A perlustration of Human Apprehension and Behavior Accedence in surroundings

Monika[1], Madhulika[2]

[1] *Student, Computer Science & Engineering, Amity University, Uttar Pradesh, India*

[2] *Assistant Professor, Computer Science & Engineering, Amity University, Uttar Pradesh, India*

Abstract

In surveillance system various methods are implemented but have some limitation and flaws. This paper presents a systematic description of the research on relevant human detection and behavior-recognition methods for surveillance system to use in various civic services. Detection method includes two class classifications differentiating human versus non-human whereas recognition methods includes multi-class classification of all actions. This paper is an approach to summarize the methods to enhance the implementations and get an optimal upshot.

Key Words: *human apprehension, human behavior accedence, video diagnostics, smart surveillance system*

1. INTRODUCTION

Today, Surveillance systems are not constraint to Military and Intelligence because the rapid increase of population in metro cities directly increasing the crime and other perils. In the last decades, various surveillance systems are emerged with improvements but the means of apprehension is although not satisfactory.

To understand the scope and extent of surveillance system, let's deliberate for Delhi the capital of India has the second most populated city in world bears a 22 million of population, and if we reminisced with the peoples in civic area then average ridership of 1.6 million commuters of Delhi metro whereas IGI airport has 37 million passengers annually and 11.2 million vehicles on road of Delhi daily.

Delhi metro with its 146 stations only have 300 CCTV. Thus, the surveillance system for Delhi metro requires around a 300 of employees for daily operations on the current record basis [1]. Similarly, if other civic areas are added it might be possible that more than millions of surveillance cameras require which makes a surveillance system dependent for human which will be operationally not feasible.

Since amount of data in form of video increases, most of the surveillance system available with the only facility to save, capture and share the video and also to analyze that data they are solely dependent on human analysts. Detecting specific actions in a live digital data (can be result from a CCTV) or exploring in video records (i.e., video diagnostics) mostly dependent on expensive and limited human utility. Presently, multiple analysts are employed to detect multiple actions in real-time video data by simultaneously watching the same video stream.

Thus, we are assigning a portion of video to each analyst to give generate list of actions (behaviors) or substances which will be found suspicious. Afterwards, the analyst will report an alert message to the appropriate authorities if some of the suspicious actions or a substance is recognized. This, task is severe to labours (video analysts), exhausting and also full of faults.

Table 1: Related Literature Survey Summary

First Author	Yr	Topic	Ref
Tetsuhi	13	Pedestrian Identification	2
Rupali S. Rakibe	13	Background subtraction algorithm based human motion detection.	3
Yang	13	Human detection and behaviour recognition on depth sensor.	4
Lu Xia	11	Human Detection using depth sensor by KINECT.	6
Joshua	10	Understanding Transit Scenes.	9

The potential to speedily react and explore on huge amount of videos or real-time footages presents an impressive capabilities to surveillance societies because up till now research development only endeavour to detect human and detecting some typical behaviours while the objective is to formulate a system which is capable of identifying inauspicious activities in civic areas basically human aggressive behaviour by change of temperature of a human body, facial expression , way of walking etc. similarly monitoring on waif things like a bag near wall and no one is carrying it.

Previous surveys underlined as ground development methods deployed in surveillance (that will be consider as "core technique," e.g., tracking and classification of human) and in comparison, emphasize is on accedence of human behavior, in essence of securing civic services. Though, for transparency, core techniques are reviewed under state- of-the-art, and also previous surveys and works in associated domains are determined (catch a glimpse of Table 1).

The latest research to the topic is 'Pedestrian Identification' [2] by Tetsuhi, in the research the pedestrian are identified by a wearable accelerometer. The research aims to locate and identify a person in crowd because accelerometer facility is also available in cell phones. Looking on the next paper from the table 'Background Subtraction for Human Detection' [3] presented by Rupali S. Rakibe, intended to detect moving substances from static background scene based on background subtraction algorithm.

The methodology is to set a dependable background where some statistical data is used to update the model, afterwards filtering is applied to reduce noise and done for smoothing, then the contour projection analysis is merged with the shape analysis in order to remove the shadow effect on the image; and in this way Rupali S. Rakibe detected the moving human bodies accurately and reliably. Point contour detection method can also be used to extract the precise contours in Human body [11]. Edges are a kind of Image Segmentation Techniques which shows the presence of edges in an image and add boundaries in that image [10].

Research presented by Yang, approaches to use a depth camera as a sensor which normalizes digital elevation map, this image is built from depth image where the intensities of pixels are indicating the elevation of the scene, built from the depth image. Basically, they use oriented templates to detect a human head by matching the predefined data set to new data set to classify and detect whether a human or not. Next research paper in reference table is presented by Lu Xia proposes human apprehension method through depth information where KINECT for XBOX 360 is used as input.

The model approaches to detect human using a 2-D head contour model and a 3-D head surface model, a segmentation algorithm is used to segment human from environment and extracting the full contour. Limitation of this method is an extremely dependent on exact head accedence. The research work discussed here are now in trend and their approaches can fulfil the aim. However this paper presents an approach to improve the former papers as listed in the Table1.

1.1 Outline

This study emphasizes to present a complete survey of vision-based processing of detecting human and accedence of their behavior's algorithms in context of securing civic services through surveillance. Before determining a human behavior we need to apply pr-processing to detect a human from the space which is presented in "core techniques" section of the paper.

Human-behavior accedence using video or image starts with apprehension of human from the environment (e.g. apprehension of human from a market because various different substances are in the environment) which is commonly achieved through environmental modelling or motion-based segmentation. These foreground substances are classified depending on the applications i.e. human vs. non-human (vehicles etc).

Finally, tracking commences the relationship among the substances and the environment. Section a concise glimpse at the core techniques, to facilitate the understanding of later sections of this paper. For organization purpose, all relevant surveys dealing

with core techniques are identified and summarized in Table 1. In Section 3, we are discussing behavior-accedence strategies whereas Section 4 explains future research directions along with the weaknesses through the current state- of-the-art and last section, Section 5 encapsulating this paper.

The latest research by Tetsuhi [2], that the pedestrian are identified by a wearable accelerometer. The research aims to locate and identify a person in crowd because accelerometer facility is also available in cell phones. Looking on the next paper from the table presented by Rupali S. Rakibe [3], intended to detecting moving objects from static background scene to based on background subtraction. Research presented by Yang [4], is to use a depth camera as a sensor which normalizes digital elevation map, where the pixel intensity indicating the elevation of the scene, built from the depth image. Next in the list by Lu Xia [6], proposes human detection method using depth information by the KINECT for XBOX 360. The model is based on an approach which detects humans using a 2-D head contour model and a 3-D head surface model, a segmentation scheme is used to segment human from surroundings and extracting the whole contour of the figure based on the detection point. The research in [5],[7] will states various methods regarding head estimation, detecting people in RGB-data respectively whereas [9] gives a detailed study of transit system security but limited to present an intelligent system.

2. CORE TECHNIQUES

2.1 CLASSIFICATION

Current surveillance system conventionally consists of a little kind of motion apprehension methods. It is implemented in order to fragment moving substances from the remaining background. Information regarding motion of a human (substances in the space/environments) is useful for classification of human from the space as well recognizing their behavior. Human Apprehension is basically a two class classification human vs. non-human. Discussion is on latest techniques like 'Background Subtraction and Temporal Differencing' [3], 'Depth Sensor' [4], 'Depth Sensor by KINECT' [6].

1) Subtraction of Background and performing temporal Differentiation: A popular segmentation strategy is background subtraction. The basis of the methodology is to extract foreground objects from the regions if a considerable distinction among the estimated and the observed image is found. Latest algorithm based on this technique is presented by Rupali S. Rakibe (2013) where the human is detected from a static background using background subtraction methods. This algorithm has following steps as shown in figure 2. The mentioned steps in diagram when applied on a static background human are extracted. To successfully make the algorithm in run the main step is to initialize the background image (input). In order to initialize the background, it is observed that a shadow problem arises, where median method is applied to set the background. Once the background image is acquired, background image is subtracted to the current frame to set the threshold value, if point appears in a moving substance it is determined else, it is in rest of the image. Moving substance after applying threshold can be detected on the basis of two judging parameters which are as given below:

- The substance area is outsized than the threshold.
- The aspect ratio of the substance region should match to the set ratio.

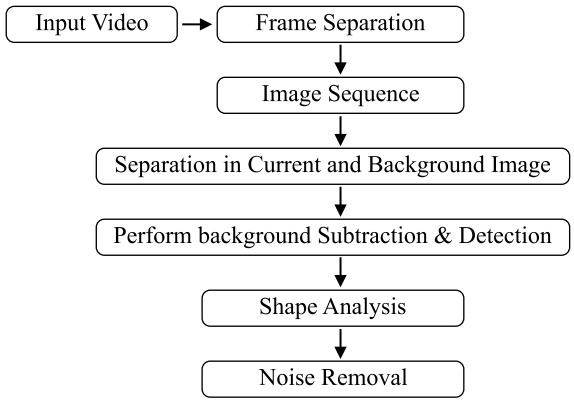

Fig. 1: Stepwise Procedure of background subtraction

On the basis of the above mentioned two conditions, the moving substance can be a moving human or non- human is conformed. The whole process is termed as moving substance mining. Since the image obtained has some sort of discrepancy due to large number of noise. Thus, the need is to filter out non-human activity areas (e.g. moving cars, flying birds) by corrosion and expansion operations while retaining the shape of human motion without any distortion.

As shadow is also a distortion; the correct extraction of the moving substance disturbed due to shadow. The method applied is to combine horizontal and vertical portions so that some sort of shadow can be eliminated. Afterwards vertical projection is analyzed in order to set the threshold value to remove pseudo-local maximum points and pseudo-local minimum points which determine the body in motion region with their width. So, that human body with movement is to be obtained with precise edge.

2) Depth Sensor: The research paper under the topic to discuss is 'Human detection and behaviour recognition on depth sensor' presented by Yang and 'Human Detection using depth sensor by KINECT ' presented by Lu Xia. According to Yang there are three main ways to get depth information:

- Binocular vision based methods
- Time-of-flight based methods
- Structured light codification methods.

On these above mentioned concepts, Yang formulated a technique called as 'Digital Elevation Map Generation' with phases as:

- Camera Calibration
- Depth Value correction
- Normalized digital elevation map generation.

To map a relationship between the world co-ordinates system image co-ordinates system camera calibration is performed. Yang introduces a new standardized method where we standardized the camera using two special correspondences with moderate precision. The introduced camera model is made up of three parts:

- Perform Rotation and translation between two co-ordinate systems.
- Perspective projection with the pinhole geometry.
- At last "sampling".

In comparison with other method named as DLT, it requires minimum of six correspondences while introduced method needs only 2 correspondences. Thus, it will efficiently standardize the depth cameras, because we can manually calculate the co-ordinates for correspondence.

According to Yang to generate a normalized digital elevation map three main steps are followed as:

According to Yang to generate a normalized digital elevation map three main steps are followed as:

Step 1: Initializing a map Elevation = *zero(L,W),*

Step 2: Updating the elevation intensity for a pixel *(u,v)* in depth image, using following equation:

$$Elevation(x_w^{u,v} \quad y_w^{u,v}) = \max(z_w^{u,v} \quad , Elevation(x_w^{u,v} \quad y_w^{u,v}))$$

Step 3: Every pixel in depth image is traversed.

Let us discuss 'Human Detection using depth sensor by KINECT' presented by Lu Xia in figure 3, this research approaches to detect people in indoor environments using depth information obtained by KINECT. At first noise is reduced from the input depth array to smooth it for later process.

To detect human, Lu Xia uses a 2 stage head apprehension process, initial step in this order is to depict the appearance of human in embedded depth array by searching boundary information, this methodology is termed as 2D chamfer distance matching. The next and last step it to verify information array by utilizing the relational depth to examine the image regions using a 3D head model, which was scanned by the 2D chamfer in initial phase.

Fig. 2: Model proposed by Lu Xia.

In order to process data some initial pre-processing is needed on input data in order to remove noise; to avoid this interference, the noisy pixels are filled by nearest neighbor interpolation algorithm to give meaningful values in pixels, and next step is to smoothen the data of depth array through a median filter. As pre-processing is completed next step is to process the data in 2D chamfer distance matching, which generates the edges of all objects in the picture, and to remove the entire set of edges whose dimensions are minor than a defined threshold. This result generates a distance map of the edges which is used to match with the threshold to detect the 'head like object'.

Next step is to computing the parameters of the head to fit them in 3D model. Once, depth and height is measured, next is to explore for head inside a circular fragment identified by a radius. To fit in the 3D model circular region is extracted and depth is normalized, while a threshold decides whether the fragmented is truly a head or not. Now, overall contour of the person is extracted to recognize the activity.

Let understand figure 3 from top to bottom, in order to process data some initial pre-processing is needed on input data in order to remove noise; to avoid this interference , the noisy pixels are filled by nearest neighbour interpolation algorithm to give meaningful values in pixels, next step is to smoothen the data of depth array through a median filter. As pre-processing is completed next step is to process the data in 2D chamfer distance matching, which generates the edges of all substances in the background, and to remove the entire edges whose dimensions will be lesser than a specific threshold.

This result generates a distance map of the edges which is used to match with the threshold to detect the 'head like substance'. Next step is to computing the parameters of the head to fit them in 3D model. Once, depth and height is measured, next is to explore for the head inside a rounded section define by radius. To fit in 3D model rounded section is extracted and depth is normalized, while a threshold decides whether region is actually head. Now, overall contour of the person is extracted to recognize the activity. Recognizing and tracking is discussed later.

2.2 TRACKING

In the context of civic service based applications, tracking is termed as the dilemma of approximate trajectory of human (as a pedestrian, moving in vehicle, walking) in the image plane when he is in an environment (railway station, airport, on road, malls etc.). Although tracking is a difficult task as many problems likes moving substances, noise, occlusions (hidden surfaces), substance intricacy, scene lighting variations, and sensor hardware. Objects move fast relative to the frame rate this is the main difficulty in tracking associate target in a video.

Tracking systems assign persistent identifications tags to tracked person, depending on the application requirements, while a system need to maintain other characteristics related to video like feature ratio, area, figure, colour information, etc. Thus, choosing fine qualities is a necessity as it can be used for future tracking or identification.

Lu Xia presented a better concept for tracking. According to Lu Xia 'Colour is basis for tracking in RGB picture, assuming colour of the same substance in different time frames must be similar; but tracking using depth sensor, need

Table 2: Comparison of Classification Techniques

PROPERTIES	Background Subtraction Algorithm	Depth Sensor Using Kinect
Basis	Deletion of background and extracting foreground objects for detection.	Detecting human through depth information or digital elevation map
Input Datasets	Static Background	Depth array in 3D form.
Resultant Output	Moving human body is extracted with precise edge.	Detection is seen in a gray scaled value while through second method whole contour is extracted.
When to Detect	Detecting inauspicious data statically.	Detecting inauspicious data dynamically.
Application Area	Analyzing on previous digital data.	Reporting on current scenarios.
Complexity	High	Comparatively low

is a 3D space which contains information of the substances, so that the measures for the movements is done in 3D space. The formula for tracking derived by Lu Xia is:

$$E = (c - c_0)^2 + (v - v_0)^2$$

In the given equation, is representing the energy, coordinates of a person in current frame is represented by and the coordinates of the person in the last frame is represented by and speed of the person in current frame and last frame is represented by v and v_0 respectively.

3. HUMAN-BEHAVIOR ACCEDENCE

Limiting visual surveillance only for detecting human or scanning human in civic area is not the need of today. Researcher needs to improve the surveillance systems to enhancing and improving the security in order to present an intelligent surveillance system which not only detects human but also recognizes the inauspicious activity from human and thus notifying the authorities to take immediate action.

In [4], with the help of the precise digital elevation map based images, accedence of behaviour are under action by detecting the position of a person by intruding the defined digital elevation map. Yang detected three behaviours accurately which are:

- *Human gathering and fast moving* is recognized by the average distance in the surveillance scene and a gathering alarm is fired if average distance is beneath the set threshold and a if a human is moving fast it will cross the set threshold value and a alarm raise for fast moving human.
- *Human Falling Down* is recognized if a person elevation is gone downwards such that the defined elevation area for a person increases significantly and the motion is under the threshold.
- *Aggressive Behavior* set up by an Energy function, which represents the activity level of a human. As an example, when a fighting is happened the number of frames increases for the movements, which significantly increases the motion's energy and system detects change as increase in number of frames in the defined time and to reflect is alarm is raised.

Today we are recognizing inauspicious activity through analyst by monitoring on cams which is not enough efficient. Although 'Human-Behavior Accedence' is not formulated like 'Human Apprehension' as it is a multiclass classification of human actions like aggressive actions. Hereby we are classifying Human-Behavior into two groups as:

- ***Person/ no interaction*** are termed as person(s) that are not mixing up with some distinct person or some other means (bus, metro, auto etc). For example, a person is waiting for something/ someone (lurking), a person is talking on a cellular phone from a long duration in public place.
- ***A several-person interaction*** includes behavior where a person involved in each element in space. For example, trailing, gathering, walking together, running in crowd, exchanging substances.

3.1 Person/ No Interaction

- ***Lurking*** is an activity of single person not interacting with others. Technically it is a person in space for more than the defined threshold time. Lurking is special interest for civic area surveillance as it is observed that it is a universal place of dope suppliers, vagrants, attackers, between others. In previous surveys, when a lurking activity is detected it is observed that before going to be in lurking the person indulged in activities like entering and leaving. One of the techniques proposed as a Gaussian's background subtraction algorithm which detects motion globule in a scene, if the time for threshold is extended than the maximum time defined.
- ***Crowd Counting:*** Accurately detecting people in civic area increases managerial efforts as places with more obstruction or signalling requires high consideration. In addition, for evaluation of crowd in under-ground transportation system (like Rajeev Chowk metro station in Delhi) we can get a fine assessment by the time needed the passengers to wait for metro in queue. Implementation can we done in this case by deploying a surveillance over the head of people, still the people are not able to

viewed then we can add camera which can generate side-views of the people in queue which can be helpful to estimate number of people in queue, but several difficulties are related to such kind of deployment model e.g. occlusion as well as segmentation i.e. on which bases two persons are distinguished. By detecting only the head portions of a human, this eased at some extent.

- ***Crowd Behavior*** relates when inauspicious behaviour is emerged in civic area like accidents(bomb blasts, vehicles hits etc) which presents a new behaviour in the scene which need to be recognized and efficiently deal by the human operators to preventing time control situations that potentially lead to imperil incidents. Researchers found common abnormal crowd characteristics are fallen person [4]. A related surveillance problem in crowd behaviour is identifying specific individual events, where movements through some other substances in the scene will be a reason for significant clutter under which algorithms fails. Behavior classification is commonly based on the person(s) various movements instead of tracking a single person.

- ***Human Pose Estimation:*** In civic surveillance-based application, human pose estimation is about the movement of a human body (e.g., a person is lying down from standing is an indication of commuters fragmenting). The latest approach for this category is to recognizing an activity by an Energy function [4], which represents the activity level of a human. According to Yang, when a fighting is happened the number of frames increases for the movements, which significantly increases the motion's energy and system detects change as increase in number of frames in the defined time and to reflect is alarm is raised and it can be used in case of removal of occluded head in crowd and different general types of tracking breakdowns.

3.2 SEVERAL-PERSON INTERACTION

Several person interactions have been deeply prompted via increasing requirement in surveillance and security in order to recognize inauspicious actions. In [3], [4] and [6], the behaviour accedence apprehension process is based on foreground segmentation, tracking and head apprehension respectively. Such as, fight is described as multiple heads centre is collectively in motion.

Thus, integrating and dividing head and the whole rapid modifications in heads features; whereas assaults are described as when two heads reaching more near towards each other, among one another head supposed to be primarily stationary, and one another head is randomly in motion. On such type of theories we can recognize a multiple behavior of a human. Some techniques including the 'Nearest neighbour classifier' by using information related to human trajectory [4] to identify individuals involvements like human gathering, lurking, human falling down, and aggressive behavior.

Here we are ending with human behaviour accedence section and moving forward for future techniques to look at the state- of art and further developments and flaws in the discussed researches.

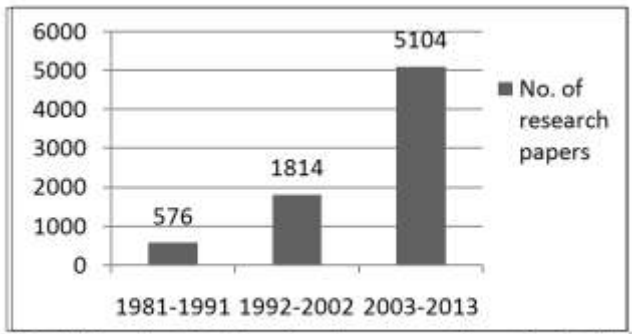

Fig. 3: Showing last three decades comparison of curiosity (on the basis of number of papers published in IEEE).

4. FUTURE DEVELOPMENTS

In this section, additional potentials work is surveyed which can be observed in figure 3. Additionally, a study on the existing modern techniques for behaviour understanding algorithm is offered and weakness related to core techniques are identified and feasible answers to the problems are described.

4.1 Core Techniques Limitations 4

A lot of restrictive reasons are spotted which are reducing the usability these core techniques in

actual civic service system based applications. Implementation of human analysts on video is practically not possible and also observed that algorithms can only be implied only on few datasets. Also, several equipments are associated like low frame rates, low resolution, or insufficient processing hardware. In [2], tracking is done through wearable accelerometer and sensor are placed in the area, which is not a feasible solution for tracking as you can't put everyone in a wearable accelerometer because every time a new person come in the environment and leaving as well.

If we proceed for [3], they are using various method only for extraction of substance (human) from the space on a static data which is not worthy in present scenario the demand is less complex and dynamic approachable methods while methodology opted in [4] is highly efficient as compare to above two as it works well in real time video but as person prototype is made by a half-terrestrial pattern, it introducing a fake person accedence. Last core technique [6], presented by Lu Xia only limitation that it is highly dependent on exact head accedence, it means that if do you to any reason the head is in occlusion or the human in the space wearing something on head , will possibly results that head is not detected.

5. CONCLUSION

Civic services bureaus are beneath demands to bestow a protected atmosphere for the public, in subway, bus, metros, and other transportation services. Most of the transportation bureaus are rapidly using visual based surveillance as a device to reduce offence, thwart violence, and enhance the security of commuters and employees. This survey presents modern developments on preprocessing of human behavior accedence termed as apprehension of human from the environment for civic services-based systems. Analysis includes apprehension, behavior, datasets, and implementation details. The behavior accedence is categorized as: a) person/no interaction and b) several person interactions.

We can say that today is need of 'Intelligent Surveillance Systems' which is capable of detecting human, recognizing inauspicious behavior and inform authorities to take primitive actions. However, still a big gap between analysts monitoring video data and state-of-art digital algorithms. Alternatively, there is an infinite effort needed to increase security workforce usefulness in extended intervals of time at the same time as well diminishing the cost on manpower. Although, numerous minds assume that computer technique is solitary solution to fill the space.

REFERENCES

[1] Official website of Delhi Metro Rail Corporation. [Online]. Available: www.delhimetrorail.com/

[2] Ikeda, Tetsushi, et al. "Pedestrian identification by associating wearable and environmental sensors based on phase dependent correlation of human walking." Journal of Ambient Intelligence and Humanized Computing (2013): 1-10.

[3] Rakibe, Rupali S., and Bharati D. Patil. "Background subtraction algorithm based human motion detection." International journal of Scientific and research publications, Volume3, May2013 (2013).

[4] Yang, Shuqiang, and Biao Li. "A novel and stable human detection and behavior recognition method based on depth sensor." 3D Research 4.2 (2013): 1-11.

[5] Marcialis, Gian Luca, Fabio Roli, and Gianluca Fadda. "A novel method for head pose estimation based on the “Vitruvian Man”." International Journal of Machine Learning and Cybernetics 5.1 (2014): 111-124.

[6] Xia, Lu, Chia-Chih Chen, and J. K. Aggarwal. "Human detection using depth information by kinect." Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on. IEEE, 2011.

[7] Spinello, Luciano, and Kai Oliver Arras. "People detection in rgb-d data."Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ

International Conference on. IEEE, 2011.

[8] Cerezo, Eva, et al. "Emotional facial sensing and multimodal fusion in a continuous 2D affective space." Journal of Ambient Intelligence and Humanized Computing 3.1 (2012): 31-46.

[9] Candamo, Joshua, et al. "Understanding transit scenes: A survey on human behavior-recognition algorithms." Intelligent Transportation Systems, IEEE Transactions on 11.1 (2010): 206-224.

[10] W. Frei and C. Chen, "Fast Boundary Detection: A Generalization and New Algorithm," IEEE Trans.Computers, vol. C-26, no. 10, pp. 988-998, Oct. 1977.

[11] Bhatia, Madhulika,Facial Expression Recognition System, Germany, LAP Lambert Academic Publishing, 2012.

BIOGRAPHIES

Monika: Graduated student of Master's degree for computer science & engineering in Amity University, major in artificial intelligence. I am graduated in computer science & engineering from MDU and Diploma in computer engineering from B.T.E.

Madhulika is working as a Assistant Professor in Department of Computer Science and Engineering at Amity University, Noida. She has total 9 years of teaching experience. She published almost 15 Research Papers in National, International conferences and journals. She is also Author of two Books. She filed one Provisional Patent also 2772/DEL/2013.