# Real-Time Facial Recognition Using SURF-FAST

**Chapter** · January 2022

DOI: 10.1007/978-981-16-2937-2_32



Showmik Setta


Shreyashee Sinha


Prasenjit Choudhury
National Institute of Technology, Durgapur

# Real-Time Facial Recognition Using SURF-FAST

**Showmik Setta, Shreyashee Sinha, Monalisa Mishra, and Prasenjit Choudhury**

**Abstract** Facial recognition is a biometric technology that uses unique facial features to identify a person. Preventing retail crime, smart advertising, unlocking phone, finding missing person, diagnosis of disease, tracking school attendance are some of the many utilizations of face recognition. Recently, novel work has been done to detect facial emotion to understand the state of mind of a learner in an online class which better helps the instructor to design courses. Since there is no contact required like it is required for fingerprints, facial recognition offers a quick, automatic, and seamless verification experience. Most of the face detection techniques use deep learning and neural networks which require huge datasets for training. Using large datasets and matching one by one is a time taking approach. On the other hand, face identification using local feature extraction provides less accuracy than what is expected. Hence, this directs us to handle huge database, focus on developing new real time algorithms with maximum accuracy. The paper proposes a dynamic real time automated face recognition system to identify face in live video using surveillance camera and label a known person by their name and unknown person as "Unknown". In the experiment, stacked-image montage is used to reduce the time for training the images. Multi-feature extraction using Fast Feature, Speeded-Up-Robust Features (SURF) and Sum of Absolute Difference (SAD) have proved beneficial in getting high accuracy using very small dataset for face recognition.

S. Setta (✉)
Department of MCA, Techno India Hooghly, Chinsurah, West Bengal, India

S. Sinha · P. Choudhury
Department of CSE, NIT Durgapur, Durgapur, West Bengal, India
e-mail: prasenjit.chowdhury@cse.nitdgp.ac.in

M. Mishra
CV Raman Global University, Bhubaneswar, Odisha, India

 
N. Sharma et al. (eds.), *Data Management, Analytics and Innovation*,
Lecture Notes on Data Engineering and Communications Technologies 71,
https://doi.org/10.1007/978-981-16-2937-2_32

## 1 Introduction

Nowadays, computer vision is rapidly developing in the field of research and its applications for its real-world problem-solving capability over the last several decades. There are various types of sub-domain computer vision like face recognition [1, 2], tracking in video [3, 4], object recognition [5, 6], 3D pose estimation [7, 8], motion estimation [9], event detection [10, 11], image restoration [12], etc. Computer vision is concerned with the analysis, automatic extraction and understanding of useful information and separates them from an individual image or a series of images.

**Face recognition**: Face recognition is a biometric technology which helps a system to take decision by its own experience which is basically learning from previous dataset and identifying or verifying a person from image and video frames. Face recognition attracted popularity because of recent success of convolutional neural networks (CNN) [13] along with deep learning (DL). DeepFace [14] technique showed prominent results on Labeled Faces in the Wild (LFW) dataset trained on convolutional neural networks (CNN) with more than 4M training images. The Labeled Faces in the Wild (LFW) accuracy increases when FaceNet [15] renders a network based on triplet loss using 200M images. This is a proof of effective performance improvement when using large amount of training data.

There are various types of research that has been explored in face recognition field. In video-based technology facial recognition technique had been least developed by the research community [16–19]. Recently, deep learning-based methods [14, 15, 20–22] shows an impressive performance with respect to recognition of still images in the wild, like those in the Labeled Faces in the Wild (LFW) dataset [23]. In video analytics one of the most challenging problems is a low-quality video which was resolved [24] by high-quality cameras, which may not always be accessible or affordable. However, deep learning techniques require very large amount of data in order to perform better than regular methods and it is extremely expensive to train due to complex data models. Moreover, deep learning requires expensive GPUs and hundreds of machines which increases cost to the users.

This paper proposes a technique which can solve the problem for limited hardware support with quick response and successfully recognizes faces from existing dataset. To reduce the time for training images montages are used, and multi-feature extraction for matching and classification is carried out.

The proposed work touches very high accuracies in very small datasets for its features. It is important to extract face features which are robust to appearance variations. However, feature extraction is one of the most difficult tasks when taking a frame from real-time video. If one is able to extract good features then it helps to match technique from a different dataset or even new dataset.

Currently, there are many feature extraction algorithms and techniques that have been proposed for face recognition. However, using local features for small datasets with high accuracy is the best policy for achieving this goal. FAST feature detection, Speeded Up Robust Features (SURF) and sum of absolute difference (SAD) give excellent results. This multi-feature extraction gives such accuracy because if any of them fail then it can factor in the other feature extraction techniques. However, making feature extractions error free is one of the challenging problems. Here montage is used to reduce time complexity to allow a quick matching then it inputs frames from real time and matches it with a montage (training ImageSets) and returns to live frame with matched label names.

## 2 Related Work

The benefit of facial recognition is that it is not intrusive. It can be done without the user's conscience of being scanned. Facial recognition is different from other biometric techniques in different ways it can be used to search for suspected terrorists, missing children and wanted criminals. It also has an edge over other biometric techniques as it can be done from a long distance without any contact with the subject.

Although there is a range of interesting research in face recognition there is limited work done to reduce time and complexity for training data and matching quickly with multiple local features. In face recognition at first, face is to be detected for identification. One of the well-known techniques for this is based on AdaBoost along with a novel algorithm proposed by Viola et al. which can identify key interest points and also can detect face frame from image [25]. Kavitha et al. uses video editing and summarization on video applied on human facial objects using SURF (Speeded Up Robust Feature) [26]. Although this paper used SURF, it is desirable to use multiple features for feature extraction. Secondly, matching images one by one is complex and takes time. Hence, this paper proposes stack image montages technique for training set. This takes short time for matching and classification and hence opens up a huge scope for future work. Further, Shoelson proposed a similar type of approach but there is no work done to identify an unknown person using this technique [27]. Hence, this paper further proposes a solution that can identify an unknown person. Lowe et al. proposed Scale Invariant Feature Transform (SIFT) [28] which have been extensively used in face recognition and object detection. Nonetheless, despite the diversity of work in face recognition to utilize SIFT features; this methodology still can't match the speed requirement enough for live video applications. On the other hand, Bay et al. proposed a new algorithm SURF [29] which is an in-plane and scale rotation-invariant descriptor and detector which is comparable or even enhances accuracy measures than the SIFT method. Like SIFT detector, in SURF, detector is first appointed in an image to search the interest points and then descriptors are used for feature extraction of each interest point. To locate the integral image of the interest points SIFT uses "Difference of Gaussians (DoG) filter" [30], whereas SURF uses Hessian matrix approximation operating, which decreases the *computational timing* drastically [31]. FAST is an algorithm proposed originally by Rosten et al. and Features from Accelerated Segment Test (FAST) by Viswanathan, 2011 [32] which is appropriate for video face recognition as it is robust and does fast corner detection and extraction.

# 3 Methodology

Basic steps for Image processing and Video Processing are described in the diagram below (Fig. 1). However, in this experiment, some changes have been done because

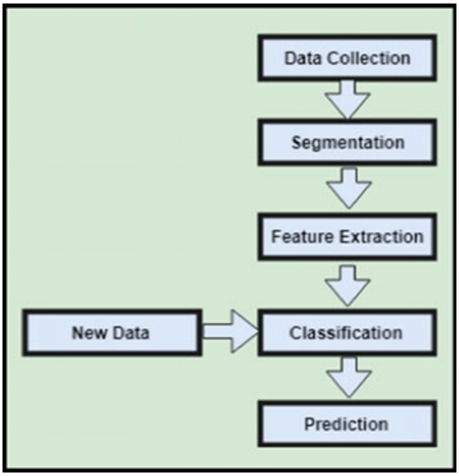


**Fig. 1** Basic block diagram for face recognition

this approach doesn't need to train huge amount of data for a small, restricted area.

## *3.1 Pre-Processing*

The term pre-processing means to convert raw data to highly understandable input data in machine learning and create a dataset for training. Here are some steps for the same:

- Start video streaming and detect face from the video using face detector. This helps to take pictures from the video of only interested faces.
- In a loop of 25 times capture a snapshot of a face frame.
- Convert it to grayscale as it helps to extract local features easily.
- Validate and crop face images from video as it is important for face matching and classification.
- Resize every face image to same size.

**Fig. 2** Example of montages

## 3.2 Training

Training is teaching, which gives the computer the knowledge to predict the results from previous experiences.

1. *ImageSets Creation, Feature detection and extraction*

(a) *Montage*
A montage is a kind of collage (Fig. 2), often assembled from different images resulting in a single overall image or in other word sit can be said that single image object which contains all the image frames. This technique uses montage to compute features for a training set in a single pass and finds the most informational of those features across the entire training set. The single-pass aggregation of features is what enables this to be fast enough to do on-the-fly training. And it helps to reduce the complexity of the application. In most of face recognition systems, matching a single person's face, one by one is time taking and is complex. Creating a montage image and matching this can be used to reduce time and make it less complex. So this paper is creating and using montage as training Image. Example of montage image is given below:

(b) *Viola-Jones Face Detection Method*
Viola-Jones face detection method is very common and accurate because of its robustness, real time (minimum of 2 frames/second) and facial detection—The target is to separate human faces from non-face areas. This face detection technique was proposed by Viola and Jones in 2001 [25]. This framework is widely used in face recognition. The four stages of this algorithm are:

(i) *Haar Feature Selection*
Every human face has some common properties. Eyes region is darker than the upper cheeks and nose bridge region is visually glossy than the eyes. The congregations of properties form equivalence face features such as size and location of eyes, edge of nose, mouth and value of oriented gradients of pixel intensities. This detection is represented by rectangular frames. By this algorithm, the four features that match are then sought in the image of a face [33].
Rectangle features:

$$\boldsymbol{value} = \sum(\boldsymbol{Pixels\,in\,black\,area}) - \sum(\boldsymbol{Pixels\,in\,white\,area})$$

Viola-Jones used two-rectangle features (Figs. 3, 4 and 5). Each individual feature is associated with a specific region in the sub-window.

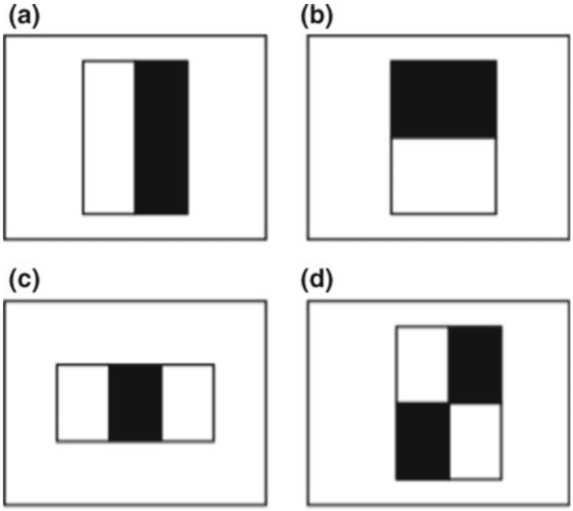


**Fig. 3** Example rectangle features shown relative to the enclosing detection window. The total number of pixels lies within the white rectangles is subtracted from the total number of pixels in the gray rectangles. Two-rectangle features are pointed out in **a**, **b**, **c**

**Fig. 4** Haar feature that looks comparable to the edge of the nose is applied on the face

**Fig. 5** Haar feature that looks comparable to the eye region which is obscure than the upper cheeks is applied on a face

(ii) *Creating an Integral Image*
Summed-area table is also known as integral image which is used by Viola and Jones. Rectangular features are analyzed in fixed time. Any rectangular sum can be computed in four array references using an integral image. The difference between two rectangular sums can be

computed in eight references. Since the two-rectangle features defined above involve adjacent rectangular sums they can be computed in six array references, eight in the case of the three-rectangle features, and nine for four rectangle features.

(iii) *Adaptive Boosting Training*
Freund et al. first introduced the Adaptive boost also known as AdaBoost is a machine learning meta-algorithm [34]. Varieties of AdaBoost are used for selecting both a minimal set of characteristics and training the classifier. In its initial form, Adaptive Boosting learning algorithm or AdaBoost is utilized in concurrency with various types of learning algorithms (weak learning algorithms) to better its performance of classification. The weak learning algorithm is developed to select the individual rectangular characteristics that can separate the negatives along with positives. For each feature or characteristics, the weak learner decides the optimal threshold of the classification function so that the least number of samples are misclassified. A weak classifier $h_j(x)$ thus consists of a feature $f_j$, a threshold $\theta_j$ and a parity $p_j$ indicating the direction of the inequality sign:

$$h_j(x) = \begin{cases} 1 \text{ if } p_j f_j(x) < p_j\theta_j \\ 0 \qquad \text{otherwise} \end{cases}$$

*Here x is sub-window of an image.*

(iv) *Cascade of Classifiers*
Cascade is done by using cost-aware AdaBoost. The Sensitivity threshold can be adjusted nearly to 100% false positives and true positives. This procedure executes further stages, until the desired accuracy time is reached. These techniques are trained with hundreds of "positive" sample views of an individual object along with arbitrary "negative" images of the equal size [25].

(c) *Speed Up Robust Features Approach*
"SURF: Speeded Up Robust Features" is a performance scale and rotation-invariant interest point detector and descriptor.
This feature gives better performance from other approaches with respect to repeatability, robustness and distinctiveness, yet can be computed and compared much quicker. SURF algorithm has similar principles as SIFT [28], but steps are different. The three parts are:

1. Interest point Detection
2. Local neighborhood description
3. Matching.

The implementation of SIFT has remarkable advantages over other descriptors [31]. On one hand, it's mixing of crudely localized information and apportion of gradient related features produce good distinctive power, at the same time, it can avoid the effects of localization error in terms of scale or space. The usage of relative strengths along with orientations of gradients deducts the effect of photometric changes. SURF descriptor has similar features, with the least complications. Firstly, it includes the fixation of an orientation that can be reproduced, following the information from a circular area around the point of interest. Secondly, a square area is constructed adjacent to the chosen area and the SURF descriptor is extracted out from it. Moreover, it was proposed that an upright descriptor (USURF) which is not constant to image rotation, due to faster computation was better suited for applications where the camera remains almost horizontal. If the integral image is used then filtering the image with a square becomes quicker

$$s(x, y) = \sum_{i=0}^{x} \sum_{j=0}^{y} I(i, j)$$

The addition of the original image inside a rectangle can be evaluated faster by using the integral image, requiring the value at the rectangle's four corners. Based on the Hessian matrix, SURF uses a blob detector to find points of interest (Fig. 6).

(i) *Interest Point Detections*

This feature helps us to find interest point at various scales, mostly because the search for correspondences often needs comparison images where they are seen at different scales. In other feature detection algorithms, the scale space is normally realized as an image pyramid. With a Gaussian filter, images are repeatedly smoothed, and then they are subsampled to get the next higher level of the pyramid. Therefore, several floors or stairs with various measures of the masks are calculated:

$$\sigma_{\text{approx}} = \text{current filter size} \times \left(\frac{\text{base filter scale}}{\text{base filter size}}\right)$$

The scale space is divided into a number of octaves, where an octave refers to a series of response maps cover a double of scale. In SURF, the lowest level of the scale space is obtained from the output of the pixel x pixel filters.

(ii) *Local Neighborhood Description*

Local neighborhood description provides a unique and robust illustration of an image feature, e.g., by describing the intensity distribution of the pixels within the neighborhood of the point of interest.

(iii) *Matching*

Comparing the descriptors obtained from different images, matching pairs can do with montages training image [29].

**Fig. 6** SURF detection

(d) *Fast Feature Detection*
This Feature Extraction technique helps to detect corner point (Fig. 7). After adding this feature with SURF this approach is getting error free accuracy. The object contains information about the feature points detected in a 2-D grayscale input image. The detectFASTFeatures function uses the Features from Accelerated Segment Test (FAST) algorithm [32] to find feature points. The FAST feature detector can perform full-frame feature detection at 400 Hz. The combination of these techniques results in a system which is capable of tracking average prediction errors of 200 pixels. This level of robustness allows us to track very rapid motions, such as 50° camera shake [35].

**Fig. 7** Fast feature detection

(e) *Sum of Absolute Differences (SAD)*
SAD is "Sum of Absolute Difference". The best guess recognition is the one that minimizes the SAD between training and test images. In image processing technology, Sum of Absolute Differences (SAD) [36] calculates the similarity between image blocks. It is also known as the Manhattan Distance metric or Taxicab. This calculation is done by summing the absolute values of the differences between pixels. Sum of Absolute Differences (SAD) can be utilized for various purposes, such as, the generation of disparity maps for stereo images,

motion estimation for video compression [37] and object recognition. The lowest SAD score determines the best position of the template within the search image. The general SAD distance metric becomes:

$$d_1(I_j, T) = \sum_{i=1}^{n} |I_{i,j} - T_i|$$

Now there is a set of features calculated for each face. This set of features can now be used as a classifier. Using Fast Feature detection technique and SURF extraction (SIFT is several time slower than the standard version of SURF and it is said by authors that SURF is more robust against various image alteration than SIFT) is tuned to be fairly "permissible"; that is, specifying low values of Minimum Quality and Minimum Contrast to return a lot of matches. This experiment then matches the test image against each training set and considers that the match is "recognized" when it minimizes this metric (Sum of Absolute Differences, or SAD). Every image is tried to be matched with every other and hence takes a lot of time. So it is better to create montages of each training set for small ImageSets. This approach creates 17 image montages (Fig). Then features are aggregated, and in that matches are evaluated to whole training sets, rather than to subsets.

### 3.3 Face Identification on Real-Time Video

Restart the streaming image capture (face recognition application) and detect faces. Crop each detected face (again, using the bounding box returned by the face detector), and detect and extract features in each. (The same preprocessing steps described above are implemented here). Match the extracted feature set to each of the feature sets extracted from the training image montages and find which training set image it matches most closely. That one is the "best guess prediction." Extract the label corresponding to that closest match from name of the parent directory for the training images. And that's it. After much experimentation and implemented a minimization of the SAD to determine the best guess. Framework (Fig. 8) and algorithm is given below.

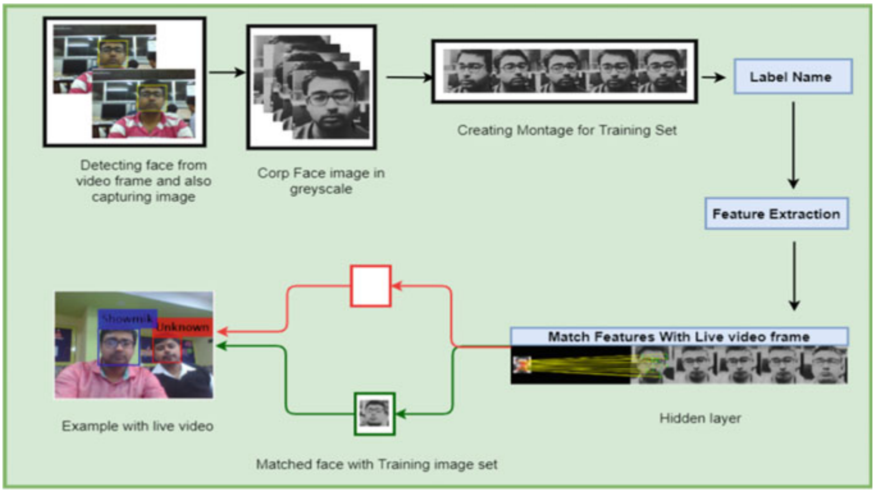


**Fig. 8** Graphical framework for video analytics

***Algorithm 1: Dataset Creating***

Step 1: Start Live Video for DataSet Creation.

Step 2: Forward to Algorithm 2. (Viola - Jones)

Step 3: Capture ”N”(N = single person multiple face images) number of face image one by one ”n” (n = different persons face images) number of persons

Loop;

Step 4: Convert Image to Grayscale

Step 5: Validate, Filter, Label creation, Resize captured face images

Step 6: Stored images in N number of folders with their respective label name (means person names).

Step 7: Create Montage from stored data for training set

Step 8: Extract features with Combined Approach (SURF + FAST)

Step 9: Forward to Algorithm 3. (Feature extraction from live Video frame and matching technique)

***Algorithm 2: Detection Face with Haar Features***

Step 1: Extract Haar Feature Selection.

Step 2: Segment human face from video frame by using mouth, nose.

Step 3: Detected 2 frames/ seconds

***Algorithm 3: Face Detection in Real-Time***

Step 1: Start Live video

Step 2: Again forward to Algorithm 2.

Step 3: extract features from detached frames

Step 4: Match with Training Set and best prediction recognition is the one that minimizes the sum of absolute differences between training and test images

Step 5: If matched

Show label Name (Person Name in detected as Blue frame)

Else

Show "Unknown" label in detected as Red frame.

Step 6: Next Frame

Loop

# 4 Experiment

Explanation of how experiments are happening.

1. Choose an option whether:
   (a) to go for live video face Recognition with already existing ImageSets, or
   (b) Create a fully new dataset adding faces, or
   (c) Add new face to existing ImageSets.
2. For [b], [c] an ImageSet is captured (of 25 images where each contains a single person's face image) for each person in the training set and verified and validated. Then ImageSets of every individual person are stored in folders and labels are given by the user.
3. Live video is started. Features are extracted from each detected face and matched with the set of features for each face in the training set. The one that minimizes the Sum of Absolute Differences (SAD) between images matrixes is assumed to be the person in the image. Labels are inserted to indicate the name of that person. If any unknown face is encountered, and then it is labeled as "UNKNOWN".

The same thing happens in fast features. Now, showing how to match SURF and Fast Features (Figs. 9 and 10).


**Fig. 9** Fast features point matching (hidden layer)

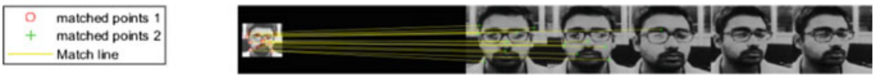


**Fig. 10** Matching with SURF points (hidden layer)

Here Single image is denoted as video frame image which is used to match the best and strongest interest points with the training set (montages).

In this Fig. 11 because of a different person's ImageSet feature is not matching. So, it checks another training set and does so in a while loop. If any of them don't match it will specify unknown on the label of that person in the video frame.

This paper is using both features because some of the interesting points is automatically taking unnecessary points as interest point to avoid this problem this paper

**Fig. 11** Not matching due to different person's image with SURF points (Hidden layer)

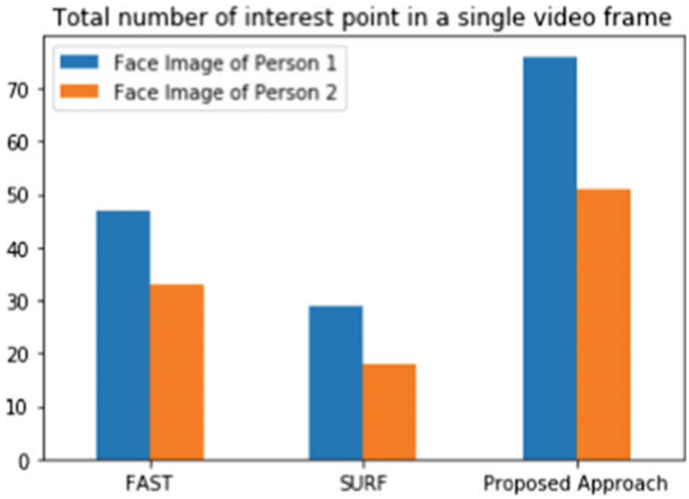


**Fig. 12** Total number of interest point in a single video frame

is using SURF and FAST features if one is making wrong decision then another will help to rectify the right decision comparison graph is given in Figs. 12 and 13.

**Datasets**: This experiment was tested with HP 15-AC024TX. Built-in webcam was used for face recognition. To create datasets from live video streaming, webcam was used to capture faces. Then it is stored in user defined and labeled folders. These folders contain individual person's face in grayscale. Each folder contains 25 images of a single person's face.

**Training Sets**: Next step for this experiment is to create training sets. So each folder image creates a montage which is going to be used as training set. That means *Total Number of folders = Total Number of Montages.*

## 5 Conclusion

By this approach for face recognition it can be said that this approach minimizes process and training time due to montages. Using images of the face from every angle increases accuracy. That means like every traditional method bigger dataset

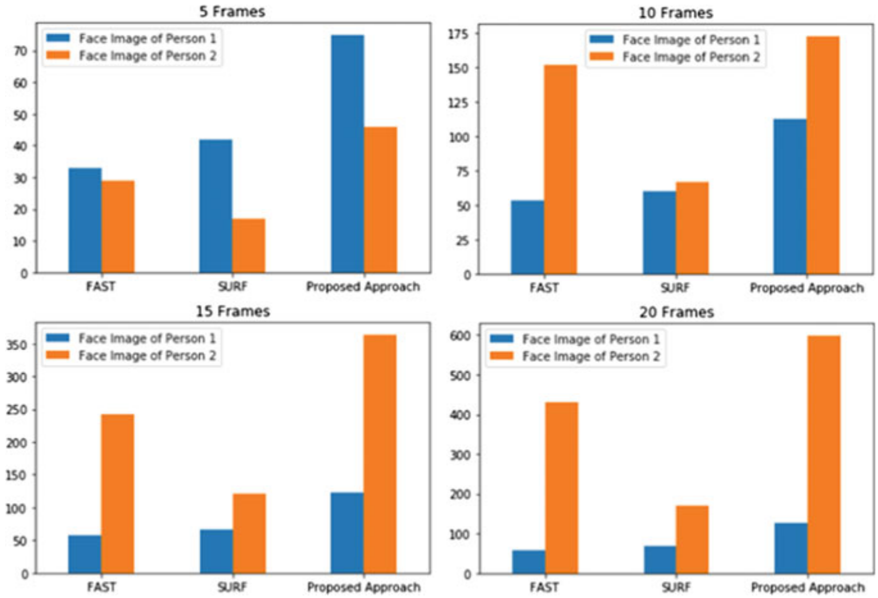


**Fig. 13** Frames wise match points

gives higher performance hence this isn't exceptional. Using local features extraction technique and with very small datasets the goal is achieved. More features may improve the performance of this proposed approach.

## 6 Future Work

Current trends in biometric technology are playing a very vital role in Information Security. This method gives superior performance with spectacles and without spectacles. However, traditional face recognition methods based on visible spectrum face challenges like object illumination, pose variation, expression changes, and facial disguises which further reduces performance. Using Infrared instead of Visual Spectrum has been found to be a substitute [38] to address all these limitations better. This approach can be further enhanced to be used in ATMs, accessing confidential files, or other sensitive materials. This can also make other security measures such as passwords and keys obsolete.

Using faces as credit cards to pay for transportation fee looks like a promising development. This could potentially streamline the process and optimize the flow of traffic drastically. The future is here.

## References

1. He R, Wu X, Sun Z, Tan T (2019) Wasserstein CNN: learning invariant features for NIR-VIS face recognition. IEEE Trans Pattern Anal Mach Intell 41(7):1761–1773
2. Madhavan S, Kumar N (2021) Incremental methods in face recognition: a survey. Artif Intell Rev 54:253–303 (2021). https://doi.org/10.1007/s10462-019-09734-3
3. Yazdi M, Bouwmans T (2018) New trends on moving object detection in video images captured by a moving camera: a survey. Comput Sci Rev 28:157–177
4. Kato H, Billinghurst M (1999) Marker tracking and HMD calibration for a video-based augmented reality conferencing system. In: Proceedings 2nd IEEE and ACM international workshop on augmented reality (IWAR'99), https://doi.org/10.1109/IWAR.1999.803809
5. Moreno P, Bernardino A, Santos-Victor J (2006) Model based selection and classification of local features for recognition using Gabor filters. Image Anal Recogn181–192
6. Gao M, Jiang J, Zou G, John V, Liu Z (2019) RGB-D-based object recognition using multimodal convolutional neural networks: a survey. IEEE Access 7:43110–43136
7. Liu J et al (2020) Feature boosting network for 3D pose estimation. IEEE Trans Patt Anal Mach Intell 42(2):494–501. https://doi.org/10.1109/TPAMI.2019.2894422
8. Savarese S, Fei-Fei L (2007) 3D generic object categorization, localization and pose estimation. In: ICCV 2007. IEEE 11th international conference on computer vision, 14–21 Oct 2007, Rio de Janeiro, Brazil
9. Moeslund T, Granum E (2001) A survey of computer vision-based human motion capture. Comput Vis Image Under 81(3):231–268
10. Lu C, Shi J, Wang W et al (2019) Fast abnormal event detection. Int J Comput Vis 127:993–1011. https://doi.org/10.1007/s11263-018-1129-8
11. Ke Y, Sukthankar R, Hebert M (2007) Event Detection in Crowded Videos. In: 2007 IEEE 11th international conference on computer vision, pp 1–8. https://doi.org/10.1109/ICCV.2007.4409011
12. Wang W, Li Z, Wu S, Zeng L (2020) Hazy image decolorization with color contrast restoration. IEEE Trans Image Proc 29:1776–1787. https://doi.org/10.1109/TIP.2019.2939946
13. Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G (2015) Recent advances in convolutional neural networks. arXiv preprint arXiv:1512.07108
14. Taigman Y, Yang M, Ranzato M, Wolf L (2014) DeepFace: closing the gap to human-level performance in face verification. In: 2014 IEEE conference on computer vision and pattern recognition, Columbus, OH, pp 1701–1708

15. Schroff F, Kalenichenko D, Philbin J (2015) FaceNet: a unified embedding for face recognition and clustering. In: 2015 IEEE conference on computer vision and pattern recognition (CVPR), pp 815–823. https://doi.org/10.1109/CVPR.2015.7298682
16. Bashbaghi S, Granger E, Sabourin R, Parchami M (2019) Deep learning architectures for face recognition in video surveillance. In: Jiang X, Hadid A, Pang Y, Granger E, Feng X (eds) Deep learning in object detection and recognition. Springer, Singapore. https://doi.org/10.1007/978-981-10-5152-4_6
17. Viola P, Jones MJ (2004) Robust real-time face detection. Int J Comput Vis 57:137–154. https://doi.org/10.1023/B:VISI.0000013087.49260.fb
18. Bolotnikova A, Demirel H, Anbarjafari G (September 01, 2017) Real-time ensemble based face recognition system for NAO humanoids using local binary pattern. Analog Integrated Circ Sig Proc: an Int J 92(3):467–475
19. Lin J, Xiao L, Wu T (2018) Face recognition for video surveillance with aligned facial landmarks learning, pp 169–178
20. Masi I, Rawls S, Medioni G, Natarajan P (2016) Pose-aware face recognition in the wild. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), Las Vegas, NV, United States, pp 4838–4846
21. Masi I, Tran A, Hassner T, Leksut JT, Medioni G (2016) Do we really need to collect millions of faces for effective face recognition? In: European conference on computer vision
22. Chen J-C, Patel VM, Chellappa R (2016) Unconstrained face verification using deep cnn features. In: IEEE winter conference on applications of computer vision (WACV), pp 1–9
23. Gary BH, Mattar M, Berg T, Learned-Miller E (2008) Labeled faces in the wild: a database for studying face recognition in unconstrained environments. Workshop on faces in 'Real-Life' images: Detection, Alignment, and Recognition, Erik Learned-Miller and Andras Ferencz and Frédéric Jurie, Oct 2008, Marseille, France
24. Burton AM, Wilson S, Cowan M, Bruce V (1999) Face recognition in poor-quality video: evidence from security surveillance. Psycholog Sci 10(3):243–248. https://doi.org/10.1111/1467-9280.00144
25. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition, CVPR 2001, Kauai, HI, USA, pI
26. Ashokkumar S, Suresh A, Kavitha MG (2019) Implication of video summarization and editing of video based on human faces and objects using SURF (speeded up robust future). Cluster Comput 22:6913–6919 (2019). https://doi.org/10.1007/s10586-018-1729-7
27. Shoelson B Face detection, training, recognition. https://in.mathworks.com/matlabcentral/profile/authors/845693-brett-shoelson
28. Lowe DG (1999) Object recognition from local scale-invariant features. In: Proceedings of the seventh IEEE international conference on computer vision, pp 1150–1157 vol.2, https://doi.org/10.1109/ICCV.1999.790410
29. Bay H, Tuytelaars T, Van Gool L (2006) SURF: speeded up robust features. In: Leonardis A, Bischof H, Pinz A (eds) Computer vision—ECCV 2006. Lecture notes in computer science, vol 3951. Springer, Berlin
30. Wang S, Li W, Wang Y, Jiang Y, Jiang S, Zhao R (2012) An improved difference of Gaussian filter in face recognition. J Multimed 7. https://doi.org/10.4304/jmm.7.6.429-433
31. Panchal P et al (2013) A comparison of SIFT and SURF. Int J Innovat Res Comput Commun Eng 1:323–327
32. Viswanathan DG, Features from accelerated segment test (FAST). http://homepages.inf.ed.ac.uk/rbf/CVonline/LOCAL_COPIES/AV1011/AV1FeaturefromAcceleratedSegmentTest.pdf
33. Owusu E, Abdulai J-D, Zhan Y (2019) Face detection based on multilayer feed-forward neural network and Haar features. Software: Pract Exper 49(1):120–129
34. Freund Y, Schapire RE (1995) A decision-theoretic generalization of on-line learning and an application to boosting. In: Computational learning theory: Eurocolt '95. Springer, Berlin, pp 23–37

35. Rosten E, Drummond T (2005) Fusing points and lines for high performance tracking. In: Tenth IEEE international conference on computer vision (ICCV'05), vol 1, Beijing, pp 1508–1515
36. Dawoud NN, Samir BB, Janier J (2011) Fast template matching method based optimized sum of absolute difference algorithm for face localization. Int J Comp Appl 18(8)
37. Richardson IEG (2004) H.264 and MPEG-4 video compression: video coding for next-generation multimedia. SERBIULA (sistema Librum 2.0)
38. Arya S, Pratap N, Bhatia K (2015) Future of face recognition: a review. Proced Comput Sci 58:578–585