FUZZY PREPROCESSING OF VIOLA-JONES ALGORITHM FOR FACE RECOGNITION IN THERMAL IMAGES

Edwyn Martínez, Edmundo Bonilla, Eduardo Sánchez, Crispín Hernández and Roberto Morales

Tecnológico Nacional de México, Apizaco, México

ABSTRACT

This paper proposes to improve the Viola-Jones algorithm to identify facial features in thermal images using a fuzzy linguistic modifier. The purpose is to improve the image quality in the infrared plane and to facilitate the recognition stage, feature selection and emotion classification. The results obtained show that adding a preprocessing stage improves the capability of the Viola-Jones algorithm in identifying facial features in thermal images.

KEYWORDS

Emotion recognition, Edge detection, Viola-Jones, Image processing

1. INTRODUCTION

Emotion recognition is an activity that has been studied by different disciplines: psychology, sociology, medicine, philosophy, and human-computer interaction.

Different research has presented different methodologies to understand this activity, so they have implemented devices from different areas, some cases using electrocardiograms[1], CW Radars (Continuous Wave), RGB Cameras[2] and thermal cameras[3] to mention a few. Recently, the study of emotional states has been widely addressed by artificial intelligence, which is known as affective computing. This new methodology deals with the design of new electromechanical systems and devices that can recognize, interpret, and process emotions[4]–[8]

2. RELATED WORKS

Emotion recognition can be applied to different areas of technology, and therefore further development and/or advances can be obtained. One of the known processes consists of selecting the facial regions from which the necessary characteristics are extracted for better emotion recognition, in the article[9]they propose to make a personalized mark for each, they mention that it is possible to obtain "a fingerprint with the characteristics that can be collected in a facial thermal image".

Within their research they present that a thermal image can present problems due to the illumination that is presented in the background of the face, so they considered the light in the place where the tests were conducted, with this, they managed to make 3D models with which you can collect data from a thermal image.

Dhinaharan Nagamalai (Eds): CGASP, VCOI, BLKCA -2023
pp. 21-28, 2023. IJCI – 2023 DOI : 10.5121/ijcseit.2023.13502

The methods used for the recognition of emotions are mostly performed by means of RGB cameras in addition to considering the positions of those involved and the lighting within the sampling site, as this can define whether a sample is relevant or not. By means of this type of cameras[10]collect 2D and 3D plane information about the face of the sample group at a distance of 0.4 to 3.5 meters away from the RGB camera, as well as adjusting the tilt angle, placing 10 people with different characteristics and poses, they selected a tilt of 45 with respect to the head of each person.

The collection of information for emotion recognition is mostly focused on those characteristics that can be found in the facial area, so tools such as RGB and thermal cameras are used to obtain such information.

Although a good emotion recognition can be obtained by monitoring a person's face, there is research that proposes to monitor different areas of the body.By means of two cameras connected to a PC respectively,[11] captured the facial area and the rest of the body, in order to record the changes that occurred. With data from both cameras, they achieved a recognition rate of 91.1% by adding the values obtained.

In different investigations they have used different devices to extract features from one or different areas such as the face and the body, although they have presented extractions with two RGB cameras in these directions,[2]presented a combination between a camera and a CW radar (radar that transmits and receives continuous waves) in order to find a greater amount of movements that allow a better extraction of features.

For the correct collection, the participants' faces must be correctly illuminated, and the environment must be configured according to the same rule. This work obtained an accuracy with these methods of 89.30%.

With the knowledge acquired, the necessary stimuli were considered to find the selected emotional states, such mental states can be induced through the presentation of video clips[12]where scenes are shown that provoke the emotions found in Ekman's classification which are: fear, sadness, anger, joy, surprise and disgust. The selection of such material has had its respective changes throughout the years in which this type of experiments have been carried out, since the most recent situation in COVID-19 has set a new flag for these resources, due to the fact that the selection of audiovisual media that were already used as a reference previously may not have the same intensity when it comes to achieving the main objective.Likewise, the different backgrounds that may exist within the selected group of people affect the way in which the stimuli take effect[13]. In order to have a better selection of the material with which to induce the desired emotions, in addition to the elements presented in the images, videos and/or audios, elements such as resolution, brightness, field of vision and contrast should be taken into account, all this to achieve a better stimulation[14].

3. METHODS AND MATERIALS

3.1. Methodology

To carry out the process of this experiment, different variables were evaluated to obtain the best results. In a group of 25 engineering students at the Technological Institute of Apizaco composed of 12 women and 13 men with an age range of 20 to 26 years were exposed to visual stimuli to collect the necessary samples, for this the following points were prepared: the environment where the tests were collected, the material used to stimulate the different moods, as well as the necessary preparations before and after.

In the first point, an environment was configured using a laptop to control the order of presentation of the stimuli, a Samsung LED monitor and being complemented with headphones equipped with noise cancellation, in order to have a greater stimulus in the participants. To collect samples within the configured environment, two cameras were used, a thermal camera (Fluke thermal Camera Ti32) and an RGB camera (Sony Ciber-Shot DSC-H300).The thermal camera was placed at a distance of 1 meter from the back of the monitor to capture the changes in the face of the participants, and the RGB camera was placed to one side of it, which in addition to capturing the facial area, focused on the upper part of the body to capture the movements made by some participants. The participants were placed 50 centimetres away from the monitor in order to obtain a better view of the clips presented.

The sampling process was as follows:

- Take the participants to the place where they were previously prepared.
- Explain what sampling consists of.
- Give a short questionnaire to know the condition in which the participants were carrying, as well as ask for their authorization (by placing name and signature on the questionnaire) to use the image collected.
- Display the selected multimedia material.
- After showing the material, ask about the emotional state in which the participants were, in order to know if there was a change of emotion.
- Give thanks for participating.

Sampling was carried out from 10:00 to 13:00 approximately, for two reasons: it was considered that this is a time when students are in the best conditions to be exposed to emotional states such as stress and boredom; likewise, in this range of hours, daylight does not suffer significant changes, so, being close to a window, there were no changes in lighting in the samples collected.

3.2. Materials

The material to stimulate the different emotional states was taken from the LATEMO-E video base, which has a total of 70 video clips of emotions such as; disgust, anger, fear, sadness, fun, tenderness and neutral, these videos are fragments of movies from different countries dubbed into Spanish. To start with this selection process, the emotions that were to be induced were considered, in principle, those emotions and clips that had a higher percentage of the target emotion were taken into account. The table presented in the article [12]shows that the clips with the highest percentage of the target is fear with 100%, followed by neutral, anger, fun, sadness, disgust and tenderness with 96%, 93%, 91%, 90%, 90%, 90% and 79% respectively. Although emotions such as anger and disgust had a high percentage, the clips presented showed scenes that could be considered "strong" or "very explicit" for a certain group of people, so these emotions were discarded, the neutral emotion was not considered for this sampling since we wanted to find the changes that can be found in different emotions, so 4 emotions were selected; fun, fear, tenderness and sadness, taking those clips with the highest percentage in the target emotion and thus presenting it to the participants. The selected material is presented in the following table.

Table 1. Clips selected by emotion, as well as the success rate displayed in the database

Emotion	**Clip**	**Percentage of excitement achieved**
Fun	Blended (LMF)	85%
	The hangover (QPAI)	60%
	The hangover III (QPAIII)	87%
	The proposal (LPR)	78%
Fear	Mirrors (ESI)	93%
	The conjuring 2 (CON2)	100%
	The conjuring 3 (CON3)	75%
	The conjuring 4 (CON4)	93%
Tenderness	He’s just not that into you (SQN)	79%
	Her 2	53%
	Les choristes (LCO)	61%
	Pride and prejudice (OYP)	65%
	The notebook (DUPI)	70%
Sadness	My sister’s keeper (DMD)	89%
	Never let me go (NMA)	75%
	The boy in the striped pajamas (NPR)	90%
	The impossible (LIM)	85%

4. PROPOSED MODEL

To improve the process of identifying facial features in thermal images, the following model is proposed (Figure 2). In a first stage, the thermal image base is read; this process is described in detail in section III.I. For each image in the infrared plane, the 3 RGB color channels are converted into grayscale using the following formula:

$$Image_{Gray} = Canal_{Red} * 0.299 + \; Canal_{Green} * 0.587 + \; Canal_{Blue} * 0.114 \quad (1)$$

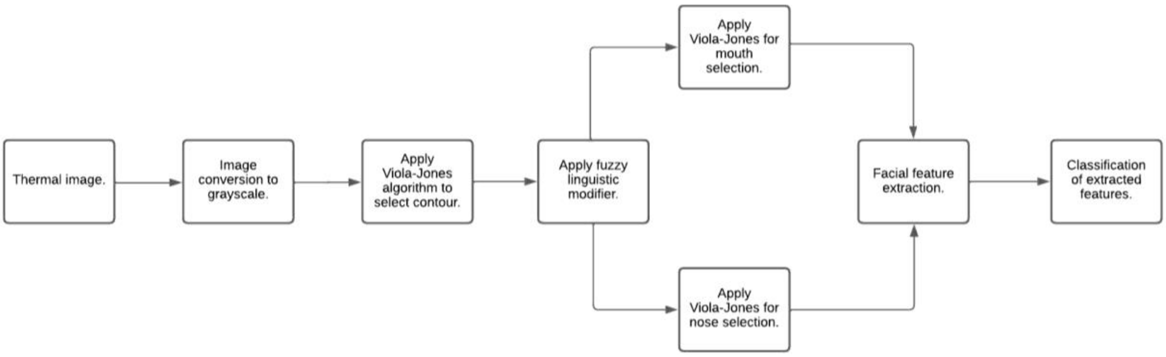
Thermal image.
Image conversion to grayscale.
Apply Viola-Jones algorithm to select contour.
Apply fuzzy linguistic modifier.
Apply Viola-Jones for mouth selection.
Apply Viola-Jones for nose selection.
Facial feature extraction.
Classification of extracted features.

Fig 1. Proposed model described by steps.

The next step is to apply the Viola-Jones algorithm to detect the contour of the person. Once the region of interest is located, which in this case is the face, we proceed to convert the image into interval [0,1] by dividing only the pixel values by 255, to apply the fuzzy linguistic modifierEc_{Image}:

$$Ec_{Image} = \min(INT\left(Image_{Gray}\right), 1 - INT\left(Image_{Gray}\right)^{6})) \quad (2)$$

Where INT is the intensification operator, which is defined as follows [15]:

$$\mu_{INT(A)}(x) = \begin{cases} 2(\mu_A(x)^2) & \mu_A(x)) \leq 0.5 \\ 1 - 2(1 - \mu_A(x)^2) & \mu_A(x) > 0.5, \end{cases} \quad (3)$$

The operator Ec_{Image} is intended to perform a double intensification to increase the contrast level in thermal images. The first one highlights the contrast in the image in gray levels, then a second one raised to the power 6 is applied to increase the image resolution. We take from these two intensified images the minimum membership value and obtain an equalized image that facilitates the Viola-Jones algorithm to identify other regions of interest such as the mouth, nose and eyes. In this paper we focus on the location of the nose and mouth regions. The nose to find a type of pathology associated with a POST-COVID effect and the mouth to classify basic emotions stimulated through audio-video, described in Section 3. 2.

The next stage is to extract facial features. For example, the Harris algorithm is applied to the mouth to find the most representative points of the mouth contour and evaluate how open it is to classify some type of emotion such as joy or fear.

5. RESULTS ANALYSIS

The following shows the results obtained with the proposal to apply a fuzzy linguistic modifier namedEc_{Image} the Viola-Jones algorithm is used in the preprocessing stage to improve the low contrast when working with thermal images.

Figure 3 shows the results of applying only the Viola-Jones algorithm once the face is located. The nose region is not located, and the mouth region is misidentified. As can be seen once the thermal image is converted into gray levels, it is difficult to locate the regions that make up the face such as the eyes, nose and mouth. The Viola-Jones algorithm fails 96% to locate a facial region. If it does detect, the window of the region of interest is in a different area, such as the hair region, cheeks, forehead or even eyebrows.

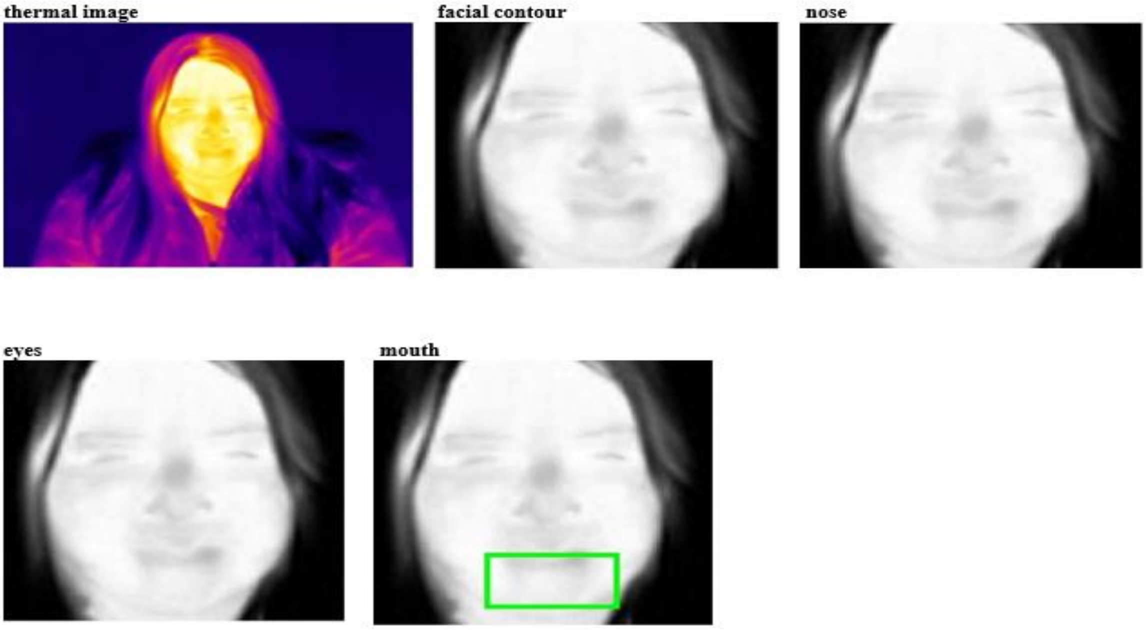

Fig 2. In this image Viola-Jones is unable to correctly detect mouth region.

Figure 3 shows the results of applying the proposed operator to improve contrast distribution in a thermal image. With this operator, 76% of the main facial regions are located: nose, eyes, and mouth. In other words, this percentage was obtained in 19 of the 25 images collected.

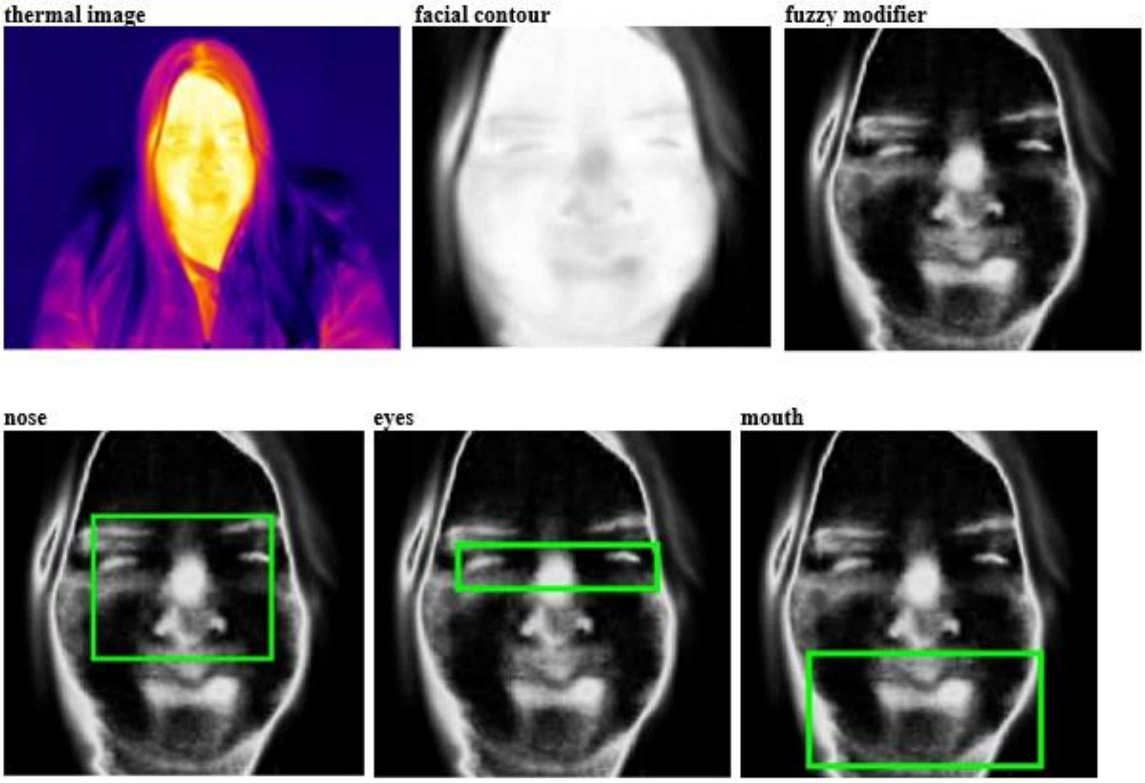

Fig 3. Result applying the fuzzy modifier as preprocesing and Viola-Jones algorithm for facial feature detection.

Of the 19 images from which the nose and mouth regions were located. A binarization of each region was applied to extract features. Figure 4 shows the binarized mouth region.

Fig 4. Binarization of region of interest: mouth.

Figure 5 shows the collection of features obtained with the Harris detector[16]. With this corner detector, it is applied to know the mouth opening, for these the main 18 features are selected and determine the height and proceed to classify the emotion. If the audiovisual stimulation was a happy clip, then it is determined that there has been a change in the corners of the mouth and the person is classified as reactive to the stimulus. If the mouth remains closed it indicates that the person has not been stimulated by the video clip.

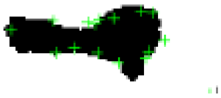

Fig 5. Main characteristics of the region of interest: mouth.

Based in these main characteristics, we extract points of height and width to calculate how much the mouth is opening and to determine if the subject is smiling or not.
For this task we define the following fuzzy rules:

- IF height IS low THEN emotion IS slight happy.
- IF height IS medium THEN emotion is More or Less Happy.
- IF height IS high THEN emotion is Fairly Happy.

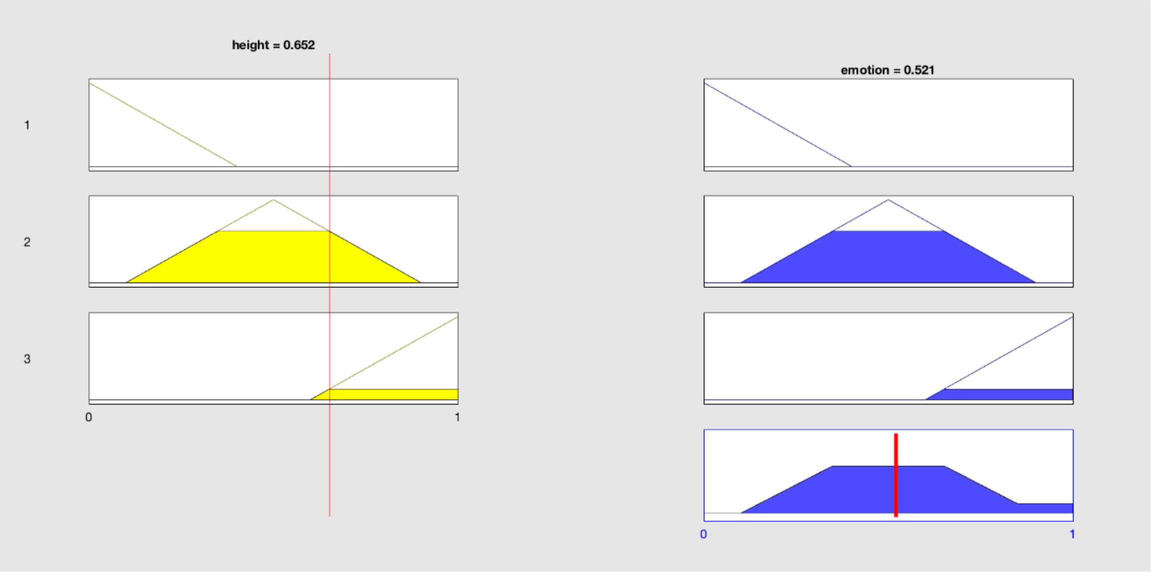

Fig 6. Fuzzy inference rules.

According to the height obtained from Figure 5 and using a fuzzy inference system (shown in the Figure 6) the subject, emotion is More or Less Happy.

6. CONCLUSIONS

Emotion recognition in the visible plane has been extensively studied for detecting facial features with the Viola-Jones algorithm; however, it has been little used with images in the infrared plane due to the limitations of the low contrast produced by thermal cameras. Thermal images can be used to analyse thermal variations that can help a classifier to better categorize facial features. In this paper we have proposed to solve the region identification problem with the Viola-Jones algorithm using a linguistic operator. The results obtained are promising and, in the future, we propose to use an emotion classifier algorithm such as a recommender system or a convolutional method to improve this activity.

REFERENCES

[1] N. S. Suhaimi, J. Mountstephens, and J. Teo, "A Dataset for Emotion Recognition Using Virtual Reality and EEG (DER-VREEG): Emotional State Classification Using Low-Cost Wearable VR-EEG Headsets," Big Data and Cognitive Computing, vol. 6, no. 1, Mar. 2022, doi: 10.3390/bdcc6010016.

[2] L. Zhang et al., "Non-Contact Dual-Modality Emotion Recognition System by CW Radar and RGB Camera," IEEE Sens J, vol. 21, no. 20, pp. 23198–23212, Oct. 2021, doi: 10.1109/JSEN.2021.3107429.

[3] R. Ashrafi, M. Azarbayjania, and H. Tabkhi, "A Novel Fully Annotated Thermal Infrared Face Dataset: Recorded in Various Environment Conditions and Distances From The Camera," Apr. 2022, [Online]. Available: http://arxiv.org/abs/2205.02093

[4] R. W. Picard, "Affective Computing." [Online]. Available: http://www.media.mit.edu/˜picard/

[5] J. Tao and T. Tan, "LNCS 3784 - Affective Computing: A Review," 2005.

[6] Y. Wang et al., "A Systematic Review on Affective Computing: Emotion Models, Databases, and Recent Advances," Mar. 2022, [Online]. Available: http://arxiv.org/abs/2203.06935

[7] L. Tian, S. Oviatt, M. Muszynski, B. Chamberlain, J. Healey, and A. Sano, Applied Affective Computing. Emotion-aware Human-Robot Interaction and Social Robots., 2022.

[8] T. Olugbade, L. He, P. Maiolino, D. Heylen, and N. Bianchi-Berthouze, "Touch Technology in Affective Human, Robot, Virtual-Human Interactions: A Survey," 2023.

[9] M. Akhloufi and A. Bendada, "Thermal faceprint: A new thermal face signature extraction for infrared face recognition," in Proceedings of the 5th Canadian Conference on Computer and Robot Vision, CRV 2008, 2008, pp. 269–272. doi: 10.1109/CRV.2008.43.

[10] Q. R. Mao, X. Y. Pan, Y. Z. Zhan, and X. J. Shen, "Using Kinect for real-time emotion recognition via facial expressions," Frontiers of Information Technology and Electronic Engineering, vol. 16, no. 4, pp. 272–282, Apr. 2015, doi: 10.1631/FITEE.1400209.

[11] H. Gunes and M. Piccardi, "Bi-modal emotion recognition from expressive face and body gestures," Journal of Network and Computer Applications, vol. 30, no. 4, pp. 1334–1345, Nov. 2007, doi: 10.1016/j.jnca.2006.09.007.

[12] Y. Michelini, I. Acuña, J. I. Guzmán, and J. C. Godoy, "Latemo-e: A film database to elicit discrete emotions and evaluate emotional dimensions in latin-americans," Trends in Psychology, vol. 27, no. 2, pp. 473–490, Jun. 2019, doi: 10.9788/TP2019.2-13.

[13] S. N. M. S. Ismail, N. A. A. Aziz, S. Z. Ibrahim, C. T. Khan, and M. A. Rahman, "Selecting Video Stimuli for Emotion Elicitation via Online Survey," Human-centric Computing and Information Sciences, vol. 11, 2021, doi: 10.22967/HCIS.2021.11.036.

[14] D. Gall and M. E. Latoschik, "Visual angle modulates affective responses to audiovisual stimuli," Comput Human Behav, vol. 109, Aug. 2020, doi: 10.1016/j.chb.2020.106346.

[15] V. N. Huynh, T. B. Ho, and Y. Nakamori, "A parametric representation of linguistic hedges in Zadeh's fuzzy logic," 2002. [Online]. Available: www.elsevier.com/locate/ijar

[16] S. Mistry and A. Patel, "Image Stitching using Harris Feature Detection," International Research Journal of Engineering and Technology, 2016, [Online]. Available: www.irjet.net