International Journal of Electrical and Computer Engineering (IJECE)
Vol. 13, No. 2, April 2023, pp. 1499~1509
ISSN: 2088-8708, DOI: 10.11591/ijece.v13i2.pp1499-1509

A novel sketch based face recognition in unconstrained video for criminal investigation

Napa Lakshmi[1], Megha P. Arakeri[2]
[1]Department of Computer Science and Engineering, Presidency University, Bangalore, India
[2]Department of Information Science and Engineering, Ramaiah Institute of Technology, Bangalore, India

Article Info

Article history:

Received Mar 26, 2022
Revised Sep 14, 2022
Accepted Oct 10, 2022

Keywords:

Chokepoint dataset
Face recognition
Feature optimization
Video summarization
Video surveillance

ABSTRACT

Face recognition in video surveillance helps to identify an individual by comparing facial features of given photograph or sketch with a video for criminal investigations. Generally, face sketch is used by the police when suspect's photo is not available. Manual matching of facial sketch with suspect's image in a long video is tedious and time-consuming task. To overcome these drawbacks, this paper proposes an accurate face recognition technique to recognize a person based on his sketch in an unconstrained video surveillance. In the proposed method, surveillance video and sketch of suspect is taken as an input. Firstly, input video is converted into frames and summarized using the proposed quality indexed three step cross search algorithm. Next, faces are detected by proposed modified Viola-Jones algorithm. Then, necessary features are selected using the proposed salp-cat optimization algorithm. Finally, these features are fused with scale-invariant feature transform (SIFT) features and Euclidean distance is computed between feature vectors of sketch and each face in a video. Face from the video having lowest Euclidean distance with query sketch is considered as suspect's face. The proposed method's performance is analyzed on Chokepoint dataset and the system works efficiently with 89.02% of precision, 91.25% of recall and 90.13% of F-measure.

Corresponding Author:

Napa Lakshmi
Department of Computer Science and Engineering, Presidency University
Bangalore, India
Email: lakshmi.napa@gmail.com

1. INTRODUCTION

An automatic face recognition has shown successful results in a constrained environment but in unconstrained environment such as video surveillance, face detection and recognition is a complex task because of low resolution, partial occlusion, pose variation, and poor illumination [1]. Face recognition in video surveillance plays an important role in criminal investigations [2]. Before advancements in facial recognition in criminal investigations, police used to search for criminal in a video manually going through each frame. This process took a lot of time and the accuracy of recognizing a suspect depended on human expertise. If suspect is new and if his photo is not available, then police personnel collect information about suspect from eyewitnesses.

Based on this information, sketch is generated in two ways. Sketch which is drawn by forensic artists is called as forensic sketch and the sketch which is generated by software tools is called as composite sketch [3] as presented in Figure 1. This sketch is compared with the mug shot data available with the police and the closest match will be displayed. Comparing a sketch with the normal photo in a video frame is not an easy task as the modality of the sketch is different from that of a photo. The surveillance video is usually very lengthy and requires huge storage and high computational processing power. Recognizing a face in the video by going through each frame consumes more time. A video will have many redundant frames. These duplicate frames need to be removed without losing any important data.

Figure 1. Examples of forensic and composite sketches [4]

Video summarization techniques will reduce duplicate frames in a video and helps in recognizing a face in a summarized video in less time compared to the original video [5]. After summarization of the frames, the face is detected and recognized in the unconstrained video. New challenges such as low resolution, partial occlusion, pose variations, poor illumination is raised when face recognition applications progress from controlled scenarios to uncontrolled scenarios like video surveillance.

To overcome these drawbacks, this paper proposes sketch-based face recognition in a surveillance video in an unconstrained environment. In the proposed method, the surveillance video and sketch of a suspect are considered as input. Firstly, the input video is converted into frames and summarized using the proposed quality indexed three-step cross search algorithm. Next, the face is detected using the proposed modified Viola-Jones algorithm. Later, the features are extracted, and the necessary features are selected using the proposed salp-cat optimization algorithm. Finally, these features are fused with scale-invariant feature transform (SIFT) features, and Euclidean distance is computed between the feature vector of sketch and of each face in a video. Based on a similarity measure, the frames with the input image are displayed.

Block matching motion estimation techniques are widely used for video summarization because they are easy and simple to apply. In block matching algorithms (BMA), a video frame is divided into non-overlapping blocks. Then for each block, a motion vector is calculated by comparing block in the current frame with the matched block of the reference frame. Motion vector is the displacement of pixel positions from one frame to another frame. Full search BMA performs exhaustive search which checks all the candidate blocks (search points) that requires more computations. To minimize the computational effort of full search algorithm, other BMAs namely three-step search (TSS), cross-search, diamond search, adaptive rood pattern search (ARPS) [6] are developed.

Koga *et al.* [7] presented a TSS algorithm in which a 16x16 micro block is selected as the search window. From the center location (0, 0) eight search points are chosen. After finding the search point with a minimum sum of absolute difference, again eight search points are chosen around this new search location with a reduced step size. This repeats until the step size becomes 1. This method requires 25 search comparisons to find the correct match. The drawback of TSS is it considers a uniform searching pattern which is inefficient for small motions.

Li *et al.* [8] proposed a new TSS algorithm by considering the fact that motion vectors are highly center biased and it is used for small motion video sequences. This search algorithm requires 18 comparisons to find best matched block in the current frame with respect to the reference frame. Ghanbari [9] proposed cross-search motion estimation algorithm in which four search points are considered in cross form instead of eight search points that are used in three step search algorithm. Cross search requires 17 comparisons to find the best match and hence is faster compared to TSS and new TSS (NTSS).

Kamble *et al.* [10] has discussed a modified TSS algorithm for motion estimation that uses two cross search patterns and two hexagon search patterns. This method requires 13 search points in best case. The limitation of this technique is it requires 33 search points to identify the top match between consequent frames in the worst case. Zhu and Ma [11] presented a diamond search motion estimation algorithm which takes both large diamond search pattern (LDSP) and small diamond search pattern (SDSP). The LDSP performs a coarse search to identify a small area of optimal motion vector and SDSP performs a fine search to calculate the best motion vector in the located small area. This technique suffers over search or under search when there is a mismatch between pattern size and the amount of actual motion.

Mukherjee *et al.* [12] discussed the ARPS motion estimation technique in which search pattern is dynamically adapted to the magnitude of the target motion vector. It assumes that the motion vector of the current macro block is similar to the motion vectors of the macro blocks around the current macro block. Bhandari and Vyas [13] has developed three step cross search (TSCS) algorithm by combining three step search and cross search motion estimation algorithms. TSCS performed well compared to TCS and CS. In this algorithm, motion vector is calculated based on minimum sum of absolute difference (SAD) between current frame and matching block in the previous frame. The drawback of this algorithm is that it will not consider the quality of the frame.

After video summarization, appearance based [14] and geometric based [15] methods can be used to extract relevant features. Murala *et al.* [16] has presented local tetra pattern (LTrP) feature extraction method which is an extension of the local binary pattern (LBP) [17] and local ternary pattern (LTP) [18]. LBP and LTP compute gray level difference between referenced pixel and its surrounding pixels. LTP helps in extracting features that helps in face recognition in varying light conditions. LTrP finds more details than LBP.

Irrelevant and redundant features can be eliminated with the help of feature reduction algorithms [19] or feature optimization techniques [20]. Faris *et al.* [21] has discussed the salp swarm optimization algorithm for feature selection. The limitation of this algorithm is that it is inefficient on higher dimension problem and shows a poor convergence rate. Abualigah *et al.* [22] has given a comprehensive survey on variants of the salp swarm optimization algorithm and its applications. Ahmed *et al.* [23] has presented overview of cat swarm optimization algorithm and its variants. This algorithm is inspired by the resting and tracing behavior of cats. The drawback of this algorithm is the premature convergence problem and the static value of mixture ratio (MR) which indicates the number of cats to be chosen in seeking mode and tracing mode.

Wang *et al.* [24] discussed particle swarm optimization (PSO). It is reset with a group of solutions and then examines for the ideal solution using local best which is the fitness value of the solution and global best which is the best fitness value of the remaining solutions. The limitation of PSO is premature convergence on the local optimal solution. Mangla *et al.* [25] proposed a weighted component-based approach to recognize sketch to photo with accuracy of 88.23% on CHUK face sketch database. Dalal *et al.* [26] implemented feature-based sketch recognition using histogram of oriented gradients (HoG) features with 82.33% accuracy on CHUK face sketch database. Patil and Shibhangi [3] developed composite sketch face recognition using AdaBoost algorithm and multi-scale local binary patterns. It is observed from the reviewed research work that there is a need to improve the accuracy of video summarization and feature selection methods for effective sketch based face recognition in an unconstrained surveillance video. Therefore, next section explains the proposed methodology for accurate sketch based face recognition.

2. METHOD

In criminal investigations, to overcome the challenges associated with identifying a suspect using a sketch image in an unconstrained video, this paper proposes the following methodology as shown in Figure 2. Initially, the input surveillance video is converted into frames. Next, to avoid the repeated frames, the frames are summarized using the proposed quality indexed three-step-cross search (QITC) method. Then, the face images are segmented from the summarized frames using the proposed modified viola jones algorithm (MVJA).

After that, features such as LTrP, Gabor Wavelet, HoG, and Shape features (solidity, eccentricity, and circularity) are extracted from each facial image to get good representation of the facial images. Then, the necessary features are selected using the proposed salp cat swarm optimization (SCO) algorithm. Also, the facial points are derived from the segmented face with the help of SIFT. Later, the extracted features and the facial points are fused in a single vector. All the above phases, such as segmentation, features extraction, features selection, and facial point extraction are also done for the sketch query image. Next, the similarity is measured between the feature vector of the sketch image and the feature vector of each face in the video frames using the Euclidean distance metric. Based on the minimum distance, the frames containing the matched input image are displayed.

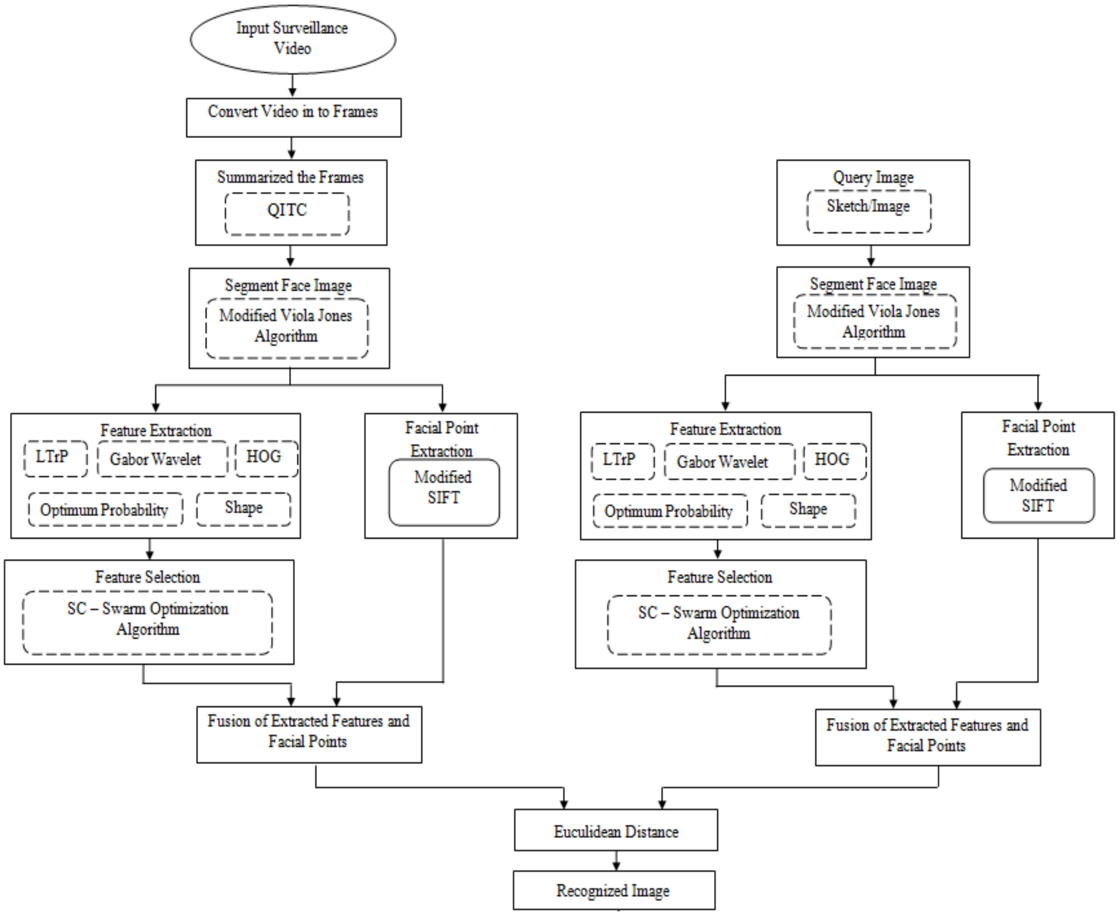

Figure 2. Methodology for face recognition using sketch

2.1. Video summarization

The notion of video summarization is to create a summary of a longer video by selecting the most informative and quality frames for the users. To compress video, temporal redundancy between frames should be reduced by using motion estimation. This paper proposes an effective video summarization technique called QITC algorithm. In the existing TSCS algorithm, the input video frame is divided into macro blocks of size M x N pixels. Each macro block is assigned with a motion vector. The motion vector is the transposition between the collocated block and the best match block as depicted in Figure 3.

The best-matched block is searched based on the minimum mean absolute difference (MAD) between the target block of the current frame with the candidate blocks of the reference frame. This kind of searching may give poor quality frames because the motion vector is calculated based on minimum MAD value without considering the quality of the macro block. But the higher quality frame is needed for recognizing the face because if the quality of the frame is poor then it gives a higher recognition error.

Therefore, to overcome this drawback of three-step cross search algorithm and to find a motion vector based on the quality of the macro block, the QITC algorithm is proposed. In QITC, the position of the best matched macro block is searched based on the quality index of the frame in a video. The quality index value is the combination of brightness, sharpness and illumination. Initially, the quality index threshold value is taken as 0.5; if the quality index value of searched location matches with the threshold value, then the location is selected; otherwise, neglect the location and again search the location with better quality index.

Consider reference frame $x = \{x_i \mid i=1, 2, 3..., N\}$ and current frame $y= \{y_i \mid i=1, 2, 3..., N\}$. Quality index of the current frame with respect to the reference frame is calculated using the (1) [27]:

$$Q=\frac{\sigma_{xy}}{\sigma_x\sigma_y} \times \frac{2\ \bar{x}\ \ \bar{y}}{\bar{x}^2+\bar{y}^2} \times \frac{2\sigma_x\sigma_y}{\sigma_x{}^2\sigma_y{}^2} \tag{1}$$

where $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i\, , \bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})^2, \sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i-\bar{y})^2$$
$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}(x_i-\bar{x})(y_i-\bar{y})$$

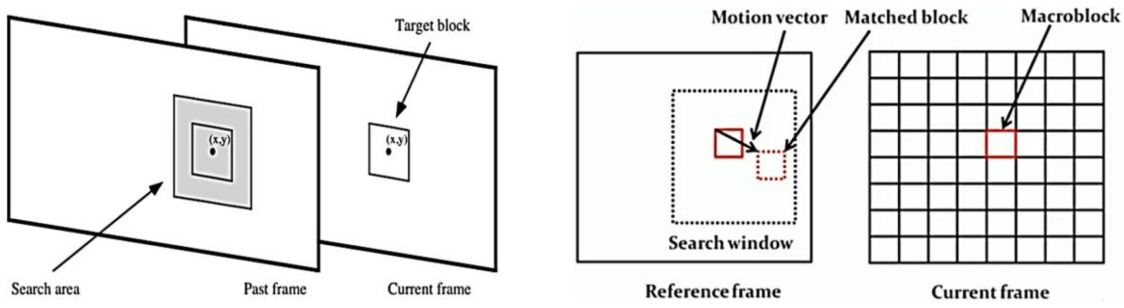

Figure 3. Search area is the region around the collocated block in the past frame

This quality index is a based on loss of correlation, change in contrast and change in luminance. The correlation coefficient lies in the range [-1, 1]. Luminance distortion is average luminance between current frame and the reference frame whose value lies in the range of [0, 1]. Contrast distortion finds similarity in the contrasts of the images and σ_x and σ_y represent the contrast of x and y. Its range lies in [0, 1]. The value of the quality index Q is in the range of [-1, 1] and it will be 1 when the images are identical. The following proposed algorithm extracts the quality frames from a surveillance video.

Algorithm: Quality indexed three step cross search summarization

```
Input: video with 'n' number of frames
Output: motion vector of each frame
Steps:
 1. Begin
 2. Given a video sequence with frames f1,f2,…fn
 3. for i=1 to n
 4. Read frame i and divide it into 16x16 size macro blocks
 5. Calculate the motion vector for every macro block. The block for which motion vector is
    to be found is called as target block.
 6. Find the collocated block in the reference frame corresponding to the target block of
    the current frame.
 7. Search point is the pixel at top left corner of the macro block. Center point is the
    search point of the collocated block. Start with search location at the center point.
 8. Assign search parameter p to 7.
 9. Set the step size as half of the search parameter p. So S=4.
 10. Search for four locations at a distance of +/- S pixels from the center.
 11. Calculate quality index Q for each candidate block at these four locations in the
     reference frame using (1) and choose a location with the maximum Q from four search
     locations and mark this as a new search origin.
 12. Set S=S/2 and pick another three locations around its new origin
 13. Choose the location with the maximum quality index value and make this new origin.
 14. Repeat search procedure until S=1.
 15. If S=1, explore eight new locations around the new origin and choose a location with
     maximum Q value. The motion vector is calculated as the displacement of the target
     block from the best matched block.
 16. End for
 17. End
```

2.2. Facial image segmentation

From the summarized video frames, the face images are identified using the proposed Modified Viola-Jones algorithm. The benefit of this algorithm is that detection is fast but training is slow. Viola-Jones algorithm uses Haar feature filters, so it does not use multiplications. The algorithm looks for specific Haar features of a face and if these features are found, pass the candidate to the next stage [28]. But the limitation of the Haar feature is that it is illumination-variant. Therefore, face detection performance under variable lighting is significantly degraded. To overcome this drawback, contrast based Haar (CHaar) [29] feature descriptor is included in the proposed modified Viola-Jones algorithm instead of Haar features. CHarr computes the contrast of the feature region. When the illumination changes to very high or low, the intensity variance between rectangular regions within a feature may reduce. Therefore, to obtain illumination-invariant features, the normalization factor is formulated and is defined as (2):

$$N_f = th - \left|A_f - th\right| \tag{2}$$

where, A_f denotes the average intensity of a feature region, and *th* denotes the threshold which is set to median intensity value, 128. The CHarr is assigned in the range of $0 < A_f < 2th$ and it yields 0 when $A_f = 0$ or $A_f >= 2th$ as shown in (3). Integral image is used for computing the CHaar features and AdaBoost classifier is used to choose a lesser number of features and train the classifier. By this method, the face can be detected in less time.

$$CHarr = \begin{cases} \frac{1}{N_f} . \sum_{i=1}^{R} sign(i). w_i. A_i, & \text{if } 0 < A_f < 2th \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

2.3. Feature extraction

From the segmented face image, the features such as LTrP, Gabor Wavelet, HoG, solidity, eccentricity, and circularity are extracted. LTrP are efficient texture descriptors and robust to illumination changes. Gabor Wavelets are used to find corners, edges and blobs. It is a linear filter for texture analysis. HoG descriptors can effectively represent the appearance of a face and also robust to partial occlusion [30]. In the proposed method, the shape features such as eccentricity, circularity ratio, and solidity [31] are used to describe the shape of the face. In addition to these features, SIFT [32] is used for detecting 128 feature points on face that are not variant to rotation, scale and illumination and can identify object even in partial occlusion and image clutter.

2.4. Feature selection

After feature extraction, the necessary features are selected using the proposed salp-cat swarm algorithm. It combines the salp swarm algorithm (SSA) [21] with the cat swarm algorithm (CSA). SC-Swarm is proposed to avoid low convergence precision [33]. Here, the weight factor is computed first which automatically changes as the number of iterations changes, so that the population is capable of adapting well to the current search environment. SSA algorithm is revised by improving updating phase of the population's position. This modification combines update mechanism of the CSA with the computed weight factor into the SSA which helps it to reach optimal value quickly. Proposed SC-Swarm algorithm is explained as follows.

Algorithm: Salp cat swarm optimization
Input: Feature vector
Output: Reduced feature vector with relevant features
Steps*:*
1). Initialize the population size *N* of swarms

$$S_i = \{S_1, S_2, \ldots . S_N\} \tag{4}$$

2). Initialize the location of food source *F* chased by the salp chain which is global optimum as (5).

$$F_j = \{F_1, F_2, \ldots . F_N \ldots, F_D\} \tag{5}$$

3). Then, find the fitness of all salps and select the salp with the maximum fitness as the global optimal position of the population.
4). Update the location of the leader using the (6).

$$S_1^j = \begin{cases} F_j + r_1 * \left(\left(u_{bj} - l_{bj}\right) * r_2 + l_{bj}\right) & \text{if } r_3 \geq 0.5 \\ F_j - r_1 * \left(\left(u_{bj} - l_{bj}\right) * r_2 + l_{bj}\right) & \text{if } r_3 < 0.5 \end{cases} \tag{6}$$

$$r_1 = 2e^{\left(4I/P\right)} \tag{7}$$

Here, S_1^j is the position of leader in the j^{th} dimension and $i \geq 2$. F_j denotes the position of F in j^{th} dimension. The parameters r_2 and r_3 are random numbers, which are in the range of [0, 1]. u_{bj} and l_{bj} denote the upper and lower bound of search space in the j^{th} dimension, I is the current iteration. P is the presupposed maximum number of iterations and r_1 denotes the updated expression of coefficient.
5). Update the position of the follower using CSA by computing the weight factor.

$$S_1^j = g(I) + S_1^j \quad (8)$$

$$g(I) = g\left(^I/_P\right)(gmin_{max})_{max} \quad (9)$$

Where S_i^j denotes the position of the followers in j^{th} dimension, g_{max} and g_{min} represents the upper and lower limit of the weight factor *g*, respectively and *rd* is the random number. The dynamic weight factor *g(I)* varies according to the number of iterations as shown in (9).

6). Update the global optimum position by comparing it with the fitness of each individual.

7). Stop the calculation when the results meet the end condition at maximum number of iterations and output the optimization result. Else, return to step 4 and repeat.

The fitness value of the proposed SCO is compared with the existing salp swarm optimization (SSO) and PSO as shown in Figure 4. It varies depending upon the number of iterations. For example, for 100 iterations, the proposed system has 2.1 fitness value, but the existing SSO and PSO have a fitness value of 3.1 and 4.3, which is greater compared to the proposed system.

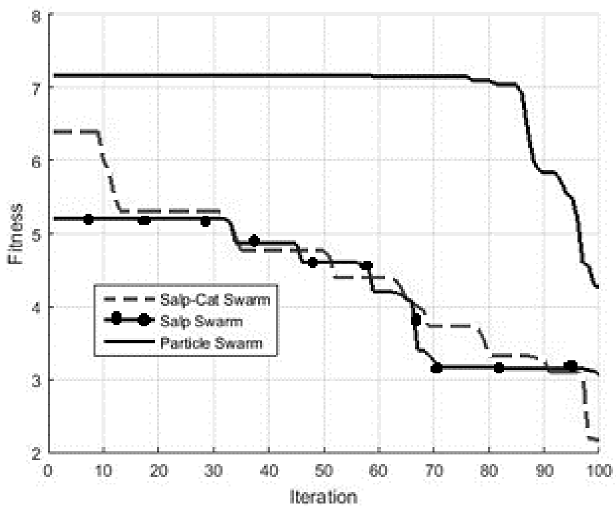

Figure 4. Fitness value of the proposed QITC-SCO with the existing methods

2.5. Euclidean distance computation

After the facial feature points are extracted, the fusion operation combines the extracted features (LTrP, Gabor wavelet, HoG, optimum probability, and shape features) and SIFT features. Then these feature extraction, feature selection, facial point extraction using SIFT, and feature fusion steps are also performed for the query image (sketch image). The Euclidean distance (ED) [34] is computed between both the feature vectors. Let us consider the fusion features of the input image as vector *p* and fusion features of the query image as vector *q* and the Euclidean distance between *p* and *q* is computed as (10).

$$ED = \sqrt{\sum_{i=1}^{n}(p_i - q_i)^2} \quad (10)$$

3. RESULTS AND DISCUSSION

The results of the proposed face recognition system which is implemented using MATLAB is demonstrated in this section. The analysis of the performance is done for the proposed QITC based SCO for face recognition (QITC-SCO) with the existing techniques. Firstly, the summarization results are shown and secondly, face detection results and later face recognition results are discussed. The results are shown on chokepoint video dataset.

3.1. Database description

The proposed system utilizes chokepoint dataset that contains “48 video sequences and 64,204 face images. The frame rate is 30 fps and the image resolution is 800 X 600 pixels” [35]. In the experiments, the

P2E_S5 video sequence of chokepoint dataset is used which represents surveillance video with the crowd. The sample frames of unconstrained video of chokepoint dataset are shown in Figure 5.

Figure 5. Sample frames of chokepoint dataset

3.2. Performance analysis

The performance of the proposed summarization method, QITC is demonstrated and compared with the existing techniques on the chokepoint dataset. The P2E_S5 video sequence chokepoint dataset has total 806 frames. QITC selected quality frames based on the quality index of the image and reduced the total number of frames to a minimum as compared to existing algorithms as displayed in Table 1. From the table it is obvious that the proposed technique has higher performance compared to existing techniques.

After summarization, face segmentation is carried out using the proposed MVJ algorithm which uses contrast Haar features instead of simple Haar features. Face segmentation is as shown in Figure 6. Later the features are extracted using LTrP, Gabor wavelet, HoG, shape features. From these features, relevant features are selected using proposed SCO technique.

Table 1. Summarization of frames

Technique	Number of summarized frames
Proposed QITC	258
TSCS [13]	324
Diamond Search (DS) [27]	543
ARPS [12]	601

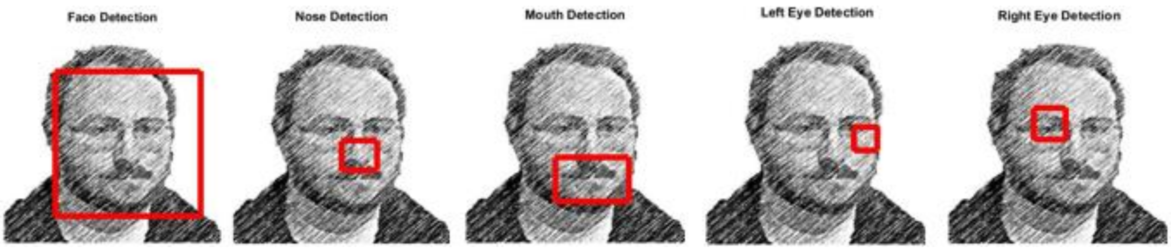

Figure 6. Result of face segmentation

Table 2 indicates that the proposed SCO selects fewer number of features compared to existing optimization techniques SSO and PSO and helps in dimensionality reduction of the feature vector. Selected features are fused with 128 facial points extracted using SIFT and these features are used in face recognition. The performance of face recognition technique is measured using “precision, recall, and F-measure” [36] which are calculated based on confusion matrix as demonstrated in Table 3.

Table 2. Comparison of feature optimization techniques

Technique	Proposed SCO	SSO	PSO
Number of features	19	22	24

Table 3. Confusion matrix

Total number of frames, N=258	Actual	
Predicted	P	N
P	TP=73	FN=7
N	FP=9	TN=169

P represents “number of frames that contains input face” and N represents “number of frames that do not contain input face”. Precision is “the ratio of correctly predicted positive observations to the total predicted positive observations”. It is represented as (11).

$$precision = {}^{TP}/_{TP+FP} \tag{11}$$

The recall is “the ratio of correctly predicted positive observations to all observations in the actual class”. It is evaluated as (12).

$$Re\,c\,all = {}^{TP}/_{TP+FN} \tag{12}$$

F-measure considers both “the precision and the recall of the test to compute the score”. It is evaluated as (13).

$$F - measure = {}^{(2 * Pr\,e\,cision * Re\,c\,all)}/_{Pr\,e\,cision + Re\,c\,all} \tag{13}$$

The performance of the proposed SCO is compared with existing methods in terms of precision, recall, and F-measure as shown in Figure 7. It is clear from the results that proposed SCO performance is better compared to the existing techniques with precision of 89.02%, recall of 91.25% and F-measure of 90.13%.

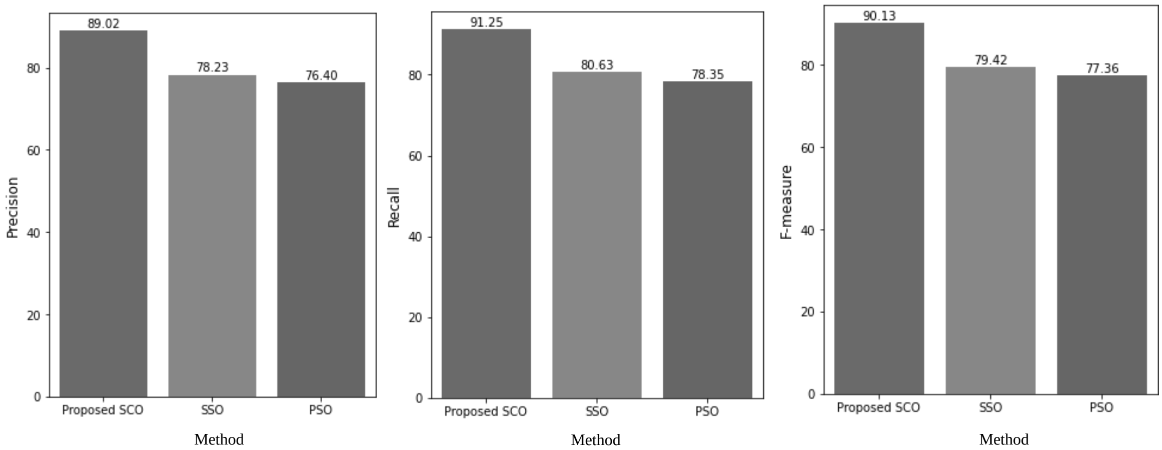

Figure 7. Performance measures of proposed methodology

4. CONCLUSION

This paper proposed a technique to recognize faces in video surveillance based on the sketch. QITC algorithm was proposed which reduced redundant frames and extracted quality frames from the unconstrained video. The face is segmented from the video frames using the proposed MVJ algorithm. Then, various features such as LTrP, Gabor wavelet, HoG, SIFT features, and Shape features (solidity, eccentricity, and circularity) are extracted from the faces and the relevant features are selected using proposed SCO optimization algorithm. It is observed in the experiments that the suggested methods perform superior to existing algorithms for sketch-based face recognition in surveillance video on Chokepoint dataset. The proposed method accomplished 89.02% of precision, 91.25% of recall, and 90.13% of F-measure. Hence the proposed method helps police in the criminal investigation.

REFERENCES

[1] R. V. Virgil Petrescu, "Face recognition as a biometric application," *Journal of Mechatronics and Robotics*, vol. 3, no. 1, pp. 237–257, Jan. 2019, doi: 10.3844/jmrsp.2019.237.257.

[2] N. Lakshmi and M. P. Arakeri, "Face recognition in surveillance video for criminal investigations: A review," in *CNC 2018: Communication, Networks and Computing*, 2019, pp. 351–364.

[3] S. Patil and D. C. Shibhangi, "Composite sketch based face recognition using ANN classification," *International Journal of Scientific & Technology Research*, vol. 9, no. 1, pp. 42–50, 2020.

[4] K. Ounachad, M. Oualla, A. Souhar, and A. Sadiq, "Face sketch recognition-an overview," in *Proceedings of the 3rd International Conference on Networking, Information Systems & Security*, Mar. 2020, pp. 1–8, doi: 10.1145/3386723.3387857.

[5] H. B. U. Haq, M. Asif, and M. Bin Ahmad, "Video summarization techniques: a review," *International Journal of Scientific and Technology Research*, vol. 9, no. 11, pp. 146–153, 2020.

[6] R. Yaakob, A. Aryanfar, A. A. Halin, and N. Sulaiman, "A comparison of different block matching algorithms for motion estimation," *Procedia Technology*, vol. 11, pp. 199–205, 2013, doi: 10.1016/j.protcy.2013.12.181.

[7] T. Koga *et al.*, "Motion compensated inter frame coding for video conferencing," in *National Telecommunication Conference, New Orleans*, 1981, p. G5.3.1– G5.3.1-5.

[8] R. Li, B. Zeng, and M. L. Liou, "A new three-step search algorithm for block motion estimation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 4, no. 4, pp. 438–442, 1994, doi: 10.1109/76.313138.

[9] M. Ghanbari, "The cross-search algorithm for motion estimation (image coding)," *IEEE Transactions on Communications*, vol. 38, no. 7, pp. 950–953, Jul. 1990, doi: 10.1109/26.57512.

[10] S. Kamble, N. Thakur, and P. Bajaj, "Modified three-step search block matching motion estimation and weighted finite automata based fractal video compression," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, no. 4, art. no. 27, Jul. 2017, doi: 10.9781/ijimai.2017.444.

[11] S. Zhu and K.-K. Ma, "A new diamond search algorithm for fast block-matching motion estimation," *IEEE Transactions on Image Processing*, vol. 9, no. 2, pp. 287–290, 2000, doi: 10.1109/83.821744.

[12] R. Mukherjee, B. Biswas, I. Chakrabarti, P. K. Dutta, and A. K. Ray, "Efficient VLSI design of adaptive rood pattern search algorithm for motion estimation of high definition videos," *Microprocessors and Microsystems*, vol. 45, pp. 105–114, Aug. 2016, doi: 10.1016/j.micpro.2016.04.003.

[13] R. Bhandari and A. Vyas, "New technique of three step search algorithm used for motion estimation in video compression," *International Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 5, pp. 760–763, 2016.

[14] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7803–7821, Mar. 2017, doi: 10.1007/s11042-016-3418-y.

[15] S. Karve, V. Shende, and R. Ahmed, "A comparative analysis of feature extraction techniques for face recognition," in *2018 International Conference on Communication information and Computing Technology (ICCICT)*, Feb. 2018, pp. 1–6, doi: 10.1109/ICCICT.2018.8325884.

[16] S. Murala, R. P. Maheshwari, and R. Balasubramanian, "Local tetra patterns: A new feature descriptor for content-based image retrieval," *IEEE Transactions on Image Processing*, vol. 21, no. 5, pp. 2874–2886, May 2012, doi: 10.1109/TIP.2012.2188809.

[17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, Jul. 2002, doi: 10.1109/TPAMI.2002.1017623.

[18] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, Jun. 2010, doi: 10.1109/TIP.2010.2042645.

[19] A. Abbad, K. Abbad, and H. Tairi, "An efficient post-processing approach for dimensionality reduction methods for face recognition," in *2017 International Conference on Advanced Technologies for Signal and Image Processing (ATSIP)*, May 2017, pp. 1–7, doi: 10.1109/ATSIP.2017.8075609.

[20] J. K. Appati, H. Abu, E. Owusu, and K. Darkwah, "Analysis and implementation of optimization techniques for facial recognition," *Applied Computational Intelligence and Soft Computing*, vol. 2021, pp. 1–13, Mar. 2021, doi: 10.1155/2021/6672578.

[21] H. Faris *et al.*, "An efficient binary salp swarm algorithm with crossover scheme for feature selection problems," *Knowledge-Based Systems*, vol. 154, pp. 43–67, Aug. 2018, doi: 10.1016/j.knosys.2018.05.009.

[22] L. Abualigah, M. Shehab, M. Alshinwan, and H. Alabool, "Salp swarm algorithm: a comprehensive survey," *Neural Computing and Applications*, vol. 32, no. 15, pp. 11195–11215, Aug. 2020, doi: 10.1007/s00521-019-04629-4.

[23] A. M. Ahmed, T. A. Rashid, and S. A. M. Saeed, "Cat swarm optimization algorithm: A survey and performance evaluation," *Computational Intelligence and Neuroscience*, vol. 2020, pp. 1–20, Jan. 2020, doi: 10.1155/2020/4854895.

[24] D. Wang, D. Tan, and L. Liu, "Particle swarm optimization algorithm: an overview," *Soft Computing*, vol. 22, no. 2, pp. 387–408, Jan. 2018, doi: 10.1007/s00500-016-2474-6.

[25] F. U. Mangla, M. Shahzad, M. I. Lali, and S. A. C. Bukhari, "Sketch-based facial recognition: a weighted component-based approach (WCBA)," *Multimedia Tools and Applications*, vol. 79, no. 37–38, pp. 27533–27553, Oct. 2020, doi: 10.1007/s11042-020-09246-1.

[26] S. Dalal, V. P. Vishwakarma, and S. Kumar, "Feature-based sketch-photo matching for face recognition," *Procedia Computer Science*, vol. 167, pp. 562–570, 2020, doi: 10.1016/j.procs.2020.03.318.

[27] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81–84, Mar. 2002, doi: 10.1109/97.995823.

[28] Y.-Q. Wang, "An analysis of the Viola-Jones face detection algorithm," *Image Processing on Line*, vol. 4, pp. 128–148, Jun. 2014, doi: 10.5201/ipol.2014.104.

[29] K.-Y. Park and S.-Y. Hwang, "An improved Haar-like feature for efficient object detection," *Pattern Recognition Letters*, vol. 42, pp. 148–153, Jun. 2014, doi: 10.1016/j.patrec.2014.02.015.

[30] A. Y. Boumedine, S. Bentaieb, A. Ouamri, and A. Mallek, "3D face identification using HOG features and collaborative representation," in *Advances in Computing Systems and Applications*, 2021, pp. 3–13.

[31] J. Liu and Y. Shi, "Image feature extraction method based on shape characteristics and its application in medical image analysis," in *Applied Informatics and Communication*, 2011, pp. 172–178.

[32] L. Lenc and P. Král, "Automatic face recognition system based on the SIFT features," *Computers & Electrical Engineering*, vol. 46, pp. 256–272, Aug. 2015, doi: 10.1016/j.compeleceng.2015.01.014.

[33] J. Wu, R. Nan, and L. Chen, "Improved salp swarm algorithm based on weight factor and adaptive mutation," *Journal of Experimental & Theoretical Artificial Intelligence*, vol. 31, no. 3, pp. 493–515, May 2019, doi: 10.1080/0952813X.2019.1572659.
[34] R. Ahdid, K. Taifi, S. Said, and B. Manaut, "Euclidean and geodesic distance between a facial feature points in two-dimensional face recognition system," *International Arab Journal of Information Technology*, vol. 14, no. 4A, pp. 565–571, 2017.
[35] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," in *CVPR 2011 WORKSHOPS*, Jun. 2011, pp. 74–81, doi: 10.1109/CVPRW.2011.5981881.
[36] W. Dhifli and A. B. Diallo, "Face recognition in the wild," *Procedia Computer Science*, vol. 96, pp. 1571–1580, 2016, doi: 10.1016/j.procs.2016.08.204.

BIOGRAPHIES OF AUTHORS

Napa Lakshmi is currently working as Assistant Professor in the department of CSE in Presidency University, Bangalore. She has 9 years of teaching experience and her areas of interest include Image processing, computer vision, machine learning and deep learning. She can be contacted at email: lakshmi.napa@gmail.com.

Megha P. Arakeri received Ph.D from NITK, Surathkal during the year 2013. She is currently working as Associate Professor in the Information Science and Engineering department of Ramaiah Institute of Technology, Bangalore. She is having total 16 years of teaching and research experience. Her areas of interests are image processing, pattern recognition, machine learning, human computer interaction and big data analytics. She can be contacted at email: meghaparakeri@msrit.edu.