International Journal of Advanced Technology & Engineering Research (IJATER)
www.ijater.com

A Study of Object Tracking Techniques

Sonal Jain, Madhav Institute of Technology and Science Gwalior, India; Prof. Rajeev Kumar Singh, Madhav Institute of Technology and Science Gwalior, India

Abstract

Face Tracking is an important & challenging research area in the field of computer vision & image processing. Tracking is the method to identify the position of a moving object in the video. Identifying the face is a much more challenging task than tracking the facial features in a moving video. Facial features that can be tracked include eyes, nose and mouth. This paper describes tracking of face and facial features in video surveillance systems. Although its main application is in security, face tracking alternatively is also used to improve the image and video quality.

Keywords:- Object Tracking, Face Tracking, Viola Jones Algorithm, and KLT Algorithm.

Introduction

Tracking can be defined as a method of measuring the movement of an object in the image plane as it moves around a scene [1]. Object tracking is of utmost importance in the field of computer vision with applications in video surveillance, human-machine interfaces, and robot perception. It involves the detection of readable moving objects in a frame of a video sequence and tracking of such objects in one after the other frames [2]. The tracking algorithm predicts the future positions of multiple moving objects according to the historical locations and the current vision features. Object tracking system has been applied in a many fields, such as the human-computer interaction, the security, and surveillance [3].

Nowadays, due to different security reasons, these surveillance systems have became more famous and crucial. A rapidly growing number of video based applications deems various objects to be detected, recognized and tracked in a specific scene in order to extract semantic information about scene activity and human behavior [4].Facial Features tracking is a crucial problem in computer vision due to its wide range of operation in Psychological facial expression analysis and human-computer interfaces. Today advances in face video processing and compression have made face-to-face communication to be practical in real-world applications. And after decades, robust and realistic real-time face tracking still poses a big challenge. The difficulty lies in a number of concerns including the real-time face feature tracking under a variety of imaging conditions (e.g., skin color, pose change, self-occlusion, and multiple non-rigid feature deformations). Discusses various developments and approaches in facial feature tracking [5]. Face detection is a computer technology that analyses human faces in digital pictures. Face detection may also refer to the mental process by which humans locate visual scenes in real life. Face detection is a specific case of class-object detection. In class-object detection, the main task is to look for the sizes and locations of all objects in a picture that is from a given class. Some examples include torsos, cars, and pedestrians [6]. Some modern applications where face detection is being used are autofocus cameras, practical security systems, people counting system, lecture attendance system etc.

Area-Based Feature Detection or Tracking

A. Eye tracking: - Eye-tracking techniques that currently exist rely on calculating electrical potentials generated by the moving eye (electro-oculography) or a metal coil in a magnetic field. Such methods are relatively cumbersome and uncomfortable for the subject. A new generation of eye trackers is now available based on the invoice recording of images of a subject's eye using infrared sensitive video technology and relying on software processing to figure out the position of the subject's eyes relative to the head [7].

(1) Eyelink Gaze tracker: - The Eye link Gaze tracker (SR Research Ltd., Mississauga, Ontario, Canada) is one of these video-based eye trackers and is used in such research fields such as psychology, ophthalmology, neurology, and ergonomics. Figure.1 represents two high-speed cameras (CCD sensors) to track both eyes together. A third camera (also a CCD sensor) tracks four infrared markers mounted on the visual incentive display, so that head motion can be calculated and gaze position can be computed. The cameras produce images at a sampling rate of 250 Hz (4- msec temporal resolution). The Eye link is used with a PC with dedicated hardware for doing the image processing necessary to resolve gaze position [8].

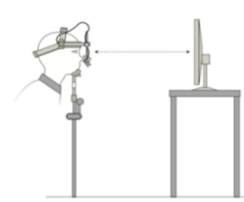

Figure.1 Eye Gazing [8]

This phenomenon is discussed in the literature as follows: -

Ahmad Delforouzi etl. propose a solution for a special challenge. Handling a variety of challenging situation of object tracking in 360-degree videos is still an unsolved problem and needs to be more considered proposed, a new structure of SURF-based object tracking is proposed which uses a train-based matching to address the challenging object tracking in 360-degree videos. The proposed tracker is able to evaluate out-of-plane rotation and occlusion during the tracking and adapt itself to handle it [9].

Liu Wancun Tang Wenyan etl. proposed a novel multi-scale behavior learning approach to analysis the motion pattern of object's location and size. The multi-object tracking problem is formulated by a generative probability model, which contains a global and a local learning procedure. The global behavior is offline learned by a neural network, while the local behavior is learned by online learning. With these combinations, the accurate prediction is made due to robust multiscale features. Experimental results validate that the proposed method reduces the IDS significantly and improves the performance considerably [10].

Diego Ayala etl. proposed the development of an integrated vision system for detection, location, and tracking of a color object, it makes use of a microprocessor to get image data, process, and perform actions according to the interpreted data. The deterring effects of uncontrolled illumination are greatly diminished by use of the rg chromaticity space enabling this system to detect, locate and track a colored object satisfactorily, while being low-cost, compact and energy-saving[11].

Weisheng Li etl. Motion vectors extracted from a compressed video file can be used to track objects in the video (Motion vectors are determined during the motion estimation step when encoding video by matching similar blocks between successive frames) and it could be efficient as motion vectors provide trajectory information of the objects. The proposed tracking multiple objects can be dynamic by using motion vectors extracted from compressed video. The system applies statistical and clustering techniques on motion vectors to track multiple objects in real video [12].

Zhangping he etl. proposed fast Fourier transform networks for object tracking called FFTNet. FFTNet for object tracking based on CF (Correlation filter). We have taken full advantage of CF, because CF has high computational performance and competitive performance, into FFTNet. FFTNet is based on deep neural networks so that it has strong capabilities of learning feature representation and matching function. FFTNet is trained end-to-end in an off-line manner instead of online fine-tuning. FFTNet is more robust to short-term occlusion, fast motion and scale variation than state-of-the-art real-time trackers [13].

Zhu Teng, Junliang Xing etl. in this they bring together the combination of the Feature Net(Feature Net is the backbone network of deep architecture to extract low-level features and does not require off-line training or online fine-tuning), Temporal Net(Temporal Net encodes the trajectory of the target and directly learns temporal correspondences to estimate the object state from a global perspective) and the Spatial Net(spatial and local information is also very important, particularly for a deep network based object tracking, since spatial information is attenuated as a deep network goes deeper) which is effective in object tracking and it is estimated on four benchmarks [14].

Yoanes Bandung etl. proposed an assimilation of object tracking technology within video conference system. This assimilation aims to provide better captured video content which can be automatically focused on key objects or individuals in a learning activity such as whiteboard, teacher or students. This system can disqualify the need of camera operator and improve quality of distance learning service. Video conference application is based on Web RTC (**Web RTC** is an open framework for the web that enables Real Time Communications in the browser. It includes the fundamental building blocks for high-quality communications on the web) Technology and Designed to support multiple video input devices**.** [15].

B.Maga etl. resolves the various methods in static and moving object detection as well as tracking of moving objects. A new proposed way is provided for efficient object tracking using Kernel and feature based tracking methods. Object detecting and tracking has a wide variety of applications in computer vision video surveillance, its gives more accuracy than other existing method like CNN, template matching. Kernel and feature based methods works better for detection in multiple objects [16].

Sun Xiaoyan etl. proposed an adapted particle filter tracker with online learning and inheriting selective model Feature-learning (feature learning or representation learning is a set of techniques that allows a system to automatically discover the representations needed for feature detection or classification from raw data) and feature inheriting help particle filter (Particle filters or Sequential Monte Carlo (SMC) methods are a set of genetic, Monte Carlo algorithms used to solve filtering problems arising in signal processing) improve the efficiency and robustness of tracking. This method can track the target quickly and accurately [17].

Wei Han, Guang-Bin Huang etl. proposed a graph learning-based tracking (graph-based trackers which construct the graph with the distance matrix, our method adaptively learns a graph of local parts of two consecutive frames to reveal the correspondence between them.) framework to handle object deformation occlusion in this deform show that this method can improve tracking robustness under large deformation and occlusion and better the state-of-the-art algorithms. In this paper algorithm optimizes the graph similarity matrix until two disconnected sub graphs divide the foreground and background nodes [18].

Wenming cao etl. proposed a short-term tracking method, which is more robust than ordinary methods for single-target tracking under occlusions. The weighted-based key points matching tracker, we employ fast key points detection and partial descriptors, which makes it run in real time. They estimate weight-based key points matching tracker for occlusion, in which we apply the geometrical similarity to supplement virtual key points and fuzzy logic to estimate the occluded degree [19].

Rani Aishwarya S N etl. proposed novel approach where the KCF (Kernalized Correlation Filters) (KCF is widely adopted in tracking algorithm because of its high speed and performance)filter is built-up by integrating it with Kalman filter. The integrated Kalman based KCF (KKCF) tracker is better than traditional KCF by performing well for outlier or failure cases which is corrected through Kalman filter. The main aim was to track moving objects more exactly and faster when compare the other approaches [20].

Miaobin Cen etl. proposed multiple form of local orientation plane (Comp-LOP) for object tracking. Comp-LOP is a simple but effective descriptor, which is robust to occlusion for object tracking. It effectively considers spatiotemporal relationship between the target and its surrounding regions in a correlation filter framework by the complex form This proposed approach outperforms state-of-the-art tracker on large benchmark data sets. This method bring outs good performance in object tracking [21].

Xuebai Zhang etl. proposed three key advertising elements (product, brand and endorser) were tracked, presented by three eye movement indicators transformed fixation time (TFT is the ratio between fixation time on AOIs and the AOIs display time), transformed fixation number (TFN is the ratio between fixation number on AOIs and the AOIs display time), and average gaze duration (AGD indicate the intensity of cognitive processing and AGD was calculated by dividing the participants' fixation time by their fixation number on the AOI (the fixation time per fixation on the AOIs). The unit of measurement used was the second. The indicated that three items are related to attitude toward ad, attitude toward brand (brand-related TFN and AGD, endorser-related TFT), and purchase intention (product related AGD, brand-related TFN and endorser-related TFN). However, only two items of them are related to recall (product-related AGD and brand-related TFN) The data were access from 61 participants, each stimulated by six video ads, via an eye tracking method and questionnaires [22].

Brian Guenter etl. present a novel, automatic eye gaze tracking scheme inspirited by smooth pursuit eye motion while playing mobile games or watching virtual reality contents. Our algorithm constantly improves an eye tracking system for a head mounted display. The algorithm comparison between corneal motion and screen space motion, and uses these to achieve Gaussian Process Regression models. A combination of those models provides a constantly mapping from corneal position to screen space position. Accuracy is nearly as good as generate with an accurate calibration step [23].

Ramona-georgiana vanghele etl. disussed target only on a small part of this research area, namely the gaze detection (i.e. This "gaze detection" system is especially sensitive to whether someone's looking directly at you (for example, whether someone's staring at you or at the clock just over your shoulder)), using the properties of images and video sequences such as brightness, contrast, RGB colors representation. The method used techniques of contour sharpening and selecting elements in the image [24].

Dionis A. Padilla etl. implementing an eye tracking system on a Field Programmable Gate Array(semiconductor devices that are based around a matrix of configurable logic blocks connected via programmable interconnects) and specifically a text-typing application. Using state machines implemented in Verilog – a hardware description language and MATLAB for verification purposes, the said eye tracking system was built._The study comprises of several algorithms such as comprising of several processes such as Thinning Algorithm and Hough Transform was accomplished [25].

Ashwani K. Thakur etl. discussed approaches to make our eye to control the cursor as an application to be used as what one can called as virtual mouse and can be used to perform many more applications by using viola Jones algorithm. This system provides fast and real time results with the accuracy of eye tracking were found to be approximately one degree of visual angle. It is specially designed for the Handicap people to use Computer or for controlling the wheelchair. Eye tracking hence can make a revolution in the everyday life of people. [26].

B. Nose tracking: - The nose feature is defined as the point on the nose surface that is the closest to the camera. This point is termed the tip of the nose [27]. Due to the symmetry and the convex shape of the nose, the nose feature is always visible in the camera, and it stays almost the same during the rotations of the head. It also does not change much with head moving towards and from the camera. Thus, the nose tip defined above can always be located. This is very an important property of the nose which does not hold for any other facial feature.

Shadman Sakib Khan etl. Human Computer Interface (HCI) is implemented with digital image processing and a new method to control personal computers with high efficiency. The main three characteristics of this interface are nose tracking cursor control, auto brightness control and display control based on the presence and detection of valid human face. The proposed system is low cost and display inherent security and power saving capabilities. The primary contribution of this paper is proposing and implementing a low-cost human computer interface using webcam which provides an easy way for disabled people to seamlessly operate personal computers [28].

Weiwei Zhang etl. proposed a yawning detection system that consists of a face detector, a nose detector, a nose tracker and a yawning detector. Deep learning algorithms are refining for detecting driver face area and nose location. A nose tracking algorithm that combines Kalman filter with a dedicate d open-source TLD (Track-Learning-Detection) tracker is developed to achieve robust tracking results under dynamic driving conditions. Experiments are conducted on real-world driving data, and results show that the deep convolutional networks can generate a satisfactory classification result for detecting driver's face and nose when compared with other pattern classification methods, and the proposed yawning detection system is effective in real-time detection of driver's yawning states. [29].

Martin B¨ohme etl. proposed a facial feature detector for time-off light (TOF) cameras that last previous work by combining a nose detector based on geometric charters with a face detector. The goal is to prevent false detections outside the area of the face. This approach to face detection was originally described by Viola and Jones and has since gained enormous popularity for a wide range of applications. They used a very simple classifier based on an axis-aligned bounding box in feature space; pixels whose feature values fall within the box are classified as nose pixels, and all other pixels are classified as "non-nose" [30].

S. Waphareetl. proposed the implementation of 2 novel algorithms named Surge-spiralx and Surge-castx on sniffer robot for odor plume tracking (The e-nose takes in samples of the environment air, where the sensors translate the odor concentration into electrical signals which are then processed to identify the odor.) in a laminar wind environment. The algorithms are developed and they have shown very good performance in terms of success ratio, while Surge-Spiralx algorithm having less distance overhead [31].

Luca Cappelletta etl. a semi-automatic system based on nostril detection is presented. The system is creating to work on ordinary front videos and to be able to recover brief nostril occlusion. Using the nostril position a motion compensated Accumulated Difference Image (ADI) is generated. This ADI is less noisy than the non-compensated one, and this leads to better mouth region tracking The VidTIMIT dataset is used and comprises of the video [37].

C. Mouth tracking: -The term mouth tracking is used to include lip tracking as the lips are a component of the mouth which contains other vital cues describing the mouth (i.e. tongue, teeth, oral cavity). The lip show ever, act as an invaluable feature for tracking the mouth as in many cases the labial area gives a very good line of demarcation between the mouth and the face background [32].

Chris Fortuna etl. proposed HTM algorithm which is designed to run continuously, actively counting bites throughout the day. They take a novel machine learning approach to customize the system to each individual user, and generate an average accuracy of 91.8%, well above the current state-of-the-art. They used us small set of 5 motion features, a Naive Bayes model. Our algorithm represents the first viable solution to counting bites. It is lightweight enough to be put in a smart device, and accurate enough to stream data 24 hours a day [33].

Sunil S. Moradeetl. proposed a novel active contour guided geometrical feature extraction approach for lip reading. Three active contour approaches are snake, region scalable fitting energy method and localized active contour model. These approaches are adopted for salient geometrical feature calculation. The experimentations were carried out for lip feature extractions which are useful for visual speech recognition. A joint feature model, obtained by combining inner area, height and width has been proposed [34].

Jie Cheng etl. proposed a novel approach for real-time mouth tracking and 3D reconstruction. This method comprises two successive processing stages. In the first stage, an AdaBoost learning algorithm and a Kalman filter are used to detect and track the mouth region in real-time under a complex background. In the second stage, the resultant 2D position of the mouth is used to determine the region where the 3D shape is reconstructed by use of a digital fringe projection and modified Fourier transform method. The main grant of this paper is the real-time dense 3D reconstruction of the mouth region, which can be useful in many applications, such as lip-reading, biometrics, 3D animation, etc [35].

Zhilin Wu etl. proposed approach consisting of a Gradient Vector Flow snake with a parabolic template as an additional external force is proposed. Based on the results of the outer lip tracking, the inner lip is tracked using a similarity function and a temporal smoothness constraint. The advantage is that the final tracking results depend on both the Gradient Vector Flow (GVF) snake field vector and the template, appropriately weighted in terms of their qualities. [36].

Algorithm Assessment

A. Viola Jones Algorithm

Viola-Jones Algorithm [38, 41] is based on the analysis of the input image by means of sub window capable of detecting features. This window is extending to detect faces of distinct sizes in the image. Viola-Jones developed a scale invariant detector which runs through the image many times, each time with particular size. Being scale-invariant, the detector requires the same number of calculations regardless of the size of the image. The system architecture [41] of Viola-Jones is based on a cascade of detectors. The first stage consists of simple detectors which ignore only those windows which do not contain faces.

Viola-Jones face detection algorithm searches the detector several times through the same image – each time with a new size [41]. The detector detects the non face area in an image and discards that area which results in detection of the face area. To reject non face area, Viola Jones takes advantage of cascading. The Viola–Jones algorithm is determined for real– time detection of faces from an image. Its real – time performance is collected by using Haar type features, computed rapidly by using integral images, feature selection using the Ada Boost algorithm (Adaptive Boost) and face detection with conditional cascade.

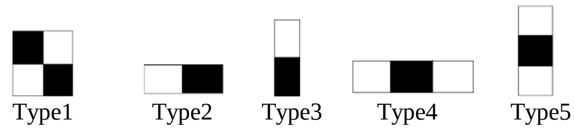

Figure.2 Haar masks used [38]

(1) Feature calculation

Starting from the common features of the faces, such as the region around the eyes, it can be seen that it is darker than the cheeks or the region of the nose is brighter than those of the eyes. For the same, five Haar masks were chosen for estimating the features, measured at different positions and sizes. Haar features are calculated as the difference between [38] the sum of the pixels from the white and the black region. In this way, it is possible to detect contrast differences.

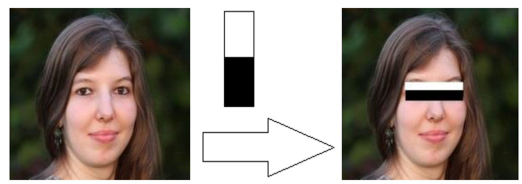

Figure.3 Type 2 Haar features from which the intensity difference between the pixels from the eyes region and the cheek region can be observed [38]

If the mask M from figure.3 Haar feature is considered, the Haar feature correlated with the image I behind the mask is defined by:

$$\sum_{1\leq i\leq N}\sum_{1\leq j\leq N} I(i,j)_{white} - I(i,j)_{black} \qquad (1)$$

The characteristic is extracted for windows with the dimensions of 24x24 pixels, which are moved on the image where faces are to be detected faces [38]. To reduce the computation time of the Haar features, which vary depending on the size and type of the feature, the integral image was used. In the figure is illustrated how an original image is obtained from the integral one and how it computes the sum of pixels within a rectangle region using the integral image. Computation from the region D using integral image.

For the location (*i*, *j*), the integral image *II* contains the sum of the pixels above and to the left of (*i*, *j*), inclusive:

$$II(i,j) = \sum_{1\leq s\leq i}\sum_{1\leq t\leq j} I(s,t)\,, 1 \leq i \leq N, 1N \qquad (2)$$

The sum of the pixels within rectangle can be measured with four array references. The value of the integral image at location 1 is the sum of the pixels in rectangle A. The value at location 2 is A + C, at location 3 is A + B, and at location

4 is A + B + C + D. The sum within D can be computed as 4 + 1 – (2 + 3).

(2) Feature selection using Ada Boost algorithm

As the number of Haar features for an image with 24 x 24 pixels is *d* = 162 336, and many of our redundant, Ada Boost algorithm [43] was used to select a smaller number of features. The basic idea is to build a complex classifier (decision rule) using a weighted linear combination of weak classifiers. Every feature f is considered a weak classifier, defined by:

$$h(x, f, p, \theta) = \begin{cases} 1, \text{if} pf(x) < p\theta \\ 0, \text{otherwise} \end{cases} \quad (3)$$

Where *x* is a 24 x 24 pixel image, Θ is a threshold and *p* is a parity.

B. Kanade LucasTomasi(KLT) Algorithm

Kanade Lucas Tomasi algorithm [38] used for feature tracking is certainly a most important one. KLT algorithm was introduced by Lucas and Kanade and their work was later extended by Tomasi and Kanade. This algorithm is used for detecting scattered feature points which have enough texture for tracking the required points in a good standard. Kanade-Lucas-Tomasi (KLT) algorithm is used here for tracking human faces constantly in a video frame. This method is accomplished by them finding the parameters that allow the reduction in dissimilarity measurements between feature points that are related to the original translational model.

Once the face has been detected it's the nature of the system to track that detected face. For tracking of the detected face by using Viola-Jones algorithm, an another and a most widely used algorithm known as Kanade-Lucas-Tomasi (KLT) algorithm was used. Kanade, Lucas, and Tomasi [38] paved the way for face tracking by introducing KLT algorithm which tracks face based on feature points of the detected face. Viola Jones can track the face but it is computationally costly. Viola Jones also cannot track the tilted face or if the face is turned.

Once the face detector locates the face, KLT tracks the feature points in the video frames which ensure exact tracking of the face. The working flow of KLT is in the order that it first detects the face, after which it identifies the facial features for tracking, then it initializes tracker for tracking those points and finally it tracks the face. The points can be lost due to lighting conditions, out of the plane rotation or articulated motion. To track points for the long period of time points need to be regained periodically.

Cascade object detector [40, 42] locates the face in the video frame. Trained classification model and Viola Jones algorithm is by default used by cascade object detector to detect faces in the input video frame. The input given to face detector is a video frame. The face detector [17] reads the video frame and runs the Viola-Jones face detection algorithm and plots the bounding box values around the face. Bounding box contains four values the x coordinate, y coordinate, length, and breadth. A rectangle is then used to plot these points around the face. The bounding box may contain different values for different shapes. According to shape, one might use the values to plot around the object.

Now further it tracks the feature points. Points=detectMinEigenFeature(rgb2gray(videoFrame),'ROI', box),converts the given video frame into grey scale image and gives this to detectMinEigen Features function. DetectMinEigen. Features function uses minimum Eigen value algorithm to find feature points. Then it shows these points in the input video frame and plots those detected points on the face. A point is an object which contains information about feature points detected in 2D grey scale input image. Then it initializes the tracker to track the points. Vision .Point Tracker uses KLT to track feature points.

The Iterative Lucas- Kanade Scheme:-

- **Location of the point on image**

  For $L = 0,\ldots,L_m$, define $u^L = [u_x^L \quad u_y^L]$, the corresponding coordinates of the point u on the pyramidal images I^L

  $$u^L = u/2^L$$

- **Spatial gradient matrix**

  $$G = \sum \begin{bmatrix} I_x^2 & I_x & I_y \\ I_x & I_y & I_y^2 \end{bmatrix}$$

  Where,

  I_x = Derivative of I_L with respect to x
  I_y = Derivative of I_L with respect to y
  G = spatial gradient matrix

- **Standard Lucas – Kanade scheme for flow computation at the level**

  L in the pyramid the goal is finding the vector d^L

L d^L

Finally, d = $d^0 + g^0$

- **Guess for next pyramid level L-1**

V = U + d

$g^{L+1} = 2(g^L + d^L)$

Year	Author	Method	Application area
2018	Xuebai Zhang Shyan-Ming Yua [22]	TFN, AGD, TFT, Dynamic	Video Advertising.
2018	Neilay Khasnabish Ketan P. Detroja [47]	Particle filter	Visual object , satisfactorily different scenarios with less no. of particles.
2017	Yoanes Bandung Kusprasapta Mutijarsa Luki B. Subekti [15]	TLD (Tracking learning Detection)	Video conference.
2017	Kang Wang QiangJi [48]	3D model-based gaze estimation methods	Natural head method and real- time eye gaze.
2017	Ahmad Delforouzi, Marcin Grzegorzek [9]	SURF-based object tracking	360 degree video.
2017	Weisheng Li, David Powers [12]	Mode reduction, K means	Multiple object tracking.
2017	Alexandru Pasarica, Radu Gabriel Bozomitu HaritonCostin, Casian Miron, CristianRotariu [49]	Circular Hough transform	Neuro-motor disabilities.
2016	Hsiau Wen Lin, Yi-Hong Lin [50]	Haar-likes features	Locating the position of each eye ball.
2016	Anjith George, Aurobinda Routray [51]	RANSAC (Random sample Consensus algorithm)	Inner eye corner.
2015	Zhang Naizhong Wen Jing, Wang Jun [52]	YCbCr color space	Hand free mouse.
2010	S. Waphare, D. Gharpure, A.Shaligram,	Surge-spiralx and Surge-castx algorithms	3 nose strategy

	B. Botre		
2010	Luca Cappelletta, Naomi Harte [37]	Kanade Lucas Tomasi (KLT) Tracker	The combination of a motion compensated ADI and nostril tracking gives a smooth track of the Speakers' mouth.

Applications

The method is based on face geometrical configuration. A face contains eyebrows, eyes, nose, and mouth; a face image is symmetric in the left and right directions; eyes are below two eyebrows; nose lies between and below two eyes; lips lie below nose; the contour of a human head can be a nearby ellipse, and so on. By using the facial components as well as the positional relationship between them the faces can be located easily [51].These real-time tracking techniques can be used to build non-intrusive vision-based user interfaces. Using the described tracking techniques, a system was built that estimates a user's head pose, and obtains the visual input for an eye-gaze tracker and a lip-reading system. These applications will be described below:

A. Head Pose Tracking: - A person's gaze direction is determined by two factors: the orientation of the head and the orientation of the eyes. Whereas the eye orientations determine the exact direction of the user's gaze, the head orientation [51] determines the overall gaze direction. Since the geometry of a face is known, estimating the orientation of the head is a pose estimation problem. In fact, the head pose can be estimated by finding correspondences between a number of head model points and their locations in the camera image. A system to generate the head pose using a full perspective model was developed. The system tracks six non-coplanar facial features in real-time.

B. Eye-Gaze Monitoring:-. The eye gaze monitoring system is a communication system which is very useful for the blind persons with the help of which they can perform their daily activities by using such as monitoring systems. Eye gaze System [52] is a direct-select vision controlled communication and control system. Eye gaze Systems are being used in homes, offices, schools, hospitals, and long-term care facilities. By looking at control keys displayed on a screen, a person can synthesize speech, control his environment (lights, appliances, etc.), type, operate a telephone, run computer software, operate a computer mouse and access the internet and e-mail.

C. Lip Reading: - It has been established that visual information can enhance the accuracy of speech recognition for both a human and a computer. However, many other lip-reading systems require a user to keep still or put special marks on his/her face. [53]

Conclusion

The detection and tracking of faces & facial features in video surveillance is a fundamental and challenging aspect in computer vision. This paper, describes object tracking of facial features such as eyes, nose & mouth in the video surveillance. It also includes the comparative study of facial features and its application in various fields. Future work includes pose estimation and 2D head tracking using the motion of the face with subsequent video frames. Furthermore, the 2D head tracking can be utilized to explore the 3D aspects of head tracking in future. The main challenge is to retain the robustness while adapting to changing environmental conditions, facial expressions and occlusions.

References

[1] Himani S.Parekh, G. Thakore, Udesang K. Jaliya, "A Survey on Object Detection and Tracking Methods, "International Journal of Innovative Research in Computer Communication Engineering, 2014.

[2] Monisha R, Muthuselvam M , "Feature Based Moving Object Detection And Tracking", International Conference On Innovative Trends In Engineering And Technology, 2017.

[3] Xiaobai Chen, Jinlong Xu, Zhiyi Yu, "A fast and energy efficient FPGA-based system for real-time object tracking", APSIPA, 2017.

[4] T. Mahalingam, M. Subramoniam, "A robust single and multiple moving object detection, Tracking and classification" ,IEEE, 2018

[5] Divya George and Arunkant A. Jose, "Face detection and tracking at different angles in video using optical flow", ARPN Journal of Engineering and Applied Sciences, 2015.

[6] Anjali Vats, R.K Bhatnagar, Shikha Vats, Antara Sarkar and Ankit, "Facial detection system(using

MATLAB)", International Journal Of Engineering And Computer Science, 2014.
[7] Frans W. Cornelissen And Enno M. Peters, "The Eyelink Toolbox: Eye tracking with MATLAB and the Psychophysics Toolbox", Behavior Research Methods, Instruments, & Computers, 2002.
[8] Frans W. Cornelissen, John J. van den Dobbelsteen, "Heading detection with simulated visual field defects" ,Visual Impairment Research,1999.
[9] Ahmad Delforouzi and Marcin Grzegorzek, "Robust and Fast Object Tracking for Challenging 360-degree Videos", International Symposium on Multimedia, 2017.
[10] Liu Wancun, Tang Wenyan, Zhang Liguo, Zhang Xiaolin, Li Jiafu, "Multi-Scale Behavior Learning for Multi-Object Tracking", IEEE, 2017.
[11] Diego Ayala, Danilo Chavez,"Low Cost Embedded Vision System For Location And Tracking of a color object",IEEE,2017.
[12] Weisheng Li, David Powers, "Multiple object tracking using motion vectors from compressed video", IEEE,2017.
[13] Zhangping He, Zhendong Zhang, Cheolkon Jung, "Fast Fourier Transform Networks for Object Tracking Based on Correlation Filter", IEEE, 2017.
[14] Zhu Teng, Junliang Xing, Qiang Wang, Congyan Lang, Songhe Feng, Yi Jin, "Robust Object Tracking based on Temporal and Spatial Deep Networks", International Conference on Computer Vision , 2017.
[15] Yoanes Bandung, Kusprasapta Mutijarsa, Luki B. Subekti, "Design and Implementation of Video Conference System with Object Tracking for Distance Learning", International Symposium on Electronics and Smart Devices , 2017.
[16] B.Maga, Mr. K.Jayasakthi Velmurgan, "An Efficient Approach for Object Detection And tracking", Third International Conference on Science Technology Engineering & Management, 2017.
[17] Sun Xiaoyan and Chang Faliang, "Moving Object Tracking with Feature Learning and Management, IEEE, 2017.
[18] Wei Han, Guang-Bin Huang, Dongshun Cui, "Deformable and Occluded Object Tracking via Graph Learning", IEEE , 2017.
[19] Wenming Cao, Yuhong Li, Zhihai He, Guitao Cao, And Zhiquan He", Supplementary Virtual Key points of Weight-Based Correspondences for Occluded Object Tracking", IEEE,2017.
[20] Rani Aishwarya S N, Vivek Maik,Chithravathi B "Robust Object Tracking Using Kernalized Correlation Filters (KCF) and Kalman Predictive Estimates", IEEE, 2017.
[21] Miaobin Cen and Cheolkon Jung, "Comp-Lop: Complex Form Of Local Orientation Plane For Object Tracking", IEEE, 2017.
[22] Xuebai Zhang, Shyan-Ming Yua, "An Eye Tracking Analysis for Video Advertising Relationship between Advertisement Elements and Effectiveness", IEEE, 2018.
[23] Subarna Tripathi and Brian Guenter, "A Statistical Approach to Continuous Self-Calibrating Eye Gaze Tracking for Head- Mounted Virtual Reality Systems", IEEE , 2017.
[24] Ramona-Georgiana Vanghele And Dumitru Stanomir, "Gaze detection / tracking", Journal of Engineering Sciences and Innovation, 2017.
[25] Dionis A. Padilla, Joseph Aaron B. Adriano, Jessie R. Balbin, "Implementation of Eye Gaze Tracking Technique on FPGA-based On-screen Keyboard System Using Verilog and MATLAB", IEEE, 2017.
[26] Ashwani K. Thakur, Tejan Gunjal, Aditya Jawalkar, Aparna More, "Eye Tracking Cursor", International Journal of Informative & Futuristic Research ,2017.
[27] Dmitry O. Gorodnic", On Importance of Nose for Face Tracking", IEEE, 2002.
[28] Shadman Sakib Khan, Md. Samiul Haque Sunny, M. Shifat Hossain, Eklas Hossain, "Nose Tracking Cursor Control for the People with Disabilities: An Improved HCI"(EICT) ,2017.
[29] Weiwei Zhang, Yi L.Murphey, Tianyu Wang, Qijie Xu, "Driver Yawning Detection based on Deep Convolutional Neural Learning and Robust Nose Tracking", IEEE, 2015.
[30] Martin Bohme, Martin Haker, Thomas Martinetz, and Erhardt Barth, "Head Tracking With Combined Face And Nose Detection", IEEE, 2009.
[31] S. Waphare, D. Gharpure and A.Shaligram, B. Botre, "Implementation Of 3-Nose Strategy In Odor Plume Tracking Algorithm", IEEE , 2010.
[32] Simon Lucey, Sridha Sridharan and Vinod Chandran, "A Suitability Metric For Mouth Tracking Through Chromatic Segmentation", IEEE, 2001.
[33] Chris Fortuna, Christophe Giraud-Carrier and Joshua West, "Hand-to-Mouth Motion Tracking in Free-Living Conditions for Improved Weight Control", International Conference on Healthcare Informatics, 2016.
[34] Sunil S. Morade and B. Suprava Patnaik, "Automatic Lip Tracking and Extraction of Lip Geometric Features for Lip Reading", International Journal of Machine Learning and Computing, 2013.
[35] Jie Cheng, Peisen Huang, "Real-Time Mouth Tracking and 3D Reconstruction,"3rd International Congress on Image and Signal Processing, 2010.

[36] Zhilin Wu, Petar S. Aleksic, and Aggelos K. Katsaggelos, "Lip Tracking for MPEG-4 Facial Animation", IEEE, 2002.
[37] Luca Cappelletta and Naomi Harte, "Nostril Detection for Robust Mouth Tracking", ISSC, 2010.
[38] Elena Alionte and Corneliu Lazar, "A Practical Implementation of Face Detection by Using Matlab Cascade Object Detector", International Conference on System Theory, Control and Computing, 2015.
[39] Cha Zhang, Zhengyou Zhang, “Boosting-based face detection and adaptation”, Synthesis Lectures on Computer Vision, 2010.
[40] P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features”, Proceeding of International Conference on Computer Vision and Pattern Recognition, 2001.
[41] Y. Wang, “An analysis of the Viola – Jones face detection algorithm”, Image Processing on Line, 2014
[42] http://www.mathworks.com/help/vision/ug/train-a-cascade-objectdetector.
[43] R. E. Schapire, “A brief introduction to boosting”, Proc. of the Sixteenth International Joint Conference on Artificial Intelligence, 1999.
[44] Jie Yang, Rainer Stiefel hagen, Uwe Meier, Alex Waibel, "Real-time Face and Facial Feature Tracking and Applications", Interactive Systems Laboratory Carnegie Mellon University, 1998.
[45] Akshay Govind Wanjari, and Swapnil S. Khode, "The Eye gaze Communication System", International Journal of Research Studies in Science, Engineering and Technology, 2014.
[46] Jie Cheng, and Peisen Huang, "Real-Time Mouth Tracking and 3D Reconstruction", 3rd International Congress on Image and Signal Processing, 2010.
[47] Neilay Khasnabish and Ketan P. Detroja, “A Stochastic Resampling Based Selective Particle Filter for Visual Object Tracking”, Indian Control Conference, 2018.
[48] Kang Wang QiangJi, “Real Time Eye Gaze Tracking with 3D Deformable Eye-Face Model” ,IEEE, 2017.
[49] Alexandru Pasarica, Radu Gabriel Bozomitu , Hariton Costin, CasianMiron, Cristi Rotariu,“Human-Computer Interface based on Eye Tracking with Dwell Time Selection”, IEEE, 2017.
[50] Hsiau Wen Lin, Yi-Hong Lin,”Face detection based on the use of eyes tracking”, International Computer Symposium, 2016.
[51] Anjith George, Aurobinda Routray, “Fast and Accurate Algorithm for Eye Localization for Gaze Tracking in Low Resolution Images”, IEEE, 2016.
[52] Zhang Naizhong ,Wen Jing, Wang Jun,” Hand-Free Head Mouse Control Based on Mouth Tracking”, International Conference on Computer Science & Education, 2015.
[53] S.Waphare, D.Gharpure and A.Shaligram, B. Botre, "Implementation of 3-Nose Strategy in Odor Plume-Tracking Algorithm”, IEEE, 2010