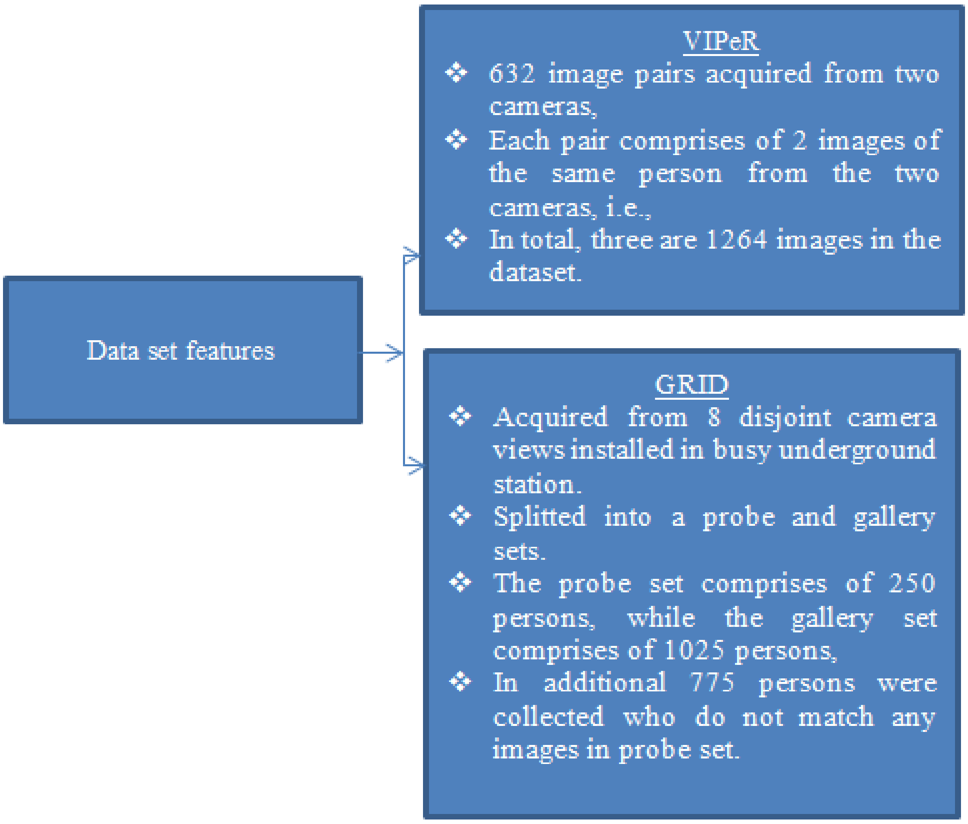
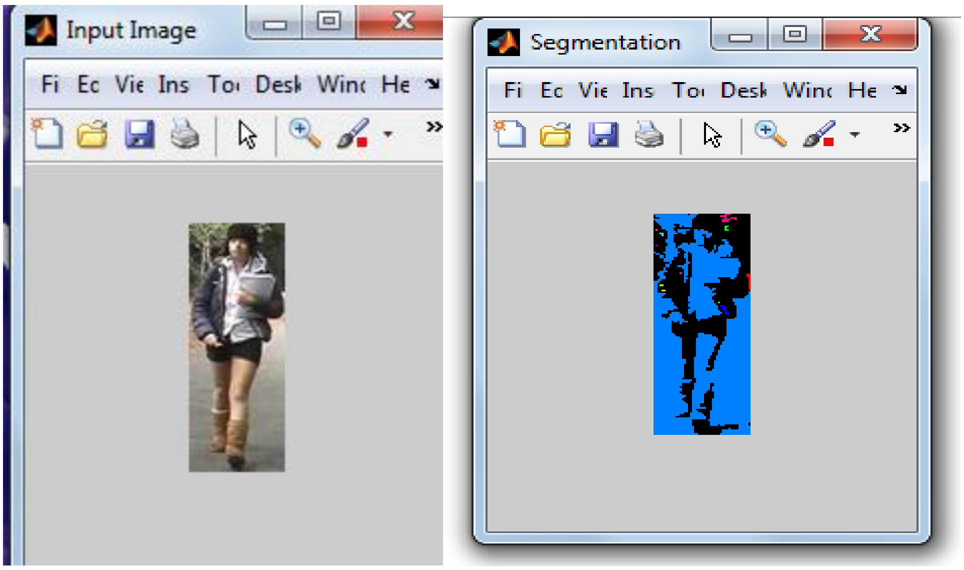


Figure 3: Features of Data sets

The dataset is challenging due to severe inter object occlusion, large viewpoint variations, and poor image quality. The algorithms are developed in MATLAB 2015 a and tested on real time standard data sets. The HSV model was utilized in the extraction of color features and quantization was done to 128 color bins. The texture feature extraction was also performed and the transform parameters were utilized to perform eight-orientation decomposition of the image at three stages of resolution.

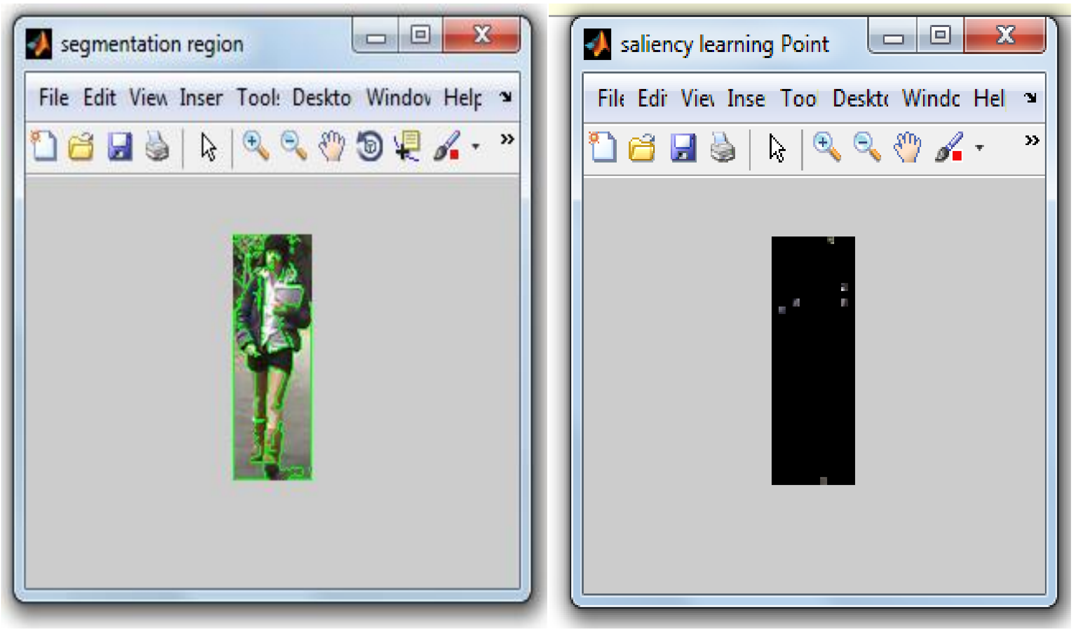


(a) (b)

Figure 4 : (a) Input Image, (b) Segmentation output

The supervised learning was employed to train the network. For training, samples are taken from the VIPER and GRID Datasets. The hybrid learning was employed for the training of network and it minimizes the error signal. The proposed system was also tested with VIPER and GRIDdatasets. After getting the query image, it used to undergo training as same in training phase.

The sample outputs from data base are depicted below. In figure 4, (a) depicts the input image and (b) depicts the segmentation output. The figure 6 depicts the boundary detection and saliency learning point.



(a) (b)

Figure 5 : (a) Boundary extraction, (b) Saliency learning point

The patched image regions and the pyramid image component are depicted in figure 6 and 7.

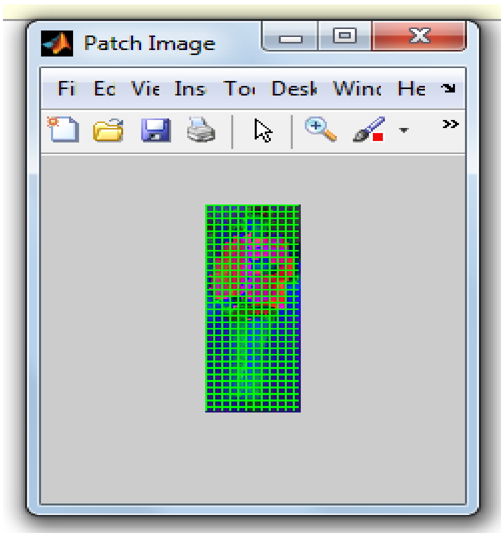


Figure 6 : Patched image regions

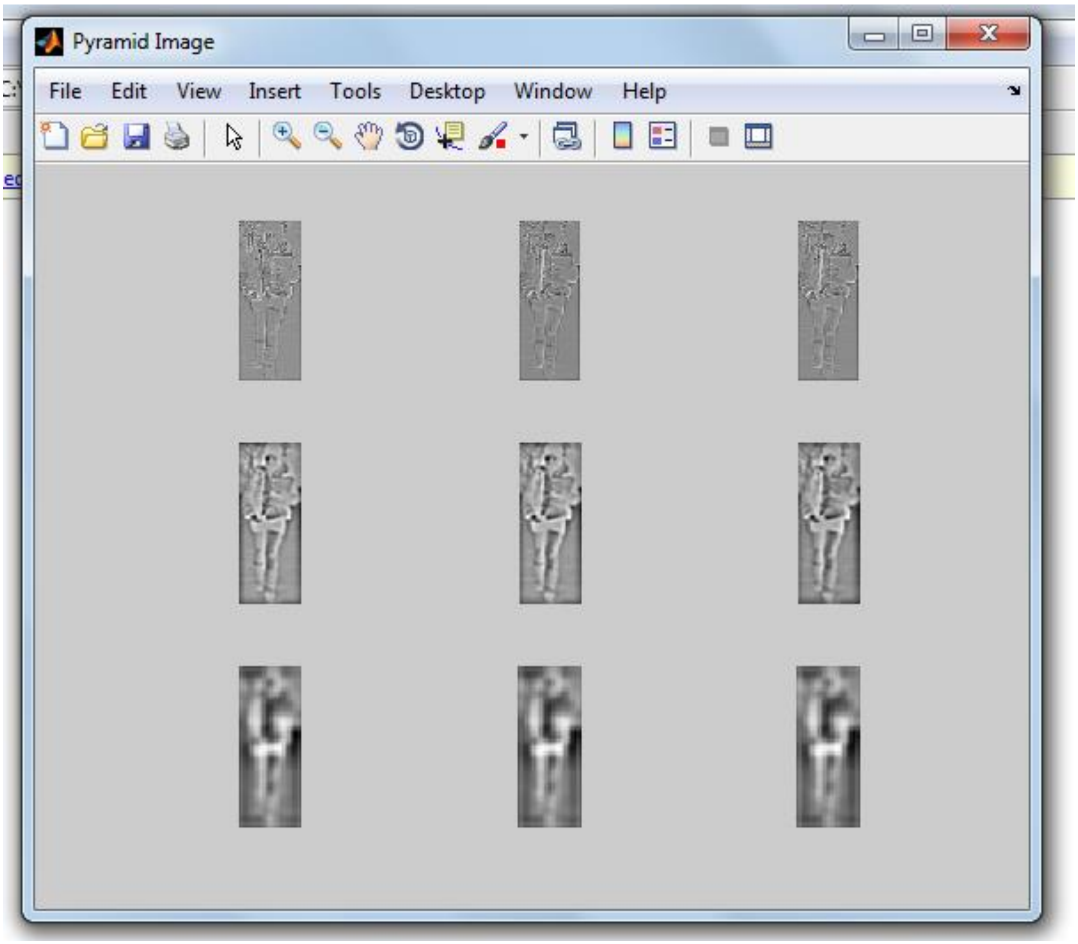


Figure 7: Pyramid image

The keypoint localization by dense and SIFT features are represented in figure 8. The figure 9 represents the matching image and the figure 10 represents the different orientation of the image.

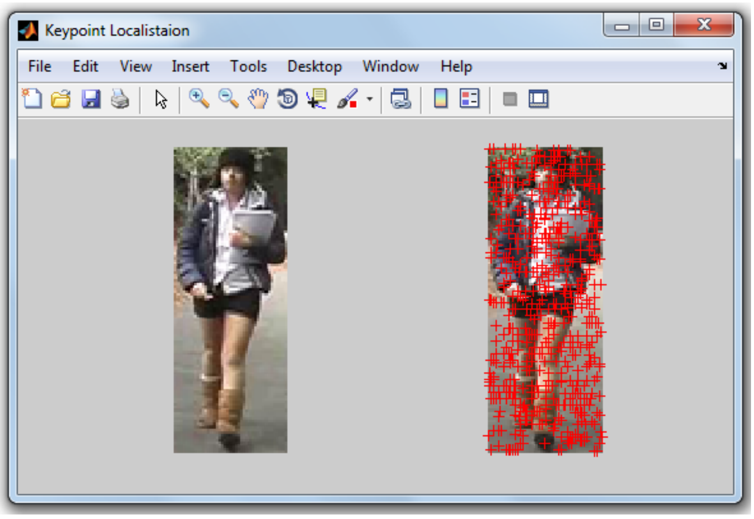


Figure 8 : Keypoint localization by color and dense SIFT features

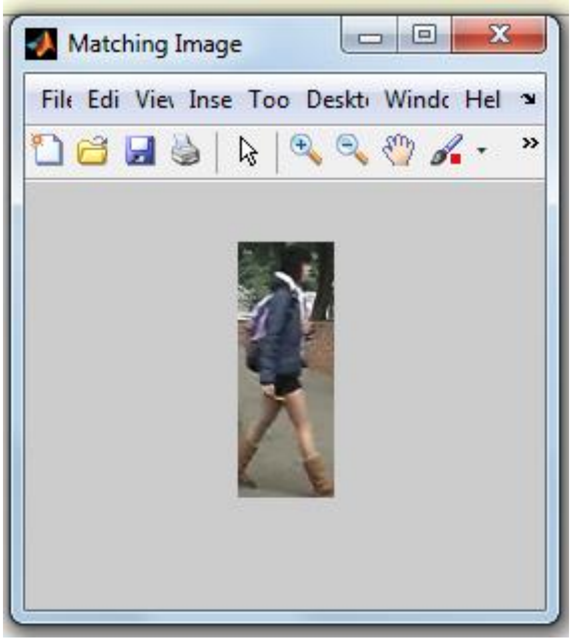


Figure9 : Matching image

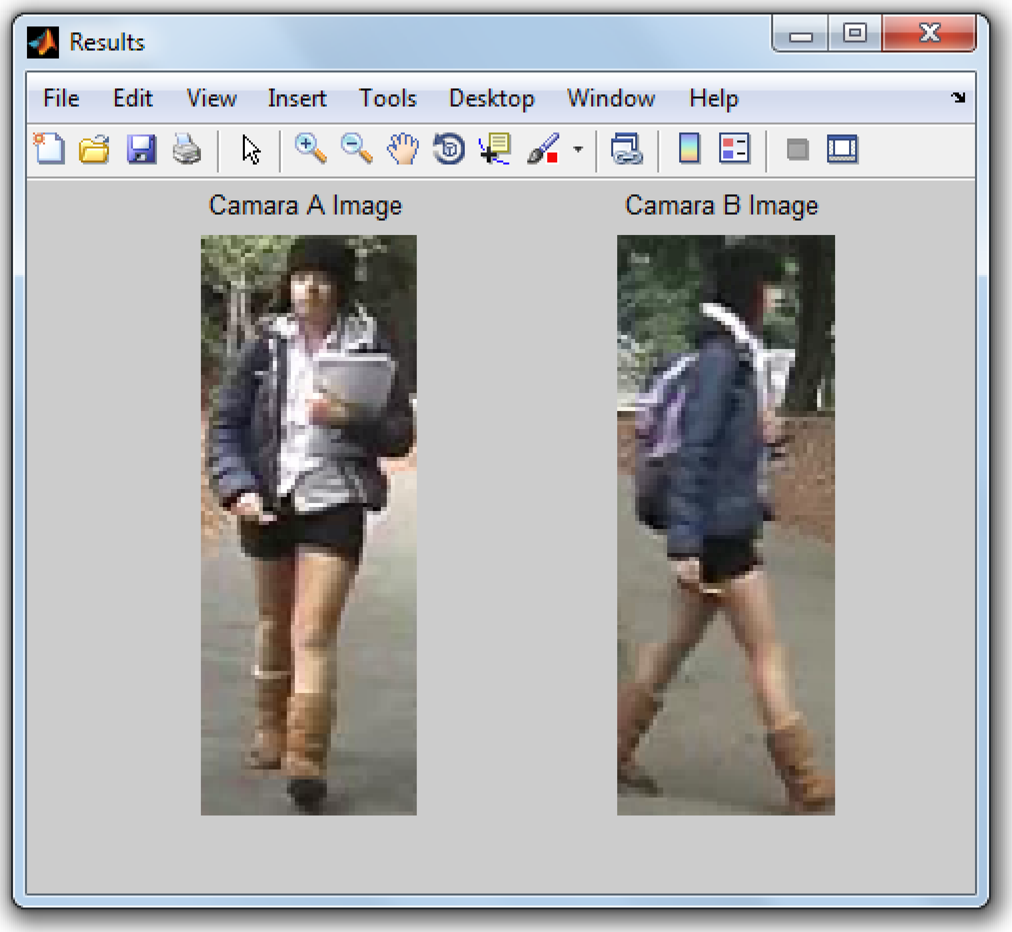


Figure 10: Different orientation of image

The comparison of different features based algorithms are there; color and dense SIFT based feature extraction technique was found to be efficient in the re identification and yields an efficiency of

91.2% for viper data set and 84.9% for Grid data set. The results reveal that, the hybrid feature combination yields superior results.The average reidentifcation score plot corresponding to VIPER data set is depicted in figure 11. The average reidentifcation score plot corresponding to GRID data set is depicted in figure 12.

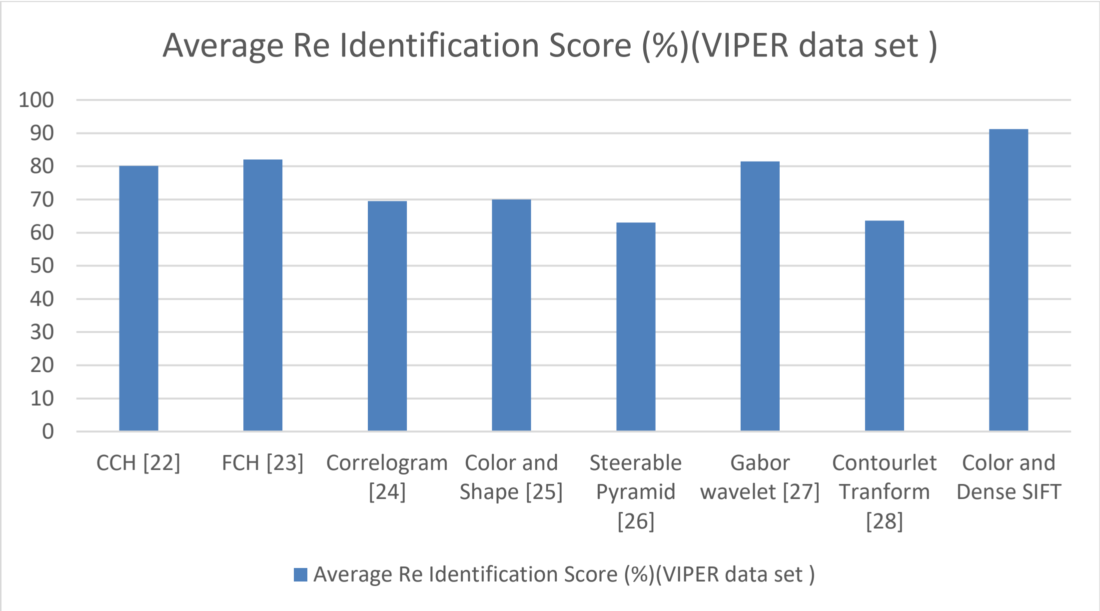


Figure 11: Average Re Identification Score (%) (VIPER data set)

The multiclass SVM was coupled with various feature extraction techniques and the results reveals that, hybrid combination of color and dense SIFT features are proficient, when compared with the other feature extraction techniques.

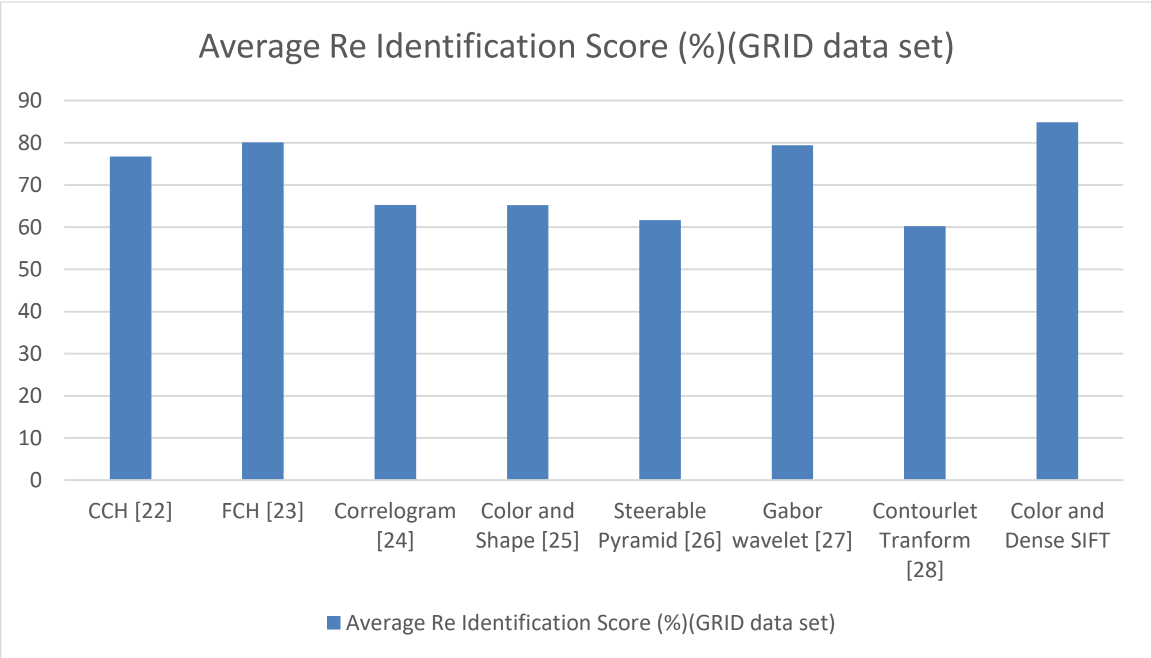


Figure 12: Average Re Identification Score (%) (GRID data set)

4. Conclusion

This chapter proposes an improved re-identification technique and efficiency has been compared with the existing techniques. The Global Color Histogram and Dense SIFT are used for feature extraction. Based on the experimental result, the proposed technique has gained the advantage of collective summarization in the false identification from the data set. The outcome of this work is applicable for real time applications and achieved 100%accuracy for training sets and 91.2 % accuracy for testing set of VIPER dataset and then 84.9% accuracy for test cases of GRID dataset. The proposed workwas found to be useful for real-world problems such as human-computer interaction, surveillanceand online-conferencing and for entertainments.

REFERENCES

[1] M. Farenzena, L. Bazzani, A. Perina, V. Murino, and M. Cristani, "Person re-identification by symmetry-driven accumulation of local features," in IEEE Conference on Computer Vision and Pattern Recognition, 2010, pp. 2360–2367

[2] D. Gray and H. Tao, "Viewpoint invariant pedestrian recognition with an ensemble of localized features," in European conference on computer vision. Springer, 2008, pp. 262–275.

[3] D. S. Cheng, M. Cristani, M. Stoppa, L. Bazzani, and V. Murino, "Custom pictorial structures for re-identification." in BMVC, vol. 1, no. 2, 2011, p. 6.

[4] B. Ma, Y. Su, and F. Jurie, "Local descriptors encoded by fisher vectors for person re-identification," in European Conference on Computer Vision. Springer, 2012, pp. 413–422.

[5] I. Kviatkovsky, A. Adam, and E. Rivlin, "Color invariants for person reidentification," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 7, pp. 1622–1634, 2013.

[6] Wiskott L, Krüger N, Kuiger N, Von Der Malsburg C. Face recognition by elastic bunch graph matching. IEEE Transactions on pattern analysis and machine intelligence. 1997 Jul;19(7):775-9.

[7] Combining Motion and Appearance for Gender Classification from Video Sequences, Pattern Recognition, 42(11): 2818-2827, 2009.

[8] L. Wiskott, J. M. Fellous, N. Kruger and C. Von der Malsburg. Face recognition and gender determination. Proceedings of IEEE Conference on Automatic Face and Gesture Recognition.pp.92-97, 1995.

[9]. M. Bicego, A. Lagorio, E. Grosso, and M. Tistarelli. On the use of SIFT features for face authentication. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06). p. 35, 2006.

[10]. H. Bay, A. Ess, T. Tuytelaars, L. V. Gool. SURF: Speeded Up Robust Features. Computer Vision and Image Understanding, Vol. 110, No. 3, pp. 346--359, 2008.

[11]. P. Dreuw, P. Steingrube H. Hanselmann, H. Ney. SURF-Face: Face Recognition Under Viewpoint Consistency Constraints. Proceedings of the British Machine Vision Conference 2007 (BMVC2007).

[12]. Wiskott L, Krüger N, Kuiger N, Von Der Malsburg C. Face recognition by elastic bunch graph matching. IEEE Transactions on pattern analysis and machine intelligence. 1997 Jul;19(7):775-9.

[13] D. G. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. IJCV, 60(2): 91-110, 2004.

[14]. Y. Ke and R. Sukthankar. PCA-SIFT: A more distinctive representation for local image descriptors. Proceedings of CVPR, pp.506–513, 2004

[15]. C. Liu, J. Yuen, A. Torralba, J. Sivic and W.T. Freeman, SIFT flow: dense correspondence across different scenes. ECCV, 2008, pp. 28-42.

[16]. K. Mikolajczyk and C. Schmid.A performance evaluation of local descriptors.IEEE Transactions on PAMI, 2005.

[17]. Wu S, Gao L. Multi-Level Joint Feature Learning for Person Re-Identification. Algorithms. 2020 May;13(5):111.

[18] An L, Chen X, Yang S. Multi-graph feature level fusion for person re-identification. Neurocomputing. 2017 Oct 11; 259:39-45.

[19] Wang G, Wang S, Chi W, Liu S, Fan D. A Person Reidentification Algorithm Based on Improved Siamese Network and Hard Sample. Mathematical Problems in Engineering. 2020 Jun 24;2020.

[20] L. Zhang, J. Chen, Y. Lu and P. Wang. Face recognition using scale invariant feature transform and support vector machine. The 9 th International Conference for Young Computer Scientist, pp. 1766-1770, 2008.

[21] Young-jun Song, Won-bae Park, Dong-woo Kim, and Jae-hyeongAhn, "Content-based image retrieval using new color histogram", Intelligent Signal Processing and Communication Systems, Proceedings of 2004 International Symposium on 18-19 Nov. 2004.

[22]. Sahoo S.K., Jain J., Prasanna S.R.M. (2011) Chain Code Histogram Based Facial Image Feature Extraction under Degraded Conditions. In: Abraham A., Mauri J.L., Buford J.F., Suzuki J., Thampi S.M. (eds) Advances in Computing and Communications. ACC 2011. Communications in Computer and Information Science, vol 192. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-22720-2_33

[23]. Ben-Ahmeida AM, Sasi AY. Improved image retrieval based on fuzzy colour feature vector. In International Conference on Graphic and Image Processing (ICGIP 2012) 2013 Mar 14 (Vol. 8768, p. 87686C). International Society for Optics and Photonics.

[24]. Venkatesan R, Chandakkar P, Li B, Li HK. Classification of diabetic retinopathy images using multi-class multiple-instance learning based on color correlogram features. In2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2012 Aug (pp. 1462-1465). IEEE.

[25]. Wang XY, Yu YJ, Yang HY. An effective image retrieval scheme using color, texture and shape features. Computer Standards & Interfaces. 2011 Jan 1;33(1):59-68.

[26]. El Aroussi M, El Hassouni M, Ghouzali S, Rziza M, Aboutajdine D. Novel face recognition approach based on steerable pyramid feature extraction. 2009, 16th IEEE International Conference on Image Processing (ICIP) 2009 Nov 7 (pp. 4165-4168). IEEE.

[27]. Ahmadian A, Mostafa A. An efficient texture classification algorithm using Gabor wavelet. InProceedings of the 25th annual international conference of the IEEE engineering in medicine and biology society (IEEE Cat. No. 03CH37439) 2003 Sep 17 (Vol. 1, pp. 930-933). IEEE.

[28]. Do MN, Vetterli M. The contourlet transform: an efficient directional multiresolution image representation. IEEE Transactions on image processing. 2005 Nov 14;14(12):2091-106.