Face Detection using Spatio-Temporal Segmentation and Tracking for Video Security

[1]Ekta Thaman, [2]Ramanpreet Kaur
[1]Chandigarh Group of College Jhanjeri
Punjab

***Abstract* - Real-time face detection and tracking is a challenging problem in many application scenarios such as faces and gesture recognition in computer vision and robotics. The ongoing work implements a robust method which can locate multiple faces simultaneously under the changing environment of light illumination and complex background in real time by using face detection and tracking in conjunction with depth information. To achieve robustness, the Kanade-Lucas-Tomasi point tracker is used and specialised to work on facial features by embedding knowledge about the configuration and visual characteristics of the face. The resulting tracker is designed to recover from the loss of points caused by tracking drift or temporary occlusion. Performance assessment experiments will be carried out on a set of video sequences of several facial expressions. It will be shown that using the original Kanade-Lucas-Tomasi trackers, some of the points are lost, whereas new method will be proposed to overcome the drawback.**

***Keywords*— Face recognition, Preprocessing, characteristics match, Image Processing etc.**

I. INTRODUCTION

There has been a lot of work on the domain of tracking-by-detection. Those methods rely on an object (people in our case) detector that generates probabilities of a person being at a point on the ground plane of the scene at each frame. Those detections are linked together to form trajectories. When no other information is available and the detections are noisy, the resulting trajectories might be inconsistent and contain identity-switches between the tracked people. By exploiting the facial characteristics we minimize the identity switches and identify the people being tracked.

The technique focuses on the automatic recovery of points lost between frames. It is specialised for robust tracking of human facial features. This new method is intended for use in robust recognition of facial information such as the identity or facial expression of a person. Such recognition applications are important components of future machine interfaces, for example. Instead of trying to improve tracking performance through the automatic selection of better features, as proposed in, here we exploit the knowledge that the tracker is working on a human face. Several constraints are applied during the initial selection and tracking of feature points.

A face detector is initially employed to estimate the position of individuals. Those positions estimates are used by the face detector to prune the search space of possible face locations and minimize the false positives. A face classifier is employed to assign identities to the trajectories. Apart from recognizing the people in the scene, the face information is exploited by the tracker to minimize identity switches. Dynamic face detection refers to detection of faces in the video frames and simultaneously tracks the movement of the face in the consecutive frames. To achieve robustness, the Kanade-Lucas-Tomasi point tracker is used and specialized to work on facial features by embedding knowledge about the configuration and visual characteristics of the face. A face detector is initially employed to estimate the position of individuals. Those positions estimates are used by the face detector to classify the search space of possible face locations and minimize the false detections. A face classifier is employed to assign identities to the trajectories. Apart from recognizing the people in the scene, the face information is exploited by the tracker to minimize identity switch. The KLT algorithm tracks a set of feature points across the video frames. Once the detection locates the face, the next step in the example identifies feature points that can be reliably tracked. The tracker calculates an optical flow for a sparse feature set using the iterative Lucas-Kanade method with pyramids. The algorithm basically looks at two consecutive video frames and determines how points move by using a straightforward least-squares optimization method. The objective of object tracking step is to compute object trajectories. This task relies on the number of matching feature points over frames between detected objects.

As the most significant external characteristics of human, face plays a vital role in communication. With human computer interaction technology turning into a newest topic within field of AI, face detection and trailing has become a serious concern analysis direction in computer vision that has broad application prospects within field of human computer interaction, video game so on. The complexness of the face results in a definite degree of issue for quick the detection and trailing. There are various ways for the face detection, for example: mathematics technique, templet matching technique, the strategy of support vector machine, the strategy of active contour model, the strategy of variability template, etc. however the a lot of thought approach for face detection is Adboost-based, however, the speed of that doesn't meet the time period in high-resolution video sequence detection. Search candidates face by mistreatment the strategy of color and contour detection, then make sure pattern matching recognition through the algorithmic rule of principal element analysis, the trailing accuracy rate of that is sort of high, however the track result is poor within the event of occlusion. In another thesis, an algorithmic rule, that is combined multiple target hunter with face observe technique, with that is ready to trace the multiple faces at a similar time. This technique has higher time period performance; however the algorithmic rule is a lot of advanced. It takes the Mean-Shift technology to trace moving target that has the quick and economical characteristic. However its trailing window is mounted. Once the target moves at the side of the direction of the camera, it had lost.

II. LITERATURE REVIEW

Petronel and Corneliu (2005) Face detection and recognition may be thought of enabling technology for spread of image technologies, authentications and advanced UIs in hand-held devices. Many modes to implement the face detection algorithmic program in hardware square measure explored and hybrid approach is found to use the most effective trade-off between the physical hardware resources and machine software system resources [1].

Sudipta N. Sinha and Jan-Michael Frahm (2007) describes novel implementations of KLT feature Recognition Or Sift Feature extraction algorithms that run on descriptions getting ready unit which is to appropriate for feature examination more and more vision frameworks. Whereas wide quickening over normal central processor usage is taken by abusing the correspondence gave to wound edge programmable style equipment, central processing unit has organized to run completely distinct calculation in parallel. Each SIFT and KLT have been utilized for intensive form of computer vision assignments to extend from structure movement, automaton route, stretch reality to face recognition, object detection and video data-mining with the quite promising results [2].

Yohan Dupuis, Xavier Savatier (2013) proposed the Bio-roused and non-customary vision frameworks square measure for the subjects. Among them, spatial relation vision frameworks have displayed their capability to increase the geometrical understanding of the scenes. In any case, number of specialists explored way to perform the object location with frameworks. Present methodologies oblige geometrical before elucidation of the icon. During this paper, tend to analysis what should be thought of and the way to method omni directional footage gave by detector. It tends to center our examination on face location and highlight approach that specific thought got to paid to descriptors thus on the effective performance face recognition on omni directional footage. It tends to exhibited however image process is true presently performed on footage. At that time to focus the thought of per user on the theoretical and sensible problems enclosed in image unwrap. We tend to exhibit the unwrapped image handling to take the regular time getting ready catadioptric image because it appears to be. Handling show an important variable, we tend to develop the last commitment with relation to procedures that are clad to exceptionally effective [3].

Muhammet Baykara and Resul Das (2013) proposed the security that comes into a lot of prominence a day. This is important for those people to stay lot of passwords in brain and convey lot of cards with themselves. Such usage in case, square measure develop to less secure and helpful, while this manner prompting associate degree increasing enthusiasm for techniques known with statistics frameworks. Statistics frameworks square measure the frameworks that store physical properties of people in electronic surroundings and empower them to perceive by the place away electronic information once needed. Statistics is recognizable proof of human. Biometric frameworks square measure structures that are often utilized as an area lately years. This biometric framework is getting under consideration the employment of few physiological parts of an individual for security. This is anticipated that biometric frameworks are an important piece of info security frameworks within nearing years [4].

Tong Zhang(2014) presents the generality of checking the cameras introduced enter open schools, spots, houses and clinics, feature examination advancements to translate the created feature substance square measure developing to to individuals' lives. On this setting, tend to boost a human-driven feature observation framework which distinguishes and tracks the people in exceedingly given scene. During this paper, a parallel pipeline is usually recommended that coordinates image handling the modules within framework, face location, for instance, individual acknowledgment and following, proficiently and simply, therefore varied people may an equivalent time followed endlessly. An individual's identifying proof and follow framework for feature security utilization cases was displayed. Notably, a data processing pipeline for feature getting ready has planned and developments were developed to image investigation modules for each productivity and sincerity [5].

Jatin Chatrath, Pankaj Gupta (2014) describes the technique for real time face detection for recognition utilized an altered variant of calculation counseled by Paul viola and Michael Jones. The paper starts with introduction to face discovery and following, trailed by the giving of Vila Jones calculation and afterward talking to concern the execution in real feature applications. Its calculation was seeable of article location to remove few explicit elements from image. This paper introduces an appointment of itemized trials on hard face recognition and following data set that has been targeted. This data set incorporates below an intensive form of conditions including: scale, brightening, camera selection and posture [6].

Cheng-Yuan blow, and Liang-Gee Chen (2014) tend to propose the algorithmic program victimization simply 2 factor webcams while not adjustment to acknowledge separation within middle of consumer and display by face location. Face recognition primarily based on the stereo coordinate calculation to spot client's separation by the face location. Typical stereo framework utilizing stereo coordinating to get client's profundity by aligned information image, top side of calculation is to get nice results with the low quality and standardization free [7].

III. PROBLEM FORMULATION

The technique focuses on the automatic recovery of points lost between frames. It is specialised for robust tracking of human facial features. This new method is intended for use in robust recognition of facial information such as the identity or facial expression of a person. Such recognition applications are important components of future machine interfaces, for example. Instead of trying to improve tracking performance through the automatic selection of better features, as proposed in, here we exploit the knowledge that the tracker is working on a human face. Several constraints are applied during the initial selection and tracking of feature points. Therefore methods are required to overcome the problems and develop a new system for detection as well as tracking.

IV. METHODOLOGY

Human face detection and tracking is an important research topic for many applications in computer vision. However, detecting human face in images or videos is a challenging problem due to high variety of possible configurations of the scenario, such as changes in the point of view, different illumination, and background complexity. The method of face detection and tracking is connection of two stages –

1. Facial feature extraction using Principal Component Analysis and,
2. Motion tracking based on points extracted.

Fig. 1 Facial feature extraction

The method described here is concerned with several issues that are considered to be important for robust recognition of facial features. These issues are rigid and non- rigid motion, variation of lighting conditions, head orientation, head tilting, and real-time tracking. The established point tracking scheme aims to tackle all those issues. The technique is designed to be robust and to recover points lost during tracking. The main constraint is for the nostrils to be visible. The tracker works on grey level image sequences of any length. The general tracking process is shown in Fig 2.

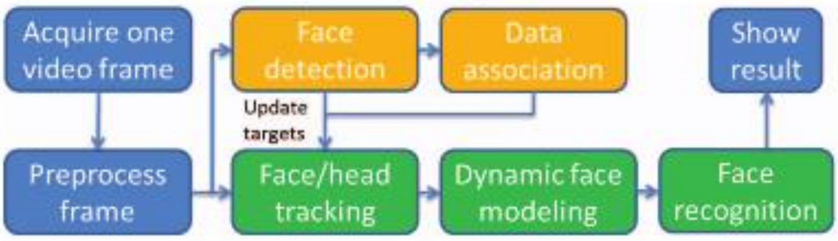


Fig 2. Flow chart

Motion tracking based on points is often important in many applications requiring time- varying image analysis. Applications may be object tracking, motion understanding, navigation, automatic speechreading, and facial feature tracking. This paper presents a new robust technique based on the Kanade-Lucas-Tomasi tracker. The technique focuses on the automatic recovery of points lost between frames. It is specialized for robust tracking of human facial features. This new method is intended for use in robust recognition of facial information such as the identity or facial expression of a person. Such recognition applications are important components of future machine interfaces, for example. Instead of trying to improve tracking performance through the automatic selec- tion of better features, as proposed in, here we exploit the knowledge that the tracker is working on a human face. Several constraints are applied during the initial selection and tracking of feature points.

Face recognition process

1. Image size normalization, histogram equalization and conversion into gray scale are used for preprocessing of the image.

2. The face library entries are normalized. Eigenfaces are calculated from the training set and stored. An individual face can be represented exactly in terms of a linear combination of eigenfaces. The face can also be approximated using only the best M eigenfaces, which have the largest eigenvalues. The eigen vectors are given as:

$$\lambda_k = \frac{1}{M}\sum_{n=1}^{M}(u_k^T \Phi_n)^2$$

Tracking process

The KLT algorithm tracks a set of feature points across the video frames. Once the detection locates the face, the next step in the example identifies feature points that can be reliably tracked. The tracker calculates an optical flow for a sparse feature set using the iterative Lucas-Kanade method with pyramids. The algorithm basically looks at two consecutive video frames and determines how points move by using a straightforward least-squares optimization method. The objective of the object tracking step is to compute object trajectories. This task relies on the number of matching feature points over frames between detected objects. Let P be the number of matching KLT features between two objects O_t^i and O^j_{t-1} . We define a similarity score between these two objects as follows:

$$S_{KLT}(o_t^i, o_{t-1}^j) = min(\frac{P}{M_{o_{t-1}^j}}, \frac{P}{M_{o_t^i}})$$

where $M_{O_t^i}$ and $M_{O_{t-1}^j}$ are respectively the total number of KLT feature points located on object O_t^i and O^j_{t-1}. The algorithm is then applied to find the best matching of objects between t − 1 and t.

Kanade-Lucas-Tracker algorithm

It is simple for human eye to follow another person's head while she or he moves around. So, the same challenge for the computer, but there is a different story. In the world of computer vision tracking it's a wonderful task. There are mainly three steps that have to take place while human face tracking:

- Finding the face in an image place
- Opting the features to track
- Tracking the features

1. Feature Selection

The main important step before detection is the selection of tracking features. How do we opt the features which are trackable? The keyword during feature selection is texture. Area with a modifying texture pattern, is mostly unique in an image, during linear or uniform intensity areas are common and not unique. It is clear that we must have to look at areas of texture or the first logical step would therefore be an investigation the image intensity gradients.

$$\mathbf{gg}^T = \begin{bmatrix} \frac{\delta(I)}{\delta x} \\ \frac{\delta(I)}{\delta y} \end{bmatrix} \begin{bmatrix} \frac{\delta(I)}{\delta x} & \frac{\delta(I)}{\delta y} \end{bmatrix} = \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix}$$

Now integrates the matrix derived over the area W, that we get :

$$Z = \iint_W \begin{bmatrix} g_x^2 & g_x g_y \\ g_x g_y & g_y^2 \end{bmatrix} \omega d\mathbf{x}$$

2. Tracking facial features

Let assume that human face is already located in an image sequence, and then the tracking is based on Kanade-Lucas tracking equation. In the end of technique can be described as follows: Firstly you have to select all the features which can be tracked from image to image in video image stream. The opting of features is based on texture. A number of fixed sized features are selected on the head of person in first image of sequence. These windows features are tracked from one image to next using the KLT method. The tracking method calculates the total sum of squared intensity differences b/w a feature in the previous image or the features in current image. The replacement of the particular feature is then defined as the replacement that minimizes sum of differences. It is done continuously between sequential pictures so; all the features can be tracked.

3. Calculating feature displacement

The main basic information has now been build to solve the replacement D of face feature from one image window to next window. For simply, we define the second window as B(x) = I(x,y,t+_) and the first window A(x-d) = I(x-d) = I(x- x, y-y,t) as where x= (x,y). Now, the relationship between the two images is given:

$$B(\mathbf{x}) = A(\mathbf{x} - \mathbf{d}) + n(\mathbf{x}).$$

In the equation, n(x) is a noise function which is caused by the ambient interference like reflections and scattered light. In this stage you can define an error function that has to be minimized in order to minimize the noise effect

$$\epsilon = \iint_W [A(\mathbf{x} - \mathbf{d}) - B(\mathbf{x})]^2 \omega d\mathbf{x}$$

4. Interpretation of the method

Now, we have covered the mathematics portion, so we look at what happens physically while the execution of the tracking algorithm. The tracking steps are: Firstly we assume that one feature window in the first picture and we want to determine the replacement of this window in the second image.

FACILITIES REQUIRED AND PROPOSED TECHNIQUE

The facilities required for the proposed technique are:

1. Set of image frames with human faces to be detected
2. Face detection method to automatically identify the faces in the frame
3. Facial feature detection algorithm for extracting facial features

For real time face detection, instead of selecting the facial features for the face directly. The proposed method is intended to implement skin color and motion cues to detect face and then automatically extract the facial features and then track the movement of the face using the tracking algorithm. The motion detector computes a motion mask by thresholding the result of frame differencing. The motion mask is applied to the skin likelihood image to obtain the the position of the face.

The main differentiating feature of this proposed method from existing work in the gesture recognition is that, our method needs neither temporal nor spatial segmentation to be performed as pre-processing. In multiple dynamic gestures recognition systems the lower level modules perform temporal and spatial segmentations and release shape and motion features. So those features are passed into recognition module that classifies the gesture. In such bottom-up methods, the recognition will fail when the results of temporal or spatial segmentation are incorrect.

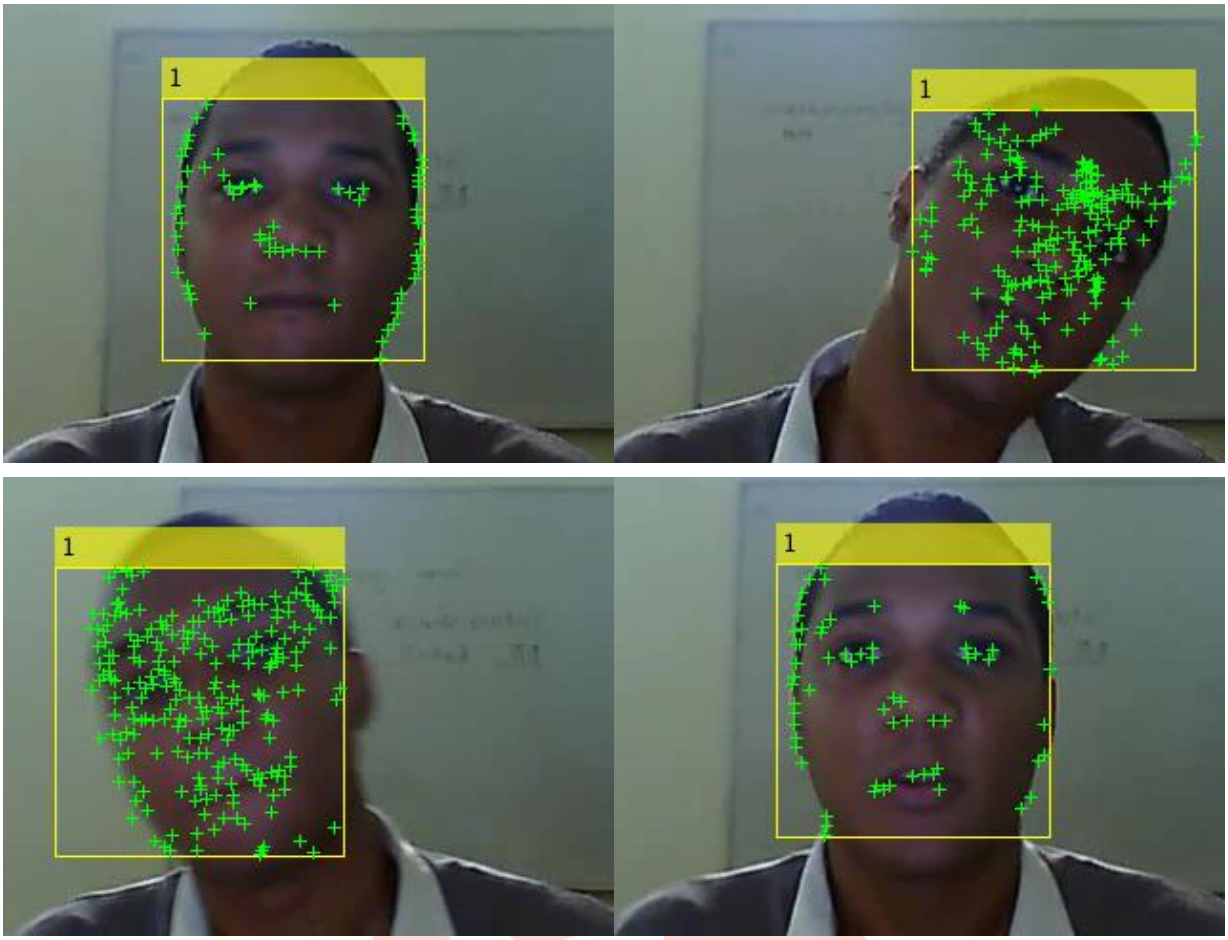
FACE TRACKING USING SPATIOTEMPORAL

The spatiotemporal matching algorithm is designed to accommodate multiple hypotheses for the face location in each frame. Thus, we can afford to use a relatively simple and efficient pre-processing step for face detection, that combines skin color and motion cues. The skin detector first computes for every image pixel a skin likelihood term. For the first frames of the sequence, where a face has still not been detected, we use a generic skin color histogram to compute the skin likelihood image. Once a face has been detected we use the mean and covariance of the face skin pixels in normalized rg space to compute the skin likelihood image.

The motion detector computes a motion mask by thresholding the result of frame differencing. The motion mask is applied to the skin likelihood image to obtain the face likelihood image. Using the integral image of the face likelihood image, we efficiently compute for every sub window of some predetermined size the sum of pixel likelihoods in that subwindow. Then, we extract the K subwindows with the highest sum such that none of the K subwindows may include the centre of another of the K subwindows. Each of the K subwindows is constrained to be of size 40 rows 30 columns. Alternatively, for scale invariance, the subwindow size can be defined according to the size of the detected face.

V. RESULTS

The results presented are there for the Viola-Jones method. The solution has implemented to create an application which works on both recorded and live videos. When software application is running, a user can interfere at any time to tag a detected face in frame. A vector feature of tagged face can be computed and saved in gallery. Due to the parallel computing in various levels and distinct modules of system, critical work has been conducted in implementation to avoid the conflicts among individual threads. In this, MATLAB's inbuilt functions are used in multiple portions of implementation. The software application with 320x240 frame size has been tested on the videos. These experiments made on various mainstream laptops, desktop and PCs, each contains up to two multi-core processors. Consistent performance has been attained from PCs. Face detection process works once every 20 frames. The calculation of facial feature for face recognition takes 44ms per frame. For tracking purpose, minimum eigen values are considered as feature points. Dynamic face modeling can be called once every seven frames. In dynamic collection, each subject has almost five clusters and each one contain maximum 20 samples. In the dynamic face modeling, subject has been identified much faster when someone reappears in video of standard benchmark videos, it is not practical to compare the detection accuracy with prior arts.



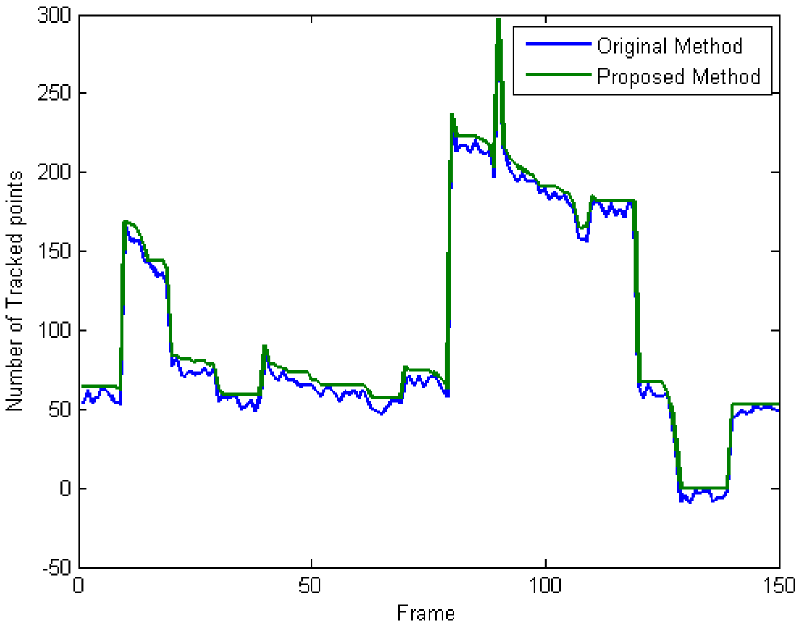


Fig 3. Graph of number of points tracked.

CONCLUSION

In this work, a facial identification and tracking system for security purpose use is presented. A novel method, of face detection using spatio-temporal segmentation is proposed and innovations were made to image examination modules for effectiveness and robustness. The results are published in this paper and are generated using MATLAB which allows immediate identification and strong detection of face in high-resolution videos.

Future research is expected to be carried out in a different of aspects to improve this work. More efforts can be made to enhance up the tracking method so that the system run faster, and specifically, the system's cost may reduce. Also, the system can be made robust against occlusion.

REFERENCES

[1] Gong, M., Langille, A., Gong, M.: Real-time image processing using graphics hardware: a performance study. In: International Conference on Image Analysis and Recognition, pp. 1217–1225 (2005).

[2] Cheng-Yuan Ko, Chung-Te Li, Chen-Han Chung, and Liang-Gee Chen, "3D hand localization by low-cost webcams," IS&T/SPIE Electronic Imaging (IS&T/SPIE EI), Jan. 2013.

[3] E. Maggio, E. Piccardo, et al. "Particle PHD filter for multitarget visual RECOGNITION," IEEE Conf. on Acoustics, Speech and Signal Processing, p.15-20, 2007.

[4] F. Qureshi and D. Terzopoulos. Surveillance camera scheduling: a virtual vision approach. In Multimedia Systems, volume 12, pages 269–283, 2006.

[5] P. Salvagnini, M. Cristani, A. D. Bue, and V. Murino. An experimental framework for evaluating ptz RECOGNITION algorithms. In International Conference on Computer Vision systems, pages 81–90, 2011.

[6] K. Smith, D. Gatica-perez, J. marc Odobez, and S. Ba. Evaluating multi-object RECOGNITION. In In Workshop on Empirical Evaluation Methods in Computer Vision, 2005.

[7] Sudipta N. Sinha ,Jan-Michael Frahm: Feature RECOGNITION and matching in video using programmable graphics hardware, Received: Springer-Verlag 2007.

[8] Cheng-Yuan Ko, and Liang-Gee Chen " Acquires user's distance by face detection" IEEE Graduate Institute of Electronics Engineering 2014.

[9] Muhammet and Resul Das" Real time face recognition and RECOGNITION system " Software Engineering Department, Fırat University Elazığ/Turkey 2013.

[10] Petronel Bigioi "Advanced Hardware Real Time Face Detector" IEEE, Corneliu 2007.

[11] Tong Zhang "Video Security With human identification And RECOGNITION" IEEE, 2014.

[12] Yohan Dupuis, Xavier Savatier "Robust Radial Face Detection for Omnidirectional Vision" IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 22, NO. 5, MAY 2013.

[13] Jatin Chatrath, Pankaj Gupta "REAL TIME HUMAN FACE DETECTION AND RECOGNITION" International Conference on Signal Processing and Integrated Networks (SgPIN) 2014.