I.J. Intelligent Systems and Applications, **2018, 8, 19-35**
Published Online August 2018 in MECS (http://www.mecs-press.org/)
DOI: 10.5815/ijisa.2018.08.03

Development of Robust Multiple Face Tracking Algorithm and Novel Performance Evaluation Metrics for Different Background Video Sequences

Ranganatha S
Asst. Professor, Dept. of Computer Science and Engineering, Government Engineering College,
Hassan-573201, Karnataka, India
E-mail: ranganath38@yahoo.co.in

Y P Gowramma
Professor, Dept. of Computer Science and Engineering, Kalpataru Institute of Technology,
Tiptur-572202, Karnataka, India
E-mail: gowrikit@gmail.com

Received: 26 October 2017; Accepted: 27 November 2017; Published: 08 August 2018

Abstract—In computer vision, face tracking is having wider opportunities for research activities using different background video sequences because of various factors and constraints. Due to the challenges that are increasing day by day, old/existing algorithms are becoming obsolete. There are many powerful algorithms that are limited to certain set of video sequences. In this paper, we are proposing an algorithm that detect and track multiple faces in different background video sequences. Viola-Jones face detection algorithm is used in such a way that, new face/first face need not to be in the starting frame of the selected video sequence. The proposed algorithm successfully detect new face(s) along with existing face(s) by keeping track of the facial data using BRISK feature points. The mean of the old points and new points are calculated based on the area of the facial data. The detected face(s) in further frames undergoes similarity check with existing facial data. If detected facial data and existing facial data mismatches, then the detected facial data is entered into face tracks structure. By using point tracker method, the proposed algorithm track those points that has been set for each of the facial data.

Index Terms—Face tracking, Different background, Video sequences, Multiple faces, Facial data, Face tracks structure.

I. Introduction

Advancement in technology is opening the path for researchers to look deeper into computer vision and image processing fields extensively. Video processing is a specific area of research in image processing that relies upon video sequences i.e. a set of frames/images. Face detection and tracking is the extensive area of research in video processing which faces various challenges [1, 32]. The technical challenges are: variations in illumination, pose, facial expressions, aging, occlusion and background conditions.

A. Face Tracking Overview

Face detection is the first step for further tracking it in video sequence. It manages separation of rushing face(s) out of background. Currently, the segmentation techniques use spatial or temporal information in the video sequence. Further, features are extracted from the region of interest i.e. from the segmented image. By doing feature extraction, we can determine numerous attributes and properties related with the face region being segmented. The importance of feature selection as well as extraction process is to minimize the dimensionality. The count of features selected for tracking also troubles the problem's time complexity and space complexity. The dimensionality housing in the problem needs to be lowered by considering some searching features out of all the features in the window or segmented portion. Finally, for a particular video sequence, based on the face(s) detected and features extracted in the last frame, face tracking process captures the moving face(s) among the next frames.

B. Categories of Tracking

Some of the possible categories of tracking methods are: *1). Intensity-based tracking*: Intensity of the pixel is actual low level input, which combines the object surface, illumination around the object, the noise, the internal and external properties of the camera involved during image acquisition. Some of the features extracted are Eigen values of the minimum bounding box, matrix trace, mean and variance, horizontal and vertical projections. The

Copyright © 2018 MECS

application of these methods can either take place for all pixels of an image or only for selected pixels defined by extracted features, e.g., the edges in an image. *2). Feature-based tracking*: The algorithms based on features can be organized further into three subdivisions matching to the type of features being selected: *i). Feature-based global* algorithms comprise areas, centroids, perimeters and few sort of colors as features. Polana et al. [2] provided a model for this subdivision of tracking. *ii). Feature-based local* algorithms include corner vertices, line and curve segments. *iii). Feature-based graph* algorithms include geometric relations among features and various type of distances. The above three subdivisions of tracking can also be federated. *3). Area-based tracking*: These algorithms work based on the fluctuations of frame areas analogous to the flowing entities. *4). Curve/Edge-based tracking*: These algorithms work based on setting the outer lines as contours and restoring them in upcoming images dynamically. *5). Prototype-based tracking*: These algorithms work based on relating estimated prototypes of objects, constructed using former knowledge of frame data.

C. Different Background

There exists algorithms to detect and track face(s) only in certain type of video sequences. But, due to rapid changes in technology, such algorithms will soon become obsolete. Face tracking problem is an important aspect in both static vision system and dynamic vision system. In case of dynamic vision system, we get different background in the video sequence. Dynamic vision system is defined as video acquired by,

1. Static camera with moving face(s)
2. Moving camera with static face(s)
3. Moving camera with moving face(s).

It is somewhat easier to detect and track face(s) in the first type of video sequences compared to the second and third type, which induces lot of challenges. Hence, it is creating the path for investigators to look into new methods for face tracking in video sequences with different backgrounds. The proposed algorithm is capable of detecting and tracking face(s) in all the three above said type of video sequences.

D. Problem Statement

Multiple faces detection and tracking in a video sequence is a much researched area as there are a very few algorithms that processes video sequences containing many faces. The proposed algorithm detect and track face(s) even if the face(s) do not appear in the first frame itself. The proposed algorithm is initiated with Viola-Jones algorithm [3, 4] to detect facial regions and compute BRISK feature points [5]. The mean of those points is computed and stored. In further frames, algorithm checks similarity of detected face(s) with the threshold value. If the value matches with the existing facial data, it plots a bounding box around it. In case if the value does not meet the threshold value or new face(s) has been introduced in the current frame, then the above said computation takes place again and the new facial data is stored. This process repeats until the last frame is processed. Finally, using the point tracker, point values are tracked throughout the video sequence.

E. Additional Contribution and Organization of the Paper

In this paper, we also propose metrics for evaluating the performance of our proposed algorithm. For testing and analysis we have considered video sequences from YouTube Celebrities dataset [22], Choke Point dataset [24] and HOHA dataset [25]. Based on the results obtained, it is definite that the proposed algorithm performs robust in most of the video sequences.

The remaining portion of paper includes related works that has been referred and utilized while carrying out this work and mentioned in section 2. Section 3 includes the methodology on how the proposed algorithm works. Section 4 includes the metrics that has been used for carrying out the evaluation of the proposed algorithm. Section 5 includes detailed results with analysis that supports the robustness of our proposed algorithm. Conclusion and discussion of the work are part of section 6 and section 7 talks about future works that can be done.

II. Related Works

The survey presented here covers those works that are in the same context as the proposed algorithm. However, for comprehensive completeness, the survey also provide brief information on some of the other techniques which are used for similar tasks.

Static part of the video frames which is never varying is considered as background, dynamic part is the moving object. Whenever the color of an object is completely different than rest other part of the image, then modelling is easy and simple thresh-holding can do the background subtraction. If it is more than one color in the non-object portion of image then a multi-point thresh-holding will do the background subtraction [6]. It is a decent approach, but highly responsive to variations in changing scenes. Due to these issues, this approach completely relies on a neat background prototype to minimize the impact of the variations [7].

Temporal differencing [8, 9] uses the pixel-wise distinctions among two to three successive frames of a video sequence to separate rushing parts. It is notably flexible to changing situations. But, normally it carries out a bad work of separating every applicable element, e.g., chances of gaps remaining inside the objects. Lipton et al. [8] conferred a two-frame differencing plan to identify targets that are moving in real video sequences. To overcome the deficiencies of two-frame scheme in few cases, differencing of three-frames can be used. For example, Jie Xia et al. [9] established a double difference approach by merging three-frame differencing.

Segmentation techniques that are optical-flow based uses flow vector properties of moving targets along time for detecting moving parts in a video sequence. However,

Copyright © 2018 MECS

most flow calculation methods are infeasible as for space and time constraints, very responsive to noise and cannot be used in real time video streams without necessary hardware. Barron's work [10] gives more detail about how optical flow method works.

The proposed algorithm uses Viola-Jones algorithm [3, 4] to detect face(s) in video sequences. The before said algorithm efficiently detect the face(s) if they are present in the first frame of video sequences. The algorithm was basically trained to detect frontal faced face(s) that are bound to certain constraints. A survey on face recognition techniques by Ranganatha S et al. [1] elaborates the constraints of face detection in detail. The work in literature [11] by Jones et al. is a continuation to the work in literature [3] by Viola et al.

In case of feature based tracking, features are used as the attribute of the face characteristics. Color, intensity (contrast, brightness), texture can be thought as primary features while some other secondary features are also given, fusing primary feature information; example Haar-like features [3, 12], Histograms of Oriented Gradients (HOG) features [13], Scale Invariant Feature Transform (SIFT) features [14], CENsus TRansform hISTogram (CENTRIST) [15], Speeded-Up Robust Features (SURF) [16] etc. Haar-like features are especially designed for face by Viola and Jones [3]. S. Leutenegger et al. [5] presented an algorithm that is more efficient than SHIFT and SURF feature points. The detected points are far better and reliable in the frames that are given as input and coined as "Binary Robust Invariant Scalable Keypoints" (BRISK) feature points. BRISK feature points are used in the proposed algorithm, which gives a vital support for detecting multiple facial areas in the given input video frame and enables to track those points efficiently using point trackers.

Lucas and Kanade [17], Tomasi and Kanade [18] have done an extensive research on tracking certain regions in video frames. Using the points present in the frames, they developed a point tracking algorithm to track those points which is popularly known as KLT algorithm. Shi and Tomasi [19] in their work "Good Features to Track" have mentioned certain features that are very less susceptible to constraints such as illumination, aging etc. The point features generated by their method are not variant to rotational and translational transformations; but, variant to affine or projective transformations.

Literature in [20] discusses the problem of detecting and tracking human face(s) in video sequences captured using a static camera. Motion in the video sequences is because of people movement. Banks and supermarkets surveillance video sequences are considered for analysis here. But, this approach fails to track human face(s) in moving camera with static face(s) and moving camera with moving face(s) video types. Ranganatha S et al. [21] have proposed a novel algorithm to track human face in video sequences by fusing KLT technique with centroid and corner points. It effectively improvised the KLT algorithm by tracking faces in few more frames in certain videos than the original algorithm. For testing and analysis, they have considered the video sequences which are available as part of the literature [22]. Their method works for static camera with moving face(s) video sequences and fails to track multiple faces. Ranganatha S et al. [26] have also developed an algorithm for face tracking by integrating improved CAMSHIFT [27, 31] and kalman filter [28-30]. Their face tracking algorithm work faster and solve the problem of illumination. But, they have used only static camera with moving face(s) video sequences from YouTube Celebrities dataset [22] for testing and analysis. Their method cannot be applied to track multiple faces.

The popular existing face(s) tracking algorithms in video sequences are lagging behind whenever they undergo tracking of multiple faces in different background video sequences. The algorithms also fail when new face(s) that are not available in the beginning frame are added in the upcoming frames. As a research area, face tracking and recognition is very active and the technology has grown to better heights since after the survey of R. Chellappa et al. [23]. However, rapid research is going on in this area to detect and track face(s) in different background video sequences.

III. Methodology

The proposed system architecture is summarized in Fig.1 below.

The technique that is being employed to detect face(s) and track them in the input video sequence involves two major parts:

1. Detect the face(s) present.
2. Track the detected face(s).

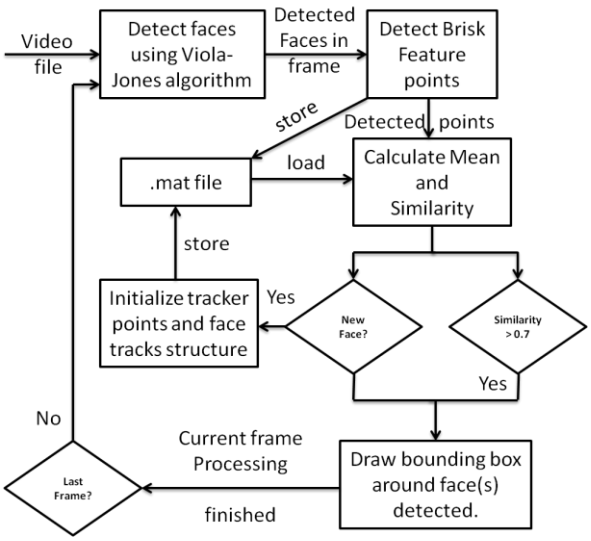


Fig.1. Proposed system architecture

A. Detection of Face(s)

This section revolves around how the face(s) are detected in a frame and how the track data is persistently stored in the external file. By using those detected data values; the facial points data in the previous frame is compared with that of the facial points data detected in

the current frame. If the match is found, then plot a bounding box surrounding the face area detected. If the current frame consists of some new face(s) whose data is not present in the file, then the points data of the new facial region that is present in the current frame is saved into the file. The same procedure is repeated until the last frame is processed.

Algorithm 1: Detection of Face(s)

Input: Different background video types with various challenges.
Output: Bounding box around detected face region or new facial point data is written onto a file.
Steps:

1. Detect face(s) present in the frame using Viola-Jones algorithm and plot bounding box around it.
2. Compute the BRISK feature points in the detected face region.
3. If it is not the first frame, then copy oldpoints from the tracks file to the points data variable.
4. If points contain equal to or more than 3 points; then go to step a. else step h.
   a. Load the newpoints from the file.
   b. Estimate geometric transform between points and newpoints.
   c. Calculate mean of newpoints.
   d. Calculate area $\Psi 1$ and $\Psi 2$.
      $\Psi 1$ ← Area of new bounding box.
      $\Psi 2$ ← Area of old bounding box.
   e. Compute intersection value φ of $\Psi 1$ and $\Psi 2$.
   f. Calculate similarity to identify facial match by,
      $$\text{Similarity} = \left(\frac{\varphi}{\Psi 1} + \frac{\varphi}{\Psi 2} \right) \diagup 2$$
   g. If similarity is more than 0.7; then, isgood variable will be set to true. Go to step i.
   h. isgood variable will be set to false.
   i. Return isgood and plot bounding box.
5. If the facial point data does not exist in the file then write the data into the file as follows.
   a. If bounding box exist around the facial region, but its point data is not present in the file then,
      New face ← Size (bbox).
   b. Iterate through all faces detected.
   c. If face is detected and it is a new face; then, check if there exists a matching ID.
   d. Initialize tracker points to the new face data in the file.
   e. Update face detected and face tracks structure by computing its data using method specified in step 4.
6. If the face detected is new face then add new face details as follows.
   a. Get details of current active tracker.
   b. Extract feature points of bounding box with respect to frame.
   c. Check if there are points present then continue with next step else terminate.
   d. Assign trackerID with current trackerID+1.
   e. Increase maximum number of trackers value by 1.
   f. Get features to track and initialize tracker.
   g. Update the face data as described in step 7.
7. If the facial point data does exist in the file and the facial data has been updated in the current frame; then it has to modify in the tracks file as follows.
   a. Verify if there exists a detected face in the ongoing frame.
   b. If the face detected is new face then add the new face data and repeat step 6.
   c. Else append face to current frame by setting its trackID, bounding box details and landmarks in and around the facial region.

B. Tracking of Face(s)

This section revolves around how the detected face(s) in each frame are tracked throughout the video sequence. This tracking methodology is dynamic as the new facial point data is updated when each frame is processed and that data is used in tracking of face(s) in further frames.

Algorithm 2: Tracking of Faces(s)

Input: Video sequence and track file.
Output: Tracking of the facial point data present in track file.
Steps:

1. Initialize point tracker.
2. Load video and file.
3. For each frames in video, load all previous facial point data from .mat file.
4. Call detection algorithm to detect face(s) in the current frame.
5. Iterate through step 3 until last frame is processed.

C. Time Complexity

The algorithm complexity can be estimated by parts as follows:

1. $O(1)$, When detecting the faces in the input frame.
2. $O(n)$, When estimating the BRISK feature points.
3. $O(n^2)$, When calculating mean and similarity between old and new points.
4. $O(n \times m)$, Where n is for plotting face around face region and m is to update face tracks structure to file.
5. $O(n)$, When extracting feature points and initializing tracker ID for them.

So, overall algorithm complexity will be the sum of the complexities mentioned above, i.e.

$$O(overall) = O(1) + O(n) + O(n^2) + O(n \times m) + O(n).$$

We ignore $O(1)$, as it contributes very less compared to other complexities. It yields

$$O(overall) = O(n) + O(n^2) + O(n \times m) + O(n);$$
$$O(overall) = 2O(n) + O(n^2) + O(n \times m).$$
As $O(n \times m) \approx O(n^2)$, we get
$O(overall) = 2O(n) + 2O(n^2)$. Which results in
$O(overall) = 2(O(n) + O(n^2))$. Hence,
$O(overall) = 2O[n + n^2]$.

IV. Metrics

The algorithm developed has the combination of BRISK points; its mean and similarity with points in the previous frame and matching for the threshold value. There is no particular algorithm that has been developed of the same type or which computes nearly similar to the proposed algorithm. Hence, to assess the attainment of our proposed algorithm, we have formulated certain equations which are explained in this section. Due to the lack of similar algorithms, this algorithm undergoes self-experimentation to check its performance.

The performance standards are calculated based on the process that the algorithm undergoes. The video sequences are taken from various datasets as follows:

1. YouTube Celebrities dataset [22].
2. Choke Point dataset [24].
3. HOHA dataset [25].

The videos chosen for the performance evaluation have certain challenges in each of them such as multiple faces, illumination changes, face(s) not being present in the first frame and so on. Also, the videos are chosen to satisfy different background conditions.

We have considered 4 metrics/evaluation techniques to analyse the performance of the algorithm during its execution. The 4 evaluation techniques are as follows:

1. Spatial Difference (SD).
2. Dimensional Ratio (DR).
3. Dimensional Error (DE).
4. Detection Factor (DF).

A. Spatial Difference (SD)

The Spatial Difference is the measure of difference in the regions between the area of bounding box that is picked as Ground Truth (GT) with the area if the bounding box of the ongoing frame taken as System Track (ST).

$$\text{i.e. } Spatial\ Difference(SD) = Area(GT, ST). \quad (1)$$

Where

$$Area(GT, ST) = \sum_{i=1}^{n} Area_i(GT, ST);$$

$$Area_i(GT, ST) = Area_i(GT) - Area_i(ST).$$

Where n = Frames count in the selected video sequence that contains the facial region, $Area_i$ = Area of the i[th] frame where i ranges from 1 to n.

For this and all the future evaluations wherever we need Ground Truth data, we had considered i-1 frame data as the Ground Truth data if the current frame under process is i. We are considering marginal threshold range for the values of SD as $-100000 < SD < 100000$ for the data to be accepted and valid, as there will be slight changes in the area of bounding box between previous frame and current frame. i.e.

$$SD = \begin{cases} 0 & ideal, \\ > -100,000 & accept, \\ < 100,000 & accept, \\ otherwise & reject. \end{cases}$$

B. Dimensional Ratio (DR)

The Dimensional Ratio is the sum of the values of individual frames (DR_i) with regard to that of total count of frames (n). The value of an individual frame is the ratio between GT and ST data of the current frame under process.

i.e.

$$Dimensional\ Ratio(DR) = \frac{\sum_{i=1}^{n} DR_i}{n}. \quad (2)$$

Where

$$DR_i = \frac{GT_i}{ST_i};$$

With

$$GT_i = \frac{Width_i}{Height_i} \quad \text{And} \quad ST_i = \frac{Width_{i-1}}{Height_{i-1}}.$$

n = Number of frames.
GT = Ground Truth.
ST = System Track.

The Dimensional Ratio is considered to be valid only if the DR results in 1 with an exemption of ±0.2 for some rare constraints such as noise, video quality and few other factors. But to consider it as optimal, the DR should be 1. As for a perfect bounding box, the ratio of their width with respect to height will always be 1. For all True Positive (TP) results, the width and height varies only if the result is False Positive (FP) and/or False Negative (FN). False Positive situation is the one in which total pixels which has the object that has no link with the Ground Truth object. False Negative is the situation in which pixels that are getting matched doesn't contain the actual object but are still being matched. Hence, when the pixels overlap with the required object, in our case it is the facial region; then the result obtained will be considered as True Positive result.

Copyright © 2018 MECS

So

$$-0.2 \le DR_i \le +0.2 .$$

C. Dimensional Error (DE)

The video sequences that are being used for evaluation and analysis are majorly of surveillance and with specific challenges. In both set of video sequences the faces which we are getting are either at the constant distance from camera or walking towards the camera, none of them are walking backwards depicting as if they are moving away from the camera. Based on this criteria, we form a theoretical analysis that the bounding box value/dimension have to be either constant or increasing but never decreasing during transition from i-1th frame to ith frame leading to what we have coined as Dimensional Error. After thorough analysis of the algorithm, we have framed an equation which if results in false; leads to the existence of Dimensional Error between the transitions. The equation framed is as follows:

$$\frac{Length_{max}\left(GT_i, ST_i\right)}{Length\left(ST_i\right)} < \frac{Length_{max}\left(GT_i, ST_i\right)}{Length\left(GT_i\right)} .$$

So

$$Dimensional\ Error\left(DE_i\right) = \begin{cases} 1 & if \quad false, \\ 0 & if \quad true. \end{cases}$$

i.e.
$$DE\% = \frac{\sum_{i=1}^{n} DE_i}{N} \times 100 . \qquad (3)$$

Where

N = Number of frames having face detected.

D. Detection Factor (DF)

The Detection Factor is the measure of total count of frames where in the face is detected with respect to the total count of frames in which that face exists. Looking into the statement, we feel it's simple as there will be faces in every frame and it will be detected. But, since the algorithm even deals with (i) detection followed by tracking of multiple faces in the selected video sequence which may come in any frame, (ii) the face(s) can also have non continuous presence throughout the video sequence; even then the face tracks structure of the algorithm keeps hold of any number of face(s) that come in between any number of frames. So while tracking multiple faces, the DF value will going to decrease due to various challenges. The DF value is calculated as follows:

$$DF = \frac{Total\ count\ of\ frames\ in\ which\ face(s)\ detected(N)}{Total\ count\ of\ frames\ in\ which\ face(s)\ exist(n)} .$$

We shall also compute the efficiency percentage of the algorithm to detect the face(s) using DF% as follows:

$$DF\% = DF \times 100 . \qquad (4)$$

By using the above mentioned techniques; SD, DR, DE, DE% and DF%, we have tested the selected video sequences that are having certain challenges associated with each of them and their results are tabulated and analysed. Since the video sequences we have considered for analysis have the number of frames ranging from 100-3000 and many number of people, we are limiting each video sequence to certain number of frames and/or certain number of faces which are explained in detail during its analysis.

V. Results and Analysis

In this section, we apply above mentioned evaluation techniques on various video sequences which belongs to the categories mentioned below:

- Static camera with moving face(s).
- Moving camera with static face(s).
- Moving camera with moving face(s).

Each of these type of video sequence will consist certain challenges in it. The algorithm being developed works on all types of video sequences as mentioned above. The height and width considered for calculation are in terms of pixels. Let us examine various set of video sequences in later parts.

Equation (1) states the difference in the area of bounding boxes between Ground Truth and System Track. We make use of (1) in our computation to show the bounding box variation throughout the analysis and results are tabulated for the three different categories of videos. Equations (2) and (3) compute Dimensional Ratio and Dimensional Error, both of them act directly on the dimensions of bounding box. The 0's obtained from (3) states that there is no Dimensional Error between consecutive frames. Equation (4) is used to measure the total number of frames in which the face has been detected with the total number of frames present. This metric gives us the idea about the robustness of the algorithm to track the detected face(s).

A. Single Face and Multi-Colored Video Sequence

The video sequence that is being taken for analysis in this section is from the YouTube Celebrities dataset [22] which comprise of clipping of an actress's interview. The video sequence named 0286_01_016_angelina_jolie.avi has been taken from the dataset for the analysis purpose. This video has a frequent illumination change in it making it difficult for certain existing algorithms to fail to keep track of the face region. It falls under the video category of static camera with moving face(s). Some of the frames of the selected video sequence which contains the face(s) detected and tracked are shown in Fig.2. The statistics of frames of that video sequence under track are shown in Table 1 below.

Copyright © 2018 MECS

Frame # 15	
Frame # 54	
Frame # 118	

Fig.2. Frames of 0286_01_016_angelina_jolie.avi

This video sequence has 186 frames, computation and evaluation has been done on all 186 frames and the results are tabulated. Due to large chunk of data we show the snippet of the computed result in tabulation.

Table 1. Statistics of 0286_01_016_angelina_jolie.avi

Frame No	Width	Height	Area	SD	DR	DE
3	81	81	6561	0	1	0
5	80	80	6400	161	1	1
7	80	80	6400	0	1	0
9	83	83	6889	-489	1	1
10	83	83	6889	0	1	0
11	83	83	6889	0	1	0
12	83	83	6889	0	1	0
13	83	83	6889	0	1	0
14	83	83	6889	0	1	0
15	83	83	6889	0	1	0
16	83	83	6889	0	1	0
17	83	83	6889	0	1	0
18	83	83	6889	0	1	0
19	83	83	6889	0	1	0
20	83	83	6889	0	1	0

The above table shows the results till frame 20 but evaluation is done till frame 186. The final overall result after the complete analysis of frames is shown as a graph in Fig.3.

B. Multiple Faces but One Face at a Time with Limited Lighting

The video sequence that is being considered for analysis in this section is from the Choke Point dataset [24]. It comprises of a surveillance video sequence where people are walking through the door one after another. This video sequence has multiple faces incoming one after another and face is not available in the first frame itself. It falls under the video sequence category of static camera with moving face(s).

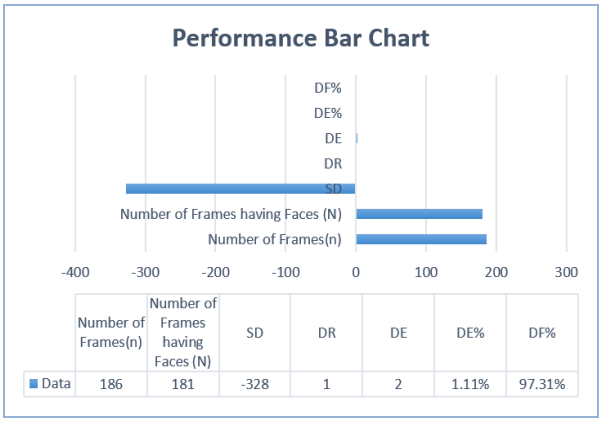


Fig.3. Bar chart of 0286_01_016_angelina_jolie.avi

The video P1ES1C1.mp4 has 1910 frames, not every face is present in all the frames and there are more than 10 faces. So, rather limiting to number of frames we limit to number of faces. Randomly chosen three frames of this video sequence in which the face(s) are detected and tracked are shown in Fig.4.

Analysis of the first 3 faces that are detected and tracked are displayed in Tables 2, 3 and 4 respectively. Due to large number of frames and faces being present in this video sequence, we shall limit our evaluation up to certain number of faces. For each face we have carried out full computation consisting of that face in the video sequence. But, only few frames are tabulated in this paper for reference based on the faces accordingly.

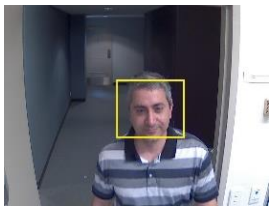
Frame # 122	
Frame # 188	
Frame # 369	

Fig.4. Frames of P1ES1C1.mp4

Face 1

In the video sequence, face 1 appears in 55 frames (n) and it is tracked in 29 frames (N). The portion of the tabulation is as follows:

Table 2. Statistics of face 1 in P1ES1C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
189	108	108	11664	0	1	0
191	119	119	14161	-2497	1	0
193	120	120	14400	-239	1	0
195	126	126	15876	-1476	1	0
196	126	126	15876	0	1	0
198	136	136	18496	-2620	1	0
200	136	136	18496	0	1	0
202	139	139	19321	-825	1	0
204	142	142	20164	-843	1	0
205	142	142	20146	0	1	0
206	142	142	20146	0	1	0
208	154	154	23716	-3552	1	0
209	154	154	23716	0	1	0
211	153	153	23409	307	1	1

Table 2 shows data starting from frame 189 from which face 1 enters into the scene and present till frame 211, but in actual large data, face 1 is present till frame 243. The final overall result after the complete analysis of 55 frames is shown as a graph in Fig.5.

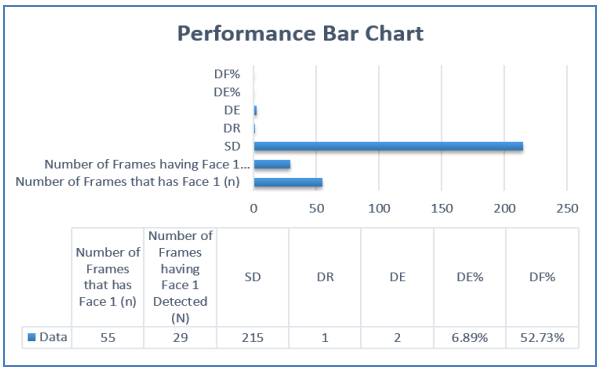


Fig.5. Bar chart of face 1 in P1ES1C1.mp4

Face 2

In the video sequence, face 2 appears in total of 33 frames (n) and it is detected in 22 frames (N). The portion of the tabulation is as follows:

Table 3. Statistics of face 2 in P1ES1C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
294	116	116	13456	0	1	0
295	124	124	15376	-1320	1	0
297	135	135	18225	-2849	1	0
299	142	142	20164	-1939	1	0
301	148	148	21904	-1740	1	0
303	152	152	23104	-1200	1	0
305	148	148	21904	1200	1	1
307	156	156	24336	-2432	1	0
308	156	156	24336	0	1	0
310	162	162	24244	-1908	1	0
311	162	162	24244	0	1	0

Table 3 shows data starting from frame 294 from which face 2 enters into the scene and present till frame 311. In actual large data, face 2 is present till frame 326. The final overall result after the complete analysis of 33 frames is shown as a graph in Fig.6.

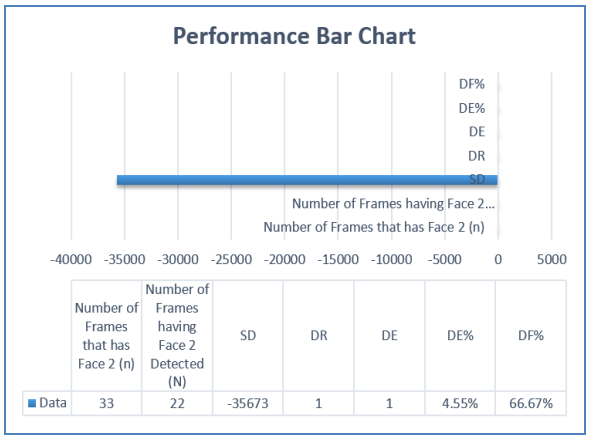


Fig.6. Bar chart of face 2 in P1ES1C1.mp4

Face 3

In the video sequence, face 3 appears in total of 53 frames (n) and it is detected in 39 frames (N). The portion of the tabulation is as follows:

Table 4. Statistics of face 3 in P1ES1C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
354	107	107	11449	0	1	0
355	110	110	12100	-651	1	0
357	110	113	12769	-669	1	0
359	110	121	14641	-1872	1	0
361	110	120	14400	241	1	1
363	110	130	16900	-2500	1	0
365	110	129	16641	259	1	1
367	110	131	17161	-520	1	0
369	110	129	16641	520	1	1
371	110	136	18496	-1855	1	0

Table 4 shows data starting from frame 354 from which face 3 enters into the scene and present till frame 371. But in actual large data, face 3 is present till frame 406. The final overall result after the complete analysis of 53 frames is shown as a graph in Fig.7.

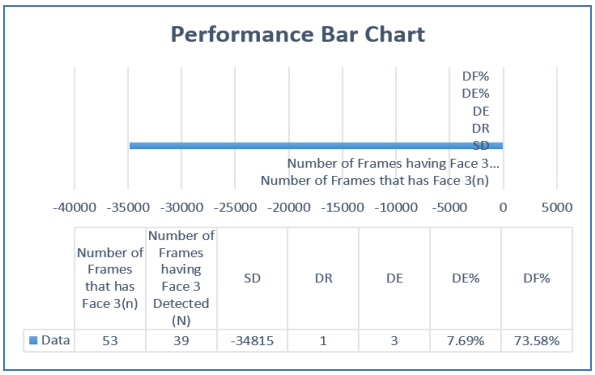


Fig.7. Bar chart of face 3 in P1ES1C1.mp4

Copyright © 2018 MECS

C. Multiple Faces but One Face at a Time with More Illumination

The video sequence that is being taken for analysis in this section is from the Choke Point dataset [24]. It comprises of a surveillance video sequence where people are walking through the door one after another. This video sequence has multiple faces incoming one after another and face is not available in the first frame itself. It falls under the video sequence category of static camera with moving face(s). The video sequence P2ES2C1.mp4 has 129 frames after removing excess frames while converting frames into video sequence from images. There are more than 10 faces and not every face is present in all the frames. So, rather limiting to number of frames we limit to number of faces say first 3 faces that are tracked for the analysis. In the scenes as shown in the pictures of Fig.8 below, we can see that there is a lot of illumination around and also reflection in the doors of the entrance. The reflection and bright light leads to the creation of False Negative tracks in certain frames. By ignoring False Negative results and considering good tracks, we carry out analysis that is shown in Tables 5, 6 and 7 respectively.

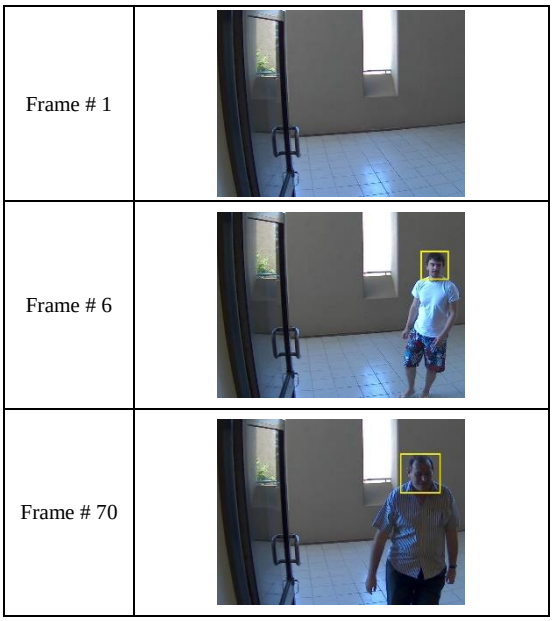


Fig.8. Frames of P2ES2C1.mp4

Face 1

In the video sequence, face 1 appears in total of 21 frames (n) and it is detected in 9 frames (N). The tabulation is as follows:

Table 5. Statistics of face 1 in P2ES2C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
3	129	129	16641	0	1	0
6	144	144	20736	-4095	1	0
8	128	128	16384	4352	1	1
13	90	90	8100	8284	1	1
15	92	92	8464	-364	1	0
17	92	92	8464	0	1	0
19	94	94	8836	-372	1	0
21	95	95	9025	-189	1	0
23	100	100	10000	-975	1	0

Table 5 shows data starting from frame 3 from which face 1 enters into the scene and present till frame 23. After the complete analysis of 21 frames, the result is shown as a graph in Fig.9 below.

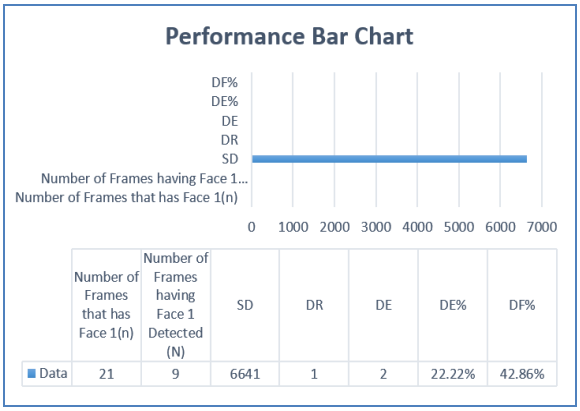


Fig.9. Bar chart of face 1 in P2ES2C1.mp4

Face 2

In the video sequence, face 2 appears in total of 21 frames (n) and it is detected in 15 frames (N). The portion of the tabulation is as follows:

Table 6. Statistics of face 2 in P2ES2C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
26	92	92	8464	0	1	0
27	92	92	8464	0	1	0
28	101	101	10201	-1737	1	0
30	105	105	11025	-824	1	0
32	106	106	11236	-211	1	0
34	113	113	12769	-1533	1	0
36	112	112	12544	225	1	1
38	120	120	14400	-1856	1	0
40	125	125	15625	-1225	1	0
41	125	125	15625	0	1	0
42	125	125	15625	0	1	0

Copyright © 2018 MECS

Table 6 shows data starting from frame 26 from which face 2 enters into the scene and present till frame 42, but in actual large data, face 2 is present till frame 46. The final overall result after the complete analysis of 21 frames is shown as a graph in Fig.10 below.

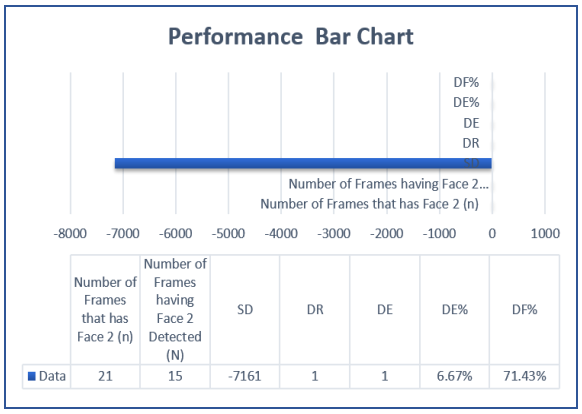


Fig.10. Bar chart of face 2 in P2ES2C1.mp4

Face 3

In the video sequence, face 3 appears in total of 27 frames (n) and it is detected in 21 frames (N). The tabulation is as follows:

Table 7. Statistics of face 3 in P2ES2C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
48	127	127	16129	0	1	0
49	136	136	18496	-2367	1	0
50	136	136	18496	0	1	0
51	136	136	18496	0	1	0
53	142	142	20164	-1668	1	0
54	142	142	20164	0	1	0
55	142	142	20164	0	1	0
56	142	142	20164	0	1	0
58	158	158	24964	-4800	1	0
59	158	158	24964	0	1	0
60	158	158	24964	0	1	0
61	158	158	24964	0	1	0
63	189	189	35721	-10757	1	0
64	189	189	35721	0	1	0
65	189	189	35721	0	1	0
66	189	189	35721	0	1	0
68	202	202	40804	-5083	1	0
70	209	209	43681	-2877	1	0
71	209	209	43681	0	1	0
73	214	214	45796	-2115	1	0
74	206	206	42436	3360	1	1

Table 7 shows data starting from frame 48 from which face 3 enters into the scene and present till frame 74. The final overall result after the complete analysis of 27 frames is shown as a graph in Fig.11 below.

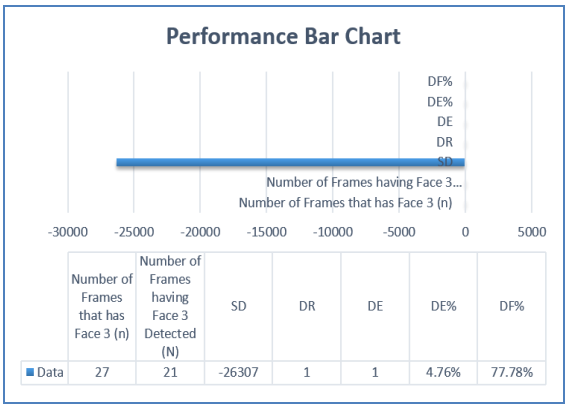
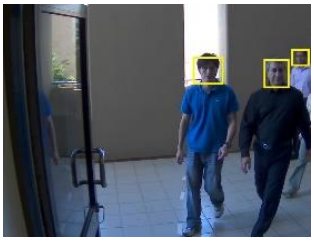


Fig.11. Bar chart of face 3 in P2ES2C1.mp4

D. Multiple Faces Appearing in Certain Frames

The video sequence that is being taken for analysis in this section is from the Choke Point dataset [24]. It comprises of a surveillance video sequence where people are walking through the door one after another. This video has multiple faces incoming at a time and also face(s) are not available in the first frame itself. It falls under the video category of static camera with moving face(s). The video P2ES5C1.mp4 has 45 frames after removing excess frames while converting frames into video sequence from images. There are more than 10 faces and not every face is present in all the frames. So, rather limiting to number of frames we limit to the number of faces; say first 3 faces that are tracked for the analysis. An important observation here is even though we have several face(s) in the selected video sequence; the total number of frames are less, because multiple faces appear in the single frame itself rather spreading across different frames. The frames are as shown in the pictures of Fig.12 below, we can see lot of people walking through the entrance. By considering only the good tracks, we carry out analysis that is shown in Tables 8, 9 and 10 respectively.

Frame # 118	
Frame # 121	

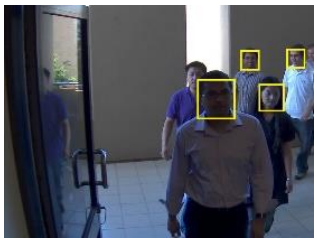
Frame # 148	

Fig.12. Frames of P2ES5C1.mp4

Face 1

In the video sequence, face 1 appears in total of 28 frames (n) and it is detected in 14 frames (N). The tabulation is as follows:

Table 8. Statistics of face 1 in P2ES5C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
109	81	81	6561	0	1	0
110	81	81	6561	0	1	0
112	86	86	7396	-835	1	0
114	88	88	7744	-348	1	0
116	87	87	7569	175	1	1
118	83	83	6889	680	1	1
123	86	86	7396	-507	1	0
128	85	85	7225	171	1	1
130	87	87	7569	-344	1	0
131	90	90	8100	-531	1	0
133	90	90	8100	0	1	0
135	92	92	8464	-364	1	0
136	96	96	9216	-752	1	0

Table 8 shows data starting from frame 109 from which face 1 enters into the scene and present till frame 136. The final overall result after the complete analysis of 28 frames is shown as a graph in Fig.13.

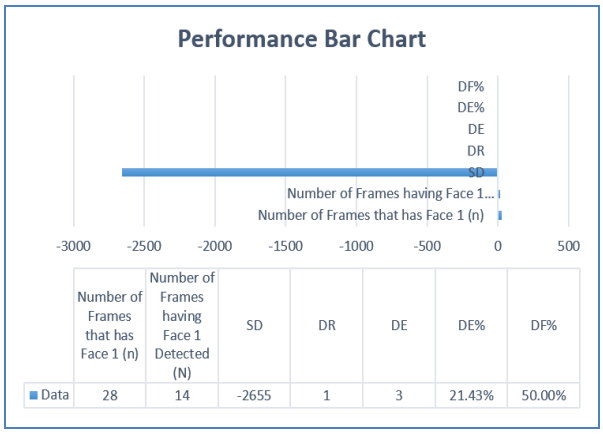


Fig.13. Bar chart of face 1 in P2ES5C1.mp4

Face 2

In the video sequence, face 2 appears in total of 17 frames (n) and it is detected in 9 frames (N). The portion of the tabulation is as follows:

Table 9. Statistics of face 2 in P2ES5C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
118	85	85	7225	0	1	0
119	85	85	7225	0	1	0
121	90	90	8100	-875	1	0
125	94	94	8836	-736	1	0
126	94	94	8836	0	1	0
130	92	92	8464	372	1	1
131	92	92	8464	0	1	0
133	93	93	8649	-185	1	0
134	96	96	9216	-567	1	0

Table 9 shows data starting from frame 118 from which face 2 enters into the scene and present till frame 134. The final overall result after the complete analysis of 17 frames is shown as a graph in Fig.14.

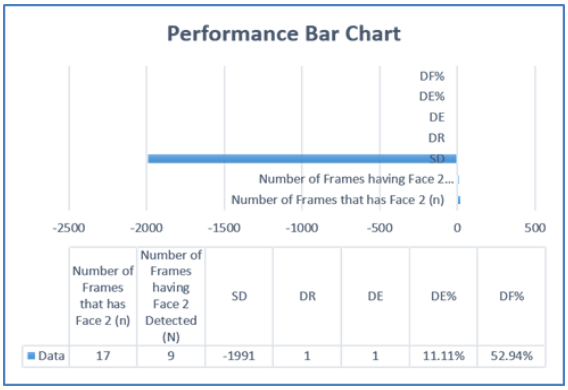


Fig.14. Bar chart of face 2 in P2ES5C1.mp4

Face 3

In the video sequence, face 3 appears in total of 14 frames (n) and it is detected in 5 frames (N). The tabulation is as follows:

Table 10. Statistics of face 3 in P2ES5C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
137	89	89	7921	0	1	0
138	93	93	8649	-728	1	0
140	95	95	9025	-376	1	0
148	98	98	9604	-579	1	0
150	104	104	10816	-1212	1	0

Table 10 shows data starting from frame 137 from which face 3 enters into the scene and present till frame 150. The final overall result after the complete analysis of 14 frames is shown as a graph in Fig.15.

Copyright © 2018 MECS

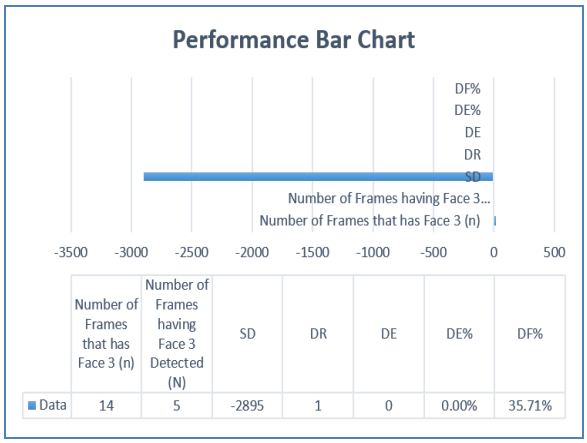


Fig.15. Bar chart of face 3 in P2ES5C1.mp4

E. Multiple Faces Appearing in Frames and Subjected to Multiple Illuminations

The video sequence that is being taken for analysis in this section is from the Choke Point dataset [24]. It comprises of a surveillance video sequence where people are walking through the door one after another. This video has multiple faces incoming together and also face(s) are not available in the first frame itself. It falls under the video category of static camera with moving face(s). The video sequence P2LS5C1.mp4 has 26 frames after removing excess frames while converting frames into video sequence from images. There are more than 10 face(s) and not every face is present in all the frames. An important observation here is even though we have several face(s) in the selected video sequence, the total number of frames are less. This is because, multiple faces appear in the single frame itself rather spreading across different frames. So, rather limiting to number of frames we limit to number of faces; say first 3 faces that are tracked for the analysis. The scenes are varied based on the lighting conditions and also we obtain False Negative results due to reflection on the glass door. As shown in the pictures of Fig.16 below, we can see lot of people walking through the entrance. By considering only the good tracks, we carry out analysis that is shown in Tables 11, 12 and 13 respectively.

Frame # 47	
Frame # 55	

Frame # 67	

Fig.16. Frames of P2LS5C1.mp4

Face 1

In the video sequence, face 1 appears in total of 17 frames (n) and it is detected in 7 frames (N). The tabulation is as follows:

Table 11. Statistics of face 1 in P2LS5C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
46	79	79	6241	0	1	0
47	83	83	6889	-648	1	0
58	85	85	7225	-336	1	0
59	85	85	7225	0	1	0
60	85	85	7225	0	1	0
61	85	85	7225	0	1	0
62	85	85	7225	0	1	0

Table 11 shows data starting from frame 46 from which face 1 enters into the scene and present till frame 62. The final overall result after the complete analysis of 17 frames is shown as a graph in Fig.17.

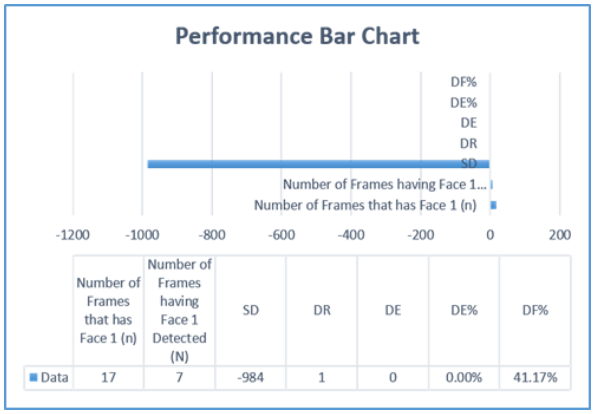


Fig.17. Bar chart of face 1 in P2LS5C1.mp4

Face 2

In the video sequence, face 2 appears in total of 14 frames (n) and it is detected in 6 frames (N). The portion of the tabulation is as follows:

Table 12. Statistics of face 2 in P2LS5C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
50	81	81	6561	0	1	0
51	82	82	6724	-163	1	0
53	84	84	7056	-332	1	0
55	84	84	7056	0	1	0
62	85	85	7224	-168	1	0
63	85	85	7224	0	1	0

Copyright © 2018 MECS

Table 12 shows data starting from frame 50 from which face 2 enters into the scene and present till frame 63. The final overall result after the complete analysis of 14 frames is shown as a graph in Fig.18.

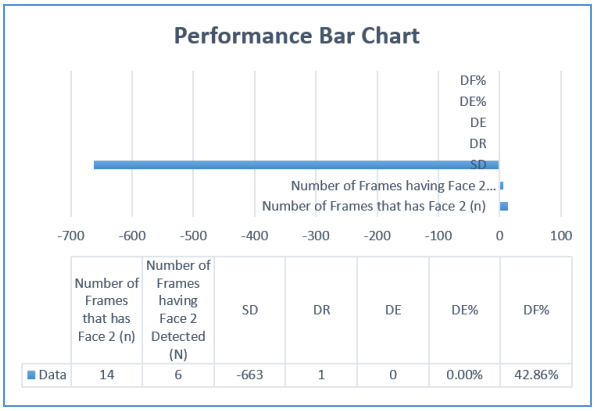


Fig.18. Bar chart of face 2 in P2LS5C1.mp4

Face 3

In the video sequence, face 3 appears in total of 8 frames (n) and it is detected in 8 frames (N). The tabulation is as follows:

Table 13. Statistics of face 3 in P2LS5C1.mp4

Frame No	Width	Height	Area	SD	DR	DE
64	84	84	7056	0	1	0
65	85	85	7225	-169	1	0
66	89	89	7921	-696	1	0
67	93	93	8649	-728	1	0
68	95	95	9025	-376	1	0
69	95	95	9025	0	1	0
70	96	96	9216	-191	1	0
71	96	96	9216	0	1	0

Table 13 shows data starting from frame 64 from which face 3 enters into the scene and present till frame 71. The final overall result after the complete analysis of 8 frames is shown as a graph in Fig.19.

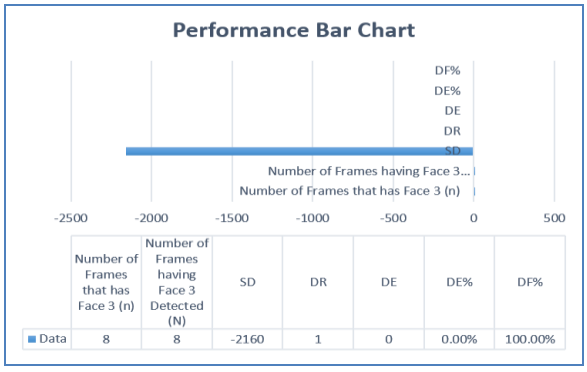


Fig.19. Bar chart of face 3 in P2LS5C1.mp4

F. Multiple Faces Appearing in Frames and Subjected to Camera as well as Face Movement

The video sequence that is being taken for analysis in this section is from the HOHA dataset [25]. It comprises of Hollywood movie clippings where scenes of different movies with certain challenges induced in them. This video has multiple faces subjected to multiple occurrences in different frames along with both camera as well as person/object are in motion. The video Dead_Poets_Society_00068.avi has 233 frames in total. There are more than 20 face(s) and not every face is present in all the frames. Some of the frames of this video sequence are shown in Fig.20. We shall consider first 3 faces as we did before rather than considering frames to carry out the analysis smoothly. It falls under the video category of moving camera with moving face(s). By considering the tracks, we carry out analysis that is shown in Tables 14, 15 and 16 respectively.

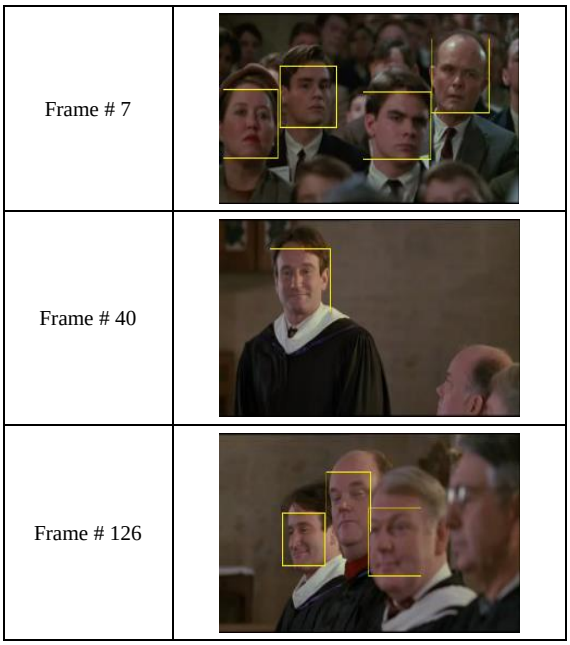


Fig.20. Frames of Dead_Poets_Society_00068.avi

Face 1

In the video sequence, face 1 appears in total of 12 frames (n) and it is detected in 9 frames (N). The tabulation is as follows:

Table 14. Statistics of face 1 in Dead_Poets_Society_00068.avi

Frame No	Width	Height	Area	SD	DR	DE
3	81	81	6561	0	1	0
7	81	81	6561	0	1	0
8	81	81	6561	0	1	0
9	81	81	6561	0	1	0
10	81	81	6561	0	1	0
11	81	81	6561	0	1	0
12	81	81	6561	0	1	0
13	81	81	6561	0	1	0
14	81	81	6561	0	1	0

Table 14 shows data starting from frame 3 from which face 1 enters into the scene and present till frame 14. The final overall result after the complete analysis of 12 frames is shown as a graph in Fig.21.

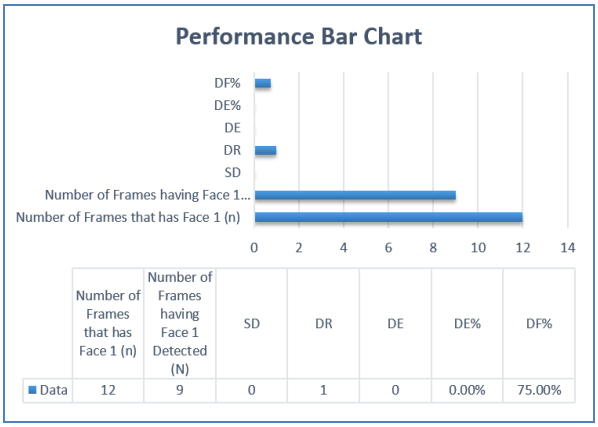


Fig.21. Bar chart of face 1 in Dead_Poets_Society_00068.avi

Face 2

In the video sequence, face 2 appears in total of 3 frames (n) and it is detected in 3 frames (N). The portion of the tabulation is as follows:

Table 15. Statistics of face 2 in Dead_Poets_Society_00068.avi

Frame No	Width	Height	Area	SD	DR	DE
2	90	90	8100	0	1	0
3	90	90	8100	0	1	0
4	90	90	8100	0	1	0

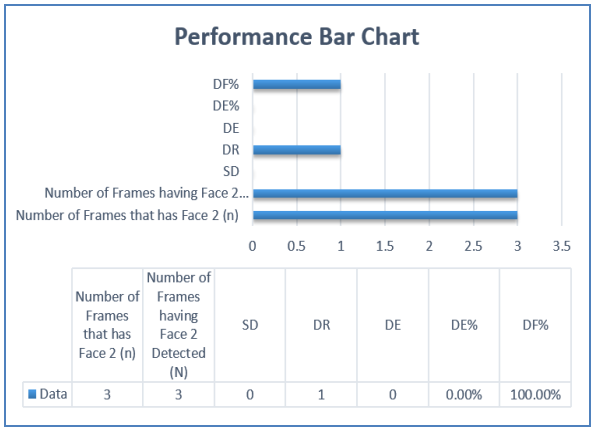


Fig.22. Bar chart of face 2 in Dead_Poets_Society_00068.avi

Table 15 shows data starting from frame 2 from which face 2 enters into the scene and present till frame 4. The final overall result after the complete analysis of 3 frames is shown as a graph in Fig.22.

Face 3

In the video sequence, face 3 appears in total of 4 frames (n) and it is detected in 4 frames (N). The tabulation is as follows:

Table 16. Statistics of face 3 in Dead_Poets_Society_00068.avi

Frame No	Width	Height	Area	SD	DR	DE
2	88	88	7744	0	1	0
3	88	88	7744	0	1	0
4	88	88	7744	0	1	0
5	88	88	7744	0	1	0

Table 16 shows data starting from frame 2 from which face 3 enters into the scene and present till frame 5. The final overall result after the complete analysis of 4 frames is shown as a graph in Fig.23.

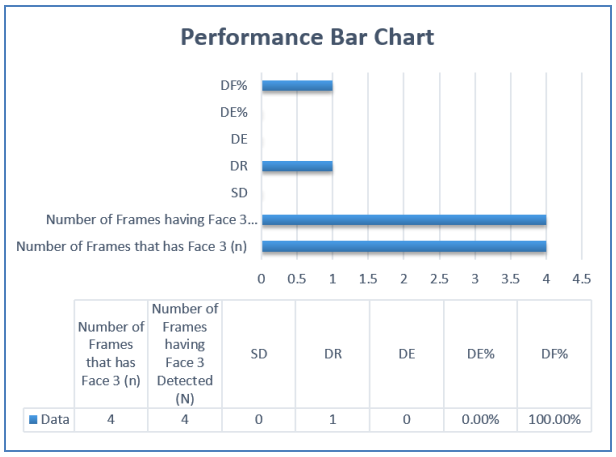


Fig.23. Bar chart of face 3 in Dead_Poets_Society_00068.avi

We can observe that out of 233 frames, we have calculated till 14 frames only; because we are not limiting frames in the video. As discussed before, we are more concerned with the number of faces over frames. Like previous set of videos we have limited to 3 faces and in this video we obtain all 3 faces within 14 frames itself. Hence, our analysis and tabulation stops at frame 14. We restricted to faces only, because analysis and tabulation of data for 233 frames is out of the scope of our work. Here, we are majorly focused over different set of videos like static camera with moving face(s), moving camera with moving face(s) and moving camera with static face(s). The first 2 types of videos have been analysed till now, next section revolves around analysis of 3rd type of video where camera is at motion but face is static.

G. Single Face Video Subjected to Camera in Motion and Static Object

The video sequence that is being taken for analysis in this section is from YouTube Celebrities dataset [22]. It comprises of a clipping where an actress is singing a song and she is in the position of rest; whereas camera is moving around her making it best suitable for our third constraint moving camera with static face(s). The video 0676_01_007_gloria_estefan.avi has 98 frames in total. Since the video is having only one face in it, we have carried out a detailed analysis and the results are tabulated in Table 17. Some of the pictures of 0676_01_007_gloria_estefan.avi where in the face is detected and tracked are shown in Fig.24.

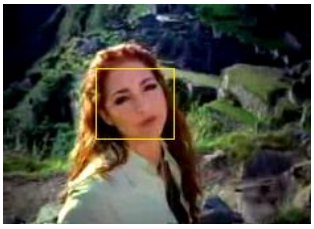
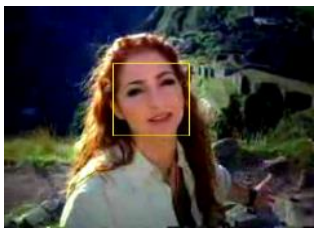
Frame # 36	
Frame # 57	
Frame # 85	

Fig.24. Frames of 0676_01_007_gloria_estefan.avi

Table 17. Statistics of 0676_01_007_gloria_estefan.avi

Frame No	Width	Height	Area	SD	DR	DE
20	86	86	7396	0	1	0
22	88	88	7744	-348	1	0
23	88	88	7744	0	1	0
24	88	88	7744	0	1	0
26	87	87	7569	175	1	1
28	89	89	7921	-352	1	0
30	88	88	7744	177	1	0
32	90	90	8100	-356	1	0
34	88	88	7744	356	1	1
36	89	89	7921	-177	1	0
38	90	90	8100	-179	1	1
39	90	90	8100	0	1	0
40	90	90	8100	0	1	0
41	90	90	8100	-704	1	0
42	90	90	8100	-704	1	0
43	90	90	8100	-704	1	0
44	90	90	8100	-704	1	0
45	90	90	8100	-704	1	0
46	90	90	8100	-704	1	0
47	90	90	8100	-704	1	0
48	90	90	8100	-704	1	0
49	90	90	8100	-704	1	0
50	90	90	8100	-704	1	0
51	90	90	8100	-704	1	0
52	90	90	8100	-704	1	0
53	90	90	8100	-704	1	0
54	90	90	8100	-704	1	0
55	90	90	8100	-704	1	0
56	90	90	8100	-704	1	0
57	90	90	8100	-704	1	0
58	90	90	8100	-704	1	0
59	90	90	8100	-704	1	0
60	90	90	8100	-704	1	0
61	90	90	8100	-704	1	0
62	90	90	8100	-704	1	0
63	90	90	8100	-704	1	0
64	90	90	8100	-704	1	0
65	90	90	8100	-704	1	0
66	90	90	8100	-704	1	0
67	90	90	8100	-704	1	0
68	90	90	8100	-704	1	0
69	90	90	8100	-704	1	0
70	90	90	8100	-704	1	0
71	90	90	8100	-704	1	0
72	90	90	8100	-704	1	0
73	90	90	8100	-704	1	0
74	90	90	8100	-704	1	0
75	90	90	8100	-704	1	0
76	90	90	8100	-704	1	0
77	90	90	8100	-704	1	0
78	90	90	8100	-704	1	0
79	90	90	8100	-704	1	0
80	90	90	8100	-704	1	0
81	90	90	8100	-704	1	0
82	90	90	8100	-704	1	0
83	90	90	8100	-704	1	0
84	90	90	8100	-704	1	0
85	90	90	8100	-704	1	0
86	90	90	8100	-704	1	0
87	90	90	8100	-704	1	0
88	90	90	8100	-704	1	0
89	90	90	8100	-704	1	0
90	90	90	8100	-704	1	0
91	90	90	8100	-704	1	0
92	90	90	8100	-704	1	0
93	90	90	8100	-704	1	0
94	90	90	8100	-704	1	0
95	90	90	8100	-704	1	0
96	90	90	8100	-704	1	0
97	90	90	8100	-704	1	0
98	90	90	8100	-704	1	0

Table 17 shows data starting from frame 20 and present till frame 98. The final overall result after the complete analysis of 79 frames is shown in Fig.25.

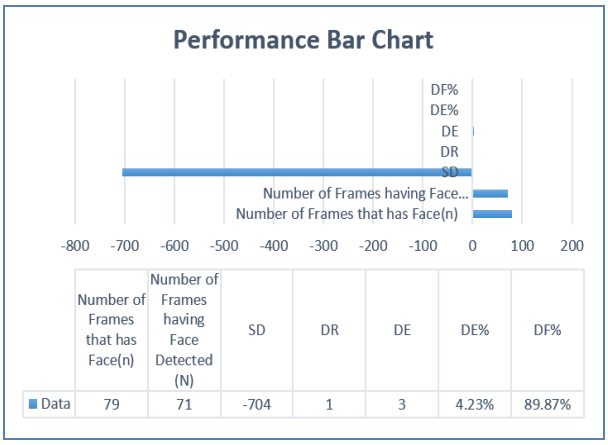


Fig.25. Bar chart of 0676_01_007_gloria_estefan.avi

VI. Conclusion and Discussion

The proposed work is capable of tracking face(s) in different background video sequences. In our proposed work, it is not mandatory for the face(s) to be present in the first frame itself. The proposed work starts with Viola-Jones algorithm and compute mean of the BRISK feature points extracted from the detected facial region(s). Further, it carries out similarity checks of detected face(s) with the threshold value extremely well. The above said process is repeated till the last frame of the video sequence.

Though there are many established metrics available, we have proposed our own metrics for assessing the attainment of the proposed algorithm. Video sequences from three different datasets are used for testing and analysis. The results obtained clearly show that the proposed algorithm's performance is better in almost every video sequence.

The video sequences being considered till now to carry out the computation and analysis are taken from particular datasets, which have certain challenges in them. The tabulations shown above are the practical analysis reports generated using 5th Gen Intel i7 processor core, we haven't considered computational time anywhere in our analysis. Computational timing is directly proportional to the hardware of the system under which the proposed algorithm executes. We even considered to evaluate the time constraint, but i7 processor takes considerable amount of time. Then in contradiction, lower end machines would take even some more time. It's not that the proposed algorithm lacks computational time efficiency; but enormous calculations taking place such as face detection, matching with existing data, adding new face entry followed by BRISK feature points detection and so on needs more computational power and processor cycles.

VII. Future Works

Future work involves avoiding false detection occurring at some places due to reflection and/or portraits that consists of face like images. This can be resolved using video stabilization which is beyond the scope of our work.

Along with the metrics used in the present work, future work on face detection and tracking can also be evaluated using recall, precision, accuracy, face tracking success rate and others.

In the future work, evaluation can be carried out on video sequences with different and more complex backgrounds from other challenging datasets.

References

[1] Ranganatha S and Dr. Y P Gowramma, "Face Recognition Techniques: A Survey", International Journal for Research in Applied Science and Engineering Technology (IJRASET), ISSN: 2321-9653, vol.3, iss.iv, pp.630-635, April 2015.

[2] R. Polana and R. Nelson, "Low Level Recognition of Human Motion", in Proc. of IEEE Workshop on Motion of Non-Rigid and Articulated Objects, Austin, TX, pp.77-82, 1994.

[3] P. Viola and M. Jones, "Rapid Object Detection Using a Boosted Cascade of Simple Features", in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Hawaii, USA, vol.1, pp.511-518, 2001.

[4] P. Viola and M. Jones, "Robust Real-Time Face Detection", International Journal of Computer Vision (IJCV), vol.57, pp.137-154, 2004.

[5] Stefan Leutenegger, Margarita Chli, and Roland Y. Siegwart, "BRISK: Binary Robust Invariant Scalable Keypoints", in Proc. of IEEE International Conference on Computer Vision, pp.2548-2555, 2011.

[6] Choi, Wongun, Caroline Pantofaru, and Silvio Savarese, "A General Framework for Tracking Multiple People from a Moving Camera", in IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), vol.35, iss.7, pp.1577-1591, 2013.

[7] C. Stauffer and W. Grimson, "Adaptive Background Mixture Models for Real-Time Tracking", in Proc. of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), vol.2, pp.246-252, 1999.

[8] A. J. Lipton, H. Fujiyoshi, and R. S. Patil, "Moving Target Classification and Tracking from Real-Time Video", in Proc. of 4th IEEE Workshop on Applications of Computer Vision, pp.8-14, 1998.

[9] Jie Xia, Jian Wu, Haitao Zhai, and Zhiming Cui, "Moving Vehicle Tracking Based on Double Difference and CAMSHIFT", in Proc. of International Symposium on Information Processing (ISIP), pp.029-032, 2009.

[10] J. Barron, D. Fleet, and S. Beauchemin, "Performance of Optical Flow Techniques", International Journal of Computer Vision (IJCV), vol.12, iss.1, pp.43-77, 1994.

[11] P. Viola and M. Jones, "Fast Multi-View Face Detection", Mitsubishi Electric Research Laboratories, TR2003-96, July 2003.

[12] Wilson, Philip Ian, and John Fernandez, "Facial Feature Detection Using Haar Classifiers", Journal of Computing Sciences in Colleges, vol.21, iss.4, pp.127-133, 2006.

[13] N. Dalal, B. Triggs, "Histograms of Oriented Gradients for Human Detection", in Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), June 2005.

[14] David G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision (IJCV), vol.60, iss.2, pp.91-110, November 2004.

[15] J. Wu and J. M. Rehg, "CENTRIST: A Visual Descriptor for Scene Categorization", in IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), vol.33, no.8, pp.1489-1501, August 2011, doi: 10.1109/TPAMI.2010.224.

[16] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool, "Speeded-Up Robust Features", Computer Vision and Image Understanding, vol.110, iss.3, pp.346-359, June 2008.

[17] Bruce D. Lucas and Takeo Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision", in Proc. of International Joint Conference on Artificial Intelligence, vol.2, pp.674-679, August 1981.

[18] Carlo Tomasi and Takeo Kanade, "Detection and Tracking of Point Features", Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

[19] Jianbo Shi and Carlo Tomasi, "Good Features to Track", in Proc. of IEEE Conference on Computer Vision and

Copyright © 2018 MECS

Pattern Recognition (CVPR), pp.593- 600, June 1994.
[20] Yadong Li, Ardeshir Goshtasby, and Oscar Garcia, "Detecting and Tracking Human Faces in Videos", in Proc. of IEEE Conference on Pattern Recognition, pp.807-810, September 2000.
[21] Ranganatha S and Y P Gowramma, "A Novel Fused Algorithm for Human Face Tracking in Video Sequences", in Proc. of IEEE International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), pp.1-6, October 2016.
[22] M. Kim, S. Kumar, V. Pavlovic, and H. Rowley, "Face Tracking and Recognition with Visual Constraints in Real-World Videos", in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2008.
[23] R. Chellappa, C. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey", in Proc. of IEEE, vol.83, iss.5, pp.705-741, May 1995.
[24] Yongkang Wong, Shaokang Chen, Sandra Mau, Conrad Sanderson, and Brian C. Lovell, "Patch-Based Probabilistic Image Quality Assessment for Face Selection and Improved Video-Based Face Recognition", in proc. of IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp.74-81, June 2011.
[25] Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld, "Learning Realistic Human Actions from Movies", in Proc. of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp.1-8, June 2008.
[26] Ranganatha S and Y P Gowramma, "An Integrated Robust Approach for Fast Face Tracking in Noisy Real-World Videos with Visual Constraints", in Proc. of IEEE International Conference on Advances in Computing, Communications and Informatics (ICACCI), pp.772-776, September 2017.
[27] G. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface", Intel Technology Journal, pp.12-21, 1998.
[28] J. Strom, T. Jebara, S. Basu, and A. Pentland, "Real-Time Tracking and Modeling of Faces: An EKF-based Analysis by Synthesis Approach", Technical Report 506, M.I.T. Media Laboratory Perceptual Computing Section, 1999.
[29] Douglas Decarlo and Dimitris N. Metaxas, "Optical Flow Constraints on Deformable Models with Applications to Face Tracking", International Journal of Computer Vision (IJCV), vol.38, no.2, pp.99-127, July 2000.
[30] G. Mallikarjuna Rao and Dr. Ch. Satyanarayana, "Object Tracking System Using Approximate Median Filter, Kalman Filter and Dynamic Template Matching", International Journal of Intelligent Systems and Applications (IJISA), vol.6, no.5, April 2014.
[31] Kamarul H. Ghazali, Jie Ma, and Rui Xiao, "Driver's Face Tracking Based on Improved CAMShift", International Journal of Image, Graphics and Signal Processing (IJIGSP), vol.5, no.1, pp.1-7, January 2013.
[32] Subrat Kumar Rath and Siddharth Swarup Rautaray, "A Survey on Face Detection and Recognition Techniques in Different Application Domain", International Journal of Modern Education and Computer Science (IJMECS), vol.6, no.8, pp.33-44, August 2014.

Authors' Profiles

Ranganatha S received B.E degree (Branch: CS & E, College: KIT, Tiptur, Karnataka, India) and M.Tech degree (Specialization: CS & E, College: RVCE, Bangalore, Karnataka, India) in the years 1998-2002 and 2004-2006 respectively from Visvesvaraya Technological University (VTU), Belagavi, Karnataka, India and currently pursuing his PhD in image processing at VTU Research Resource Centre. He has more than 15 years of teaching experience in reputed organizations at graduate and post graduate level. He is currently working as Assistant Professor, Department of Computer Science & Engineering, Government Engineering College, Hassan, Karnataka, India. He has published papers in various national and international journals and conferences. His current research interests include image processing and algorithms.

Dr. Y P Gowramma received B.E degree (Branch: CS &E, College: AIT, Chikmagalore, University: Mysore, Karnataka, India), M.Tech degree (Specialization: SACA, College: NITK, Karnataka, India) and PhD degree (University: VTU, Belagavi, Karnataka, India) in the years 1994, 1998 and 2014 respectively. She has more than 24 years of teaching experience in reputed organizations at graduate and post graduate level. She is currently working as Professor, Department of Computer Science & Engineering, Kalpataru Institute of Technology, Tiptur, Karnataka, India. She has published papers in various national and international journals and conferences. Her current research interests include image processing and algorithms.

How to cite this paper: Ranganatha S, Y P Gowramma, "Development of Robust Multiple Face Tracking Algorithm and Novel Performance Evaluation Metrics for Different Background Video Sequences", International Journal of Intelligent Systems and Applications(IJISA), Vol.10, No.8, pp.19-35, 2018. DOI: 10.5815/ijisa.2018.08.03

Copyright © 2018 MECS