

## Smart Interrogation System by Detection of Visual Focus of Attention

P. P. Debnath<sup>1\*</sup>, M. G. Rashed<sup>2</sup>, D. Das<sup>2</sup>, M. R. Basar<sup>1</sup>

<sup>1</sup>Department of Information and Communication Engineering, Bangladesh Army University of Engineering and Technology, Natore-6431, Bangladesh

<sup>2</sup>Department of Information and Communication Engineering, University of Rajshahi, Rajshahi-6205, Bangladesh

Received 10 June 2020, accepted in final revised form 30 September 2020

### Abstract

The paper presents an approach to detect and control the focus of attention of the suspect using his/her eye gaze and head movement direction to build up an automatic interrogation system- specially to detect lies. To this point, we classified interrogation conversation into different criteria and identified the fatal ones. At first, we conducted psychological experiments on the sampled population to detect the different parameters connected with various symptoms when the suspect tells lies and build our knowledgebase with the results. This knowledgebase helps the system to make strategic decisions and to optimize accuracy. A monitoring camera captures continuous interrogation and feeds the frames to our proposed system. 3D head tracker is used to track the head from image and Active Shape Model (ASM) is utilized to localize face points. Vector Field of Image Gradient (VFIG) is calculated to track the eyeball and its rotation within the eye area. Random eye and head movement, change of eyebrow at the critical level of questionnaire provide us the possibility of detecting lies. Finally, experiments are conducted in a controlled environment to validate our psychological findings.

*Keywords:* Lie detection; Visual focus of attention; Eye center localization; Gaze detection; Sustained attention.

© 2021 JSR Publications. ISSN: 2070-0237 (Print); 2070-0245 (Online). All rights reserved.

doi: <http://dx.doi.org/10.3329/jsr.v13i1.47481>

J. Sci. Res. **13** (1), 47-58 (2021)

## 1. Introduction

Head and eye gaze behavior is a very interesting research topic for many years in psychology and in the computer vision community. Especially in the field of suspects' visual focus of attention (VOFA) [1,2] detection, this study has become a part of active research in order to detect lies [3]. However, there are very little works by which suspect's visual attention may be controlled accurately through human machine interface (HMI) and Human Posture Recognition [4]. Generally, humans have very poor ability in detecting

---

\* Corresponding author: [parthapatim.ice10@gmail.com](mailto:parthapatim.ice10@gmail.com)

deceit and hostile intent, with accuracy rate of 40-60 % [5]. However, when one is being deceitful, she/he is making up something in the brain. This results in increased brain activity and rapid physiological responses which can be measured on the face (including facial blood flow pattern). Humans do not possess the ability to control these physiological responses to emotions. Stress causes abrupt changes in local skin temperature and distinctive facial signatures which provides the main key to lie detection.

Some previous researches on lie detection systems used Polygraph [6-8]. It was developed in 1921 which measures and records several physiological indices such as blood pressure, pulse, respiration and skin conductivity and tries to find correlation between these measurements. However, it is highly invasive, very slow (requires several experts) and cannot be used in covert operations (where the suspect is unaware about the experiments.) At present there are several techniques in vogue [9-14]. But most of them use body sensors or high cost thermal cameras. Their detection accuracy is also low.

Humans have a well-defined “rigid” skull structure and facial muscle structure - this means there are finite numbers of facial expressions a human can perform. These are called Facial Action Units (FACS) and there are 46 such Action Units. To this point, our proposed technique finds out the best parameters of lie detection from visual domain, such as- Random Eye Movement, changes in the orientation of head and shape of eyebrow changes etc. Based on this, we shall create our knowledgebase to store how people reacted in different stimuli. This knowledgebase will help us to detect the phase of questionnaire (performed while interrogation) and its corresponding possibility of lie. A low-cost USB camera captures continuous video of the suspect while interrogation and feeds it to our proposed system. A multistage computer vision approach has been utilized to track the most critical parameters that show the symptoms of lie. Our proposed technique is cost effective and unbiased. It also provides proper video evidence of interrogation. The proposed system can be a helpful tool to the existing interrogation system.

## 2. Lie Detection

Human characteristics are very popular in nature. It is very tough to detect the insights of a human, especially when he is telling lies. For the case of an interrogation environment, the problem becomes tougher because the suspect keeps himself ready for some common question. In our approach, we classify the answers of the victim in four major criteria as follows-

Criteria (1, 1): The suspect has done it and also admits it. It would be considered as “**True.**” For example, if the suspect has stolen and told that “Yes, I have done it.”

Criteria (1, 0): The suspect has done it but denying it. It would be considered as “**False.**” For example, if the suspect has stolen and told that “No, I have not done it.”

Criteria (0, 1): The suspect has not done it but pretending that he has done it. It would be considered as “**False.**” For example, if the suspect does not have a red shirt but if he says “Yes, I have.”

Criteria (0, 0): The suspect has not done it but is denying it. It would be considered as “**True.**” For example, if the suspect does not have a red shirt and if he says “No, I don’t have.”

In our approach Criteria (1, 1) and Criteria (0, 0) are considered TRUE and hence, they are to be bypassed by the system. But Criteria (1, 0) and Criteria (0, 1) are considered as FALSE and must be tracked by the system.

Now when the interrogation begins, even the weakest liar can overcome our test without showing any external symptoms because at first s/he is ready to fight with some of the ready answers. However, interrogation means a chain of questions and to support one lie the suspect must tell another lie. With the course of time it becomes tougher for the suspect to answer with his/her ready wit. So, weak liars cannot sustain with the “Chain of Lie.” and shows some external symptoms. And the symptoms are very obvious for the criteria (0, 1). The reason behind this fact is that, if anyone claims to do something that s/he has not done, he needs to imagine that situation at first before answering about it. While imagination there is a change in body language showing symptoms of a possible lie. Our main goal is to track those symptoms to detect a lie. To this point we have classified the symptoms in three major criteria-

**Symptom Type-1:** The suspect provides a very random movement of his eye which is a clear indication of his nervousness. S/he also moves his eyeball to the upper right corner of his eye area as an indication that s/he is thinking of something.

**Symptom Type-2:** The suspect provides a head movement. In most cases for the weak liar, it is tough for him/her to continue the conversation with direct eye contact with the interrogator. Too much loss of direct eye contact is a clear indication of telling a lie. To avoid the direct eye contact, the suspect moves his/her head downward. Sometimes s/he also makes a frequent head movement.

**Symptom Type-3:** The suspect changes the shape of his eyebrow as an indication of his worry. Now if we can detect Symptom Type-1, 2 or 3, we can say that there is a possibility that the suspect is telling a lie. To confirm the possibility, we need to compare it with our knowledgebase. The knowledgebase is the storehouse of symptoms provided by different suspects at different levels of questionnaire obtained from our manual interrogation. Each of the questions in the questionnaire is marked as Phase 1, 2, ..., N etc. The symptoms offered by the suspects are also age variant. The teenagers are generally impatient and cannot remain calm and hence provide useful and clear as well as useless symptoms while interrogations. But the aged groups provide very low levels of symptoms which are very hard to track. So to build our knowledgebase to store how people react while telling a lie, we need to consider our samples from different age groups. The

statistics from the samples will help us to get the threshold value to confirm a lie. Since, human emotion is the prime parameter of our lie detection strategy; we classify our crime pattern as follows-

**Crime Type-1:** Emotion is connected with the crime. In this situation there is no economic profit. Any type of revenge or avenge (for example killing for extra marital relationship) can be considered as Type -1 crime. Here the suspects show more symptoms of telling a lie.

**Crime Type-2:** The crime is done only for economic profit, there is no emotional issue. Any type of fraud activities such as stealing, cheating, abduction, mugging etc. can be considered as crime Type 2. Here the suspects show less symptoms. A questionnaire for this type of crime (for stealing in this case) may be as follows:

Table 1. Questionnaire for Crime Type 2.

Phase	Possible Question	Possible Answer	Symptom Possibility
1	Have you done this?	No	NO
2	Where were you last day?	At my village home	LOW
3	What were you doing at 10.00am?	Playing cards	LOW
4	Who were with you?	My other cousins.	NO
5	What does your cousin do?	They study in school.	NO
6	Last day was not a vacation. What were they doing with you in school time?	They had a fever and did not go to school.	HIGH
7	Both of them had a fever?	May be	HIGH

Now we see that the suspect provides external symptoms at different levels of questions. The changed focus of attention offered by the suspect is our key to lie detection. The overall system is depicted in Fig. 1.

### 3. Visual Focus of Attention Detection

#### 3.1. Method overview

At first a camera or sensor is needed to detect the VFOA of the suspect. Then the captured video frames are fed one by one to an intelligent system. Thereafter, each video frame is converted into grey scale for easier processing and analyzed each video frame pixel by pixel. These analyzed data is used to detect the loss of attention if the suspect changes his/her VFOA to another direction. How the head direction of the suspect is changed while telling a lie is observed from our psychological tests. The minimum deviation of the head orientation of the suspect can be considered as a clue to detect lie because suspects change the tilt/pan angle of their head while telling a lie. A step by step approach to detect the head movement and Random Eye Movement (REM) as discussed below.

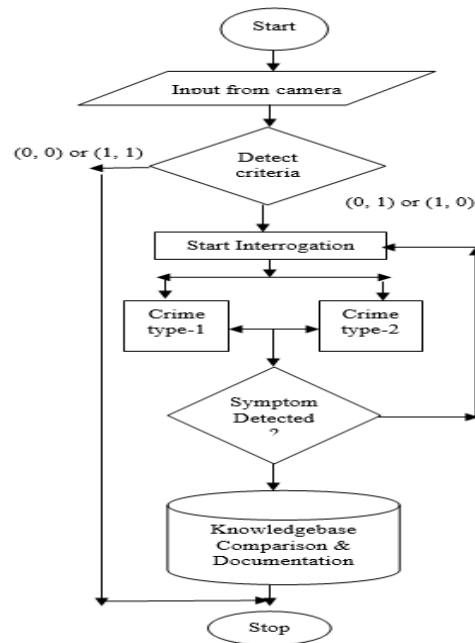


Fig. 1. Programming flow chart.

### 3.1.1. Head pose detection

The main goal of head pose detection is to track the head from the continuous image whether the head is in movement or not. In our present work, we have used the Seeing Machine’s faceAPI [15] to detect and track the head pose,  $h_p$  of the target person. To detect the head, we have used a haar cascade classifier as a 3-D head tracker [16]. We draw a rectangle outside of the head. For Symptom Type-2, the suspect moves the head within this head rectangle. The orientation of the middle of the face line within this face rectangle clearly defines the head movement at a particular direction as proposed in [17]. If the suspect turns his face downwards without moving his head then the change of the orientation of the nose point with respect to the face rectangle is considered as an alternative to measure the Symptom Type-2.

### 3.1.2. Face points extraction by active shape model

Our modelling method works by examining the statistics of the coordinates of the labeled points in the head rectangle. In order to be able to compare equivalent points from different shapes, they must be aligned with respect to a set of axes. We achieve the required alignment by scaling, rotating and translating the shape so that they correspond as closely as possible. In our presented approach, it is desirable to minimize a weighted

sum of squares of distances between equivalent points on different shapes. Finally, the facial feature points are extracted from the active shape model [18].

### 3.1.3. *Iris center detection*

The facial feature points are utilized to detect the eye regions roughly from the face. The VFIG is used to detect the iris center. The VFIG iris center detection technique is described as follows:

Let  $I_c$  be the possible iris center and  $I_{gi}$  be the gradient vector in position  $I_{xi}$ . If  $I_{di}$  is the normalized displacement vector, then it should have some absolute orientation as the gradient  $I_{gi}$ . We can determine the optical center  $I_c^*$  of the iris (darkest position of the eye) by computing the dot products of  $I_{di}$  and  $I_{gi}$  and finding the global maximum of the dot product over the eye image:

$$I_c^* = \operatorname{argmax}_{I_c} \left\{ \frac{1}{N} \sum_{i=1}^N (P) \right\} \quad (1)$$

$$\text{Where, } P = (I_{di}^T I_{gi})^2$$

$$I_{di} = (I_{xi} - I_c) / (\|I_{xi} - I_c\|_2)$$

$i = 1, 2, \dots, N$  and the displacement vector  $I_{di}$  is scaled to unit length in order to obtain an equal weight for all pixel positions in the image.

We create an eye rectangle around our eye. The fluctuation of the coordinate of the eyeball within this eye rectangle provides us a clue transient attention detection as proposed by Debnath [17] and thereby, also provides a clue to Symptom Type-1.

### 3.1.4. *Eyebrow Movement Detection*

The American psychologist Gibson gave the concept of optical flow in 1940 [19]. To materialize the visual stimulus provided to animals, he considered the pattern of apparent motion in a visual scene. The considered motion is caused by the relative speed between the observer and the scene. To estimate the optical flow of the eyebrow we need the sequences of the ordered images. This approach tries to evaluate the motion between the two consecutive images taken at time  $t$  and  $t + \Delta t$ . In our approach, we have used the Lucas-Kanade method [20]. For a pixel under consideration, it considers the optical flow constant in its neighborhoods. Using the least square criterion, it solves all the optical flow equations for all of the neighborhood pixels. To distinguish among different image elements, we have used The Shi and Tomasi corner detection algorithm [21]. In our proposed work, the eye area has been considered the Region of Interest (ROI), where the change of the shape of the eyebrow is detected.

### 3.2. Field of view of the suspect

The field of view (FOV) of the suspect is divided into three regions:

- *Central Field of View (CFV):*

This FOV exists at the center of the human FOV. This zone is set to a 30° cone shaped area (75° to 105°).

- *Near Peripheral Field of View (NPFV):*

It is defined as the 45° fan shaped area on the both sides of CFV zones. At the right side of CFV (30° to 75°) it is defined as the right near peripheral field of view (RNPFV) and at the left right side of CFV (105° to 150°) it is defined as the left near peripheral field of view (LNPFV).

- *Far Peripheral Field of View (FPFO):*

This FOV exists on both sides at the edge of the human FOV. The right side of the RNPFV (-35° to 30°) is known as the right far peripheral field of view (RFPFV) and on the left side (-145° to 150°) is known as left far peripheral field of view (LFPFV).

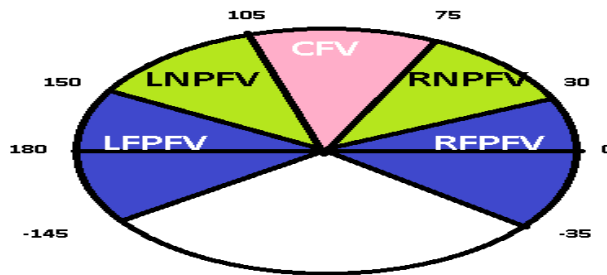


Fig. 2. Field of View (FOV) of the suspect.

## 4. Data Collection and Experiments

### 4.1. Data collection for knowledgebase creation

People of different ages show different levels of body language while telling lies. To validate it, we collected data with different age groups. There were a total of 20 participants. We divided them according to their age into four consecutive groups as follows:

Group-1 (10 to 14 years old): Four members with the age 11, 11, 12 and 13 years, respectively.

Group-2(15 to 19 years old): Four members with the age 16, 16, 18 and 19 years, respectively.

Group-3(20 to 24 years old): Four members with the age 23, 24, 24 and 24 years, respectively.

Group-4(25 to 29 years old): Four members with the age 25, 26, 27 and 28 years, respectively.

They were interrogated with the same questionnaire (priorly unknown to them) and instructed them to tell lies instantly. The resulting symptoms were tracked very carefully. The summarized results of age variant symptom analysis are given in Table 2.

Table 2. Age variant symptom analysis.

	Head Rotation	Eye Rotation
Group 1	3	2
Group 2	3	1
Group 3	2	2
Group 4	1	0

We conducted several experiments to detect the level of questionnaire (Phase) at which the suspect shows external symptoms. There were 3 unpaid participants (2 male, 1 female) who were staffs at Bangladesh Army University of Engineering and Technology (mean age 27 years and standard deviation 0.775 years). To avoid the influence of the interrogator (which may change the possibility of exhibiting external symptoms), we used Zoom application software to interrogate the subjects online. Some images taken while experiments are given below-



Fig. 3. Experiments to track eyeball rotation to detect Symptom Type-1.



They were interrogated with the same questionnaire and the external symptoms (rotation of head, eye, change of eyebrow) were monitored very carefully. We collected as follows.

Table 3. External symptoms detection.

	Suspect 1	Suspect 2	Suspect 3
Phase 1			
Phase 2	1	1	0
Phase 3			1
Phase 4			
Phase 5	2	1	2
Phase 6	1	1	1
Phase 7	2	1	1

To trace the relationship between Crime Type-1 and 2 with emotion, we conducted two separate experiments in the controlled environment. We set the threshold value of sustained symptoms to be greater than 3 sec. Less than 3 sec will be considered as transient symptoms. The summary of the collected data is given in Table 4.

Table 4. Crime Type 1 and 2 detections.

		Suspect 1	Suspect 2	Suspect 3
EXP 1	Transient	1	0	1
	Sustained	2	1	2
EXP 2	Transient	0	1	1
	Sustained	1	1	0

#### 4.2. Data collection for symptom detection

To validate the performance of the system in controlled environment, we went through several experiments. At first to detect the eyeball rotation within the eye area, the participants were asked to look within the LNPFV and RNPFV with keeping their head in stable state, that means in CFV with varying duration. The illumination of the room was 200 Lux and the distance from the camera was 0.5 m. After that, to track the head rotation, they were asked to rotate their head with varying rotation time i.e. 1, 2 and 3 sec respectively. They were also asked to move their eyebrow (shrink or expand it) so that it can be tracked by the optical flow feature. The average video duration is 3 min with average 12 frames per sec.

### 5. Performance Evaluation

In our lie detection technique, we detected Symptom Type-1 (st1) based on random fluctuation of eye movement. The Accuracy of detection while interrogation is expressed as-

$$\text{Accuracy}_{st1} = \frac{\text{Total number of eye movement tracked}}{\text{Total number of eye movement occurred}} \quad (1)$$

Based on equation (1), we get the Fig. 4 which shows the eye rotation detection accuracy for varying eye rotation time. Similarly, Symptom Type-2 is tracked based on random head movement or change in the orientation of head. The accuracy is expressed as:

$$\text{Accuracy}_{st2} = \frac{\text{Total number of head movement tracked}}{\text{Total number of head movement occurred}} \quad (2)$$

Based on equation (2), we get the Fig. 5 which shows the head rotation detection accuracy for varying head rotation time. Symptom Type-3 is tracked with the change in the shape of the eyebrow by implementing Optical Flow Feature in the ROI. The detection accuracy is expressed as:

$$\text{Accuracy}_{st3} = \frac{\text{Total number of eyebrow movement tracked}}{\text{Total number of eyebrow movement occurred}} \quad (3)$$

The accuracy of Symptom Type-3 has been illustrated in Fig. 6.

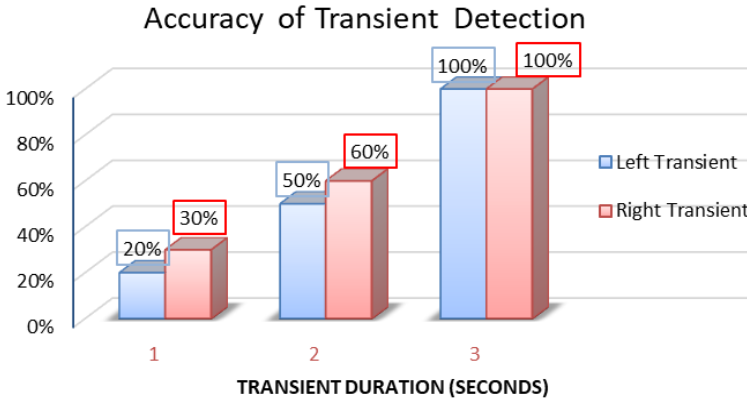


Fig. 4. Accuracy measurement of Symptom Type -1 detection.

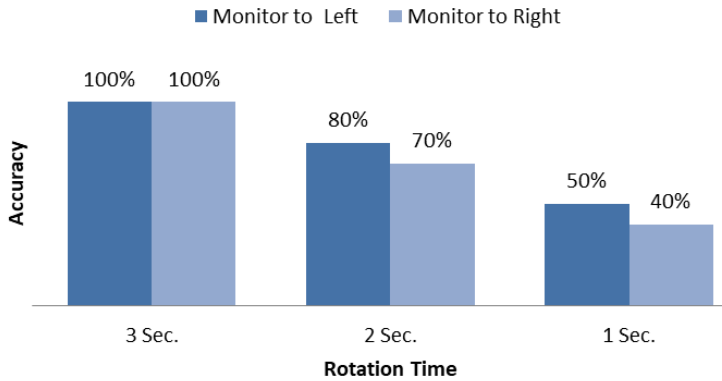


Fig. 5. Accuracy measurement of Symptom Type -2 detection.

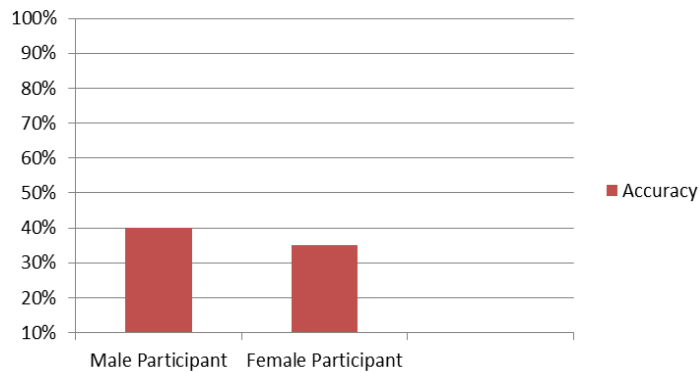


Fig. 6. Accuracy measurement of Symptom Type -3 detection accuracy.

## 6. Conclusion

We aimed at detecting lies to build up an automatic interrogation system by tracking the visual focus of attention of the suspect. At first, we created a knowledgebase gleaned from manual experiments to define different parameters of suspicious behavior during telling lies and to optimize the accuracy. From the experimental results it can be concluded that the accuracy of Symptom Types-1 and -2 is dependent on transient duration. And the accuracy increases linearly with the duration. However, Symptom Type-3 detection accuracy depends on the gender of the suspect. The reason is that the width of the female suspect's eyebrow is less than that of the male suspect. So, it becomes harder to track the change in shape of eyebrows using optical flow features. We conducted our experiments in a controlled environment which may not match with the results of real-life environments. We leave this for our future research.

## Acknowledgment

The authors would like to thank the ICT Division, Ministry of Post, Telecommunication and Information Technology, Bangladesh, for their research grants.

## References

1. T. Judd, K. Ehinger, F. Durand, and A. Torralba, 2009 IEEE 12th Int. Conf. Computer Vis. ID 16445820 (2009).
2. H. Zhang, M. R. Smith, and G. J. Witt, Human Factors **48**, 805 (2006). <https://doi.org/10.1518/001872006779166307>
3. M. Vicianova, Eur. J. Psychol. **11**, 522 (2015). <https://doi.org/10.5964/ejop.v11i3.919>
4. M. Z. Uddin and M. A. Yousuf, J. Sci. Res. **7**, 11 (2015). <https://doi.org/10.3329/jsr.v7i3.19527>
5. J. K. Burgoon, D. B. Buller, A. S. Ebesu, and P. Rockwell, Communicat. Monographs **61**, 303 (1994). <https://doi.org/10.1080/03637759409376340>
6. J. E. Reid and F. E. Inbau, Truth and Deception-The Polygraph ('Lie-Detector') Technique, 2<sup>nd</sup> Edition (Williams and Wilkins Co., USA, 1977).

7. M. R. Bhutta, M. J. Hong, Y. H. Kim, and K. S. Hong, *Front. Psychol.* **6**, 709 (2015).  
<https://doi.org/10.3389/fpsyg.2015.00709>
8. National Research Council, *The Polygraph and Lie Detection* (National Academies Press, 2003).
9. Y. V. Patil, O. A. Salvi, P. R. Waghmare, A. D. Kondilkar, and V. P. Kadroli, *Int. J. Eng. Develop. Res.* **5**, 47 (2017).
10. G. K. Chavali, S. K. N. Bhavaraju, T. Adusumilli, and V. Puripanda, Masters Thesis, Blekinge Institute of Technology, Sweden (2014).
11. M. Owayjan, A. Kashour, N. Al Haddad, M. Fadel, and G. Al Souki, The Design and Development of A Lie Detection System using Facial Micro-expressions – *Proc. 2nd Int. Conf. on Adv. in Computat. Tools for Eng. Applicat.* (IEEE, Lebanon, 2012) pp. 33-38.  
<https://doi.org/10.1109/ICTEA.2012.6462897>
12. J. Gao, Z. Wang, Y. Yang, W. Zhang, C. Tao, J. Guan, and N. Rao, *PLoS One* **8**, ID e64704 (2013). <https://doi.org/10.1371/journal.pone.0064704>
13. Y. F. Lai, M. Y. Chen, and H. S. Chiang, *Granular Comput.* **3**, 169 (2018).  
<https://doi.org/10.1007/s41066-017-0064-3>
14. G. S. Agrawal, D. R. Koche, H. D. Nagrale, and V. R. Surjuse, *Rev. Lie Detect. Syst.* **6**, 1118 (2019).
15. S. Machines, faceAPI-The Real-Time Face Tracking Toolkit for Developers and OEMs (2008).
16. L. M. Brown, 3D Head Tracking using Motion Adaptive Texture-Mapping - *Proc. Computer Soc. Conf. on Computer Vis. and Pattern Recognition* (IEEE, USA, 2001) pp. 1-1.
17. P. P. Debnath, A. R. Hasan, and D. Das, Detection and Controlling of Drivers' Visual Focus of Attention - *Proc. Int. Conf. on Electrical, Computer and Communicat. Eng.* (ECCE, 2017) pp. 301-307. <https://doi.org/10.1109/ECACE.2017.7912922>
18. T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, *Computer Vision Image Understanding* **61**, 38 (1995). <https://doi.org/10.1006/cviu.1995.1004>
19. B. K. Horn and B. G. Schunck, *Artificial Intelligence* **17**, 185 (1981).  
[https://doi.org/10.1016/0004-3702\(81\)90024-2](https://doi.org/10.1016/0004-3702(81)90024-2)
20. B. D. Lucas and T. Kanade, An Iterative Image Registration Technique with an Application to Stereo Vision - *Proc. DARPA Image Understanding Workshop* (Carnegie Mellon University, USA, 1981) pp. 121-130.
21. J. Shi and C. Tomasi, Good Features to Track - *Proc. IEEE Conf. on Computer Vis. and Pattern Recognition* (IEEE, USA, 1994).