JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

EYE CONTACT MAINTENANCE IN VIDEO CONFERENCES

[1]Kirthika K M, [2] Sanjay M, [3] Sneha M, [4] Sri Hari Raj R

[1]Assistant Professor, [2] Student, [3]Student, [4]Student,
[1,2,3,4]Computer Science and Engineering,
[12,3,4]Sri Ramakrishna Institute Of Technology, Coimbatore, India

Abstract : Video-conferencing systems were invented for the purpose of long-distance communication and have become widely accessible due to the Internet and the growth of electronic devices such as PCs. However, the problem of the face of the user not being aligned correctly in a conference still requires a solution. Many facial analysis tasks like detection or recognition can be used to rectify this fault. These tasks require the use of facial landmarks. Media Pipe, a Machine Learning-based framework can generate landmarks and can perform these tasks. In this article we propose a simple solution which is based on using the media pipe framework, which tracks the movements of the iris and gives helpful alerts to improve the subjective feeling of eye contact.

Keywords: *Machine Learning, Graphics Processing Unit, Single Shot Detector Application Programming Interface Exchangeable Image File Format Computer Vision*

I. INTRODUCTION

As of recent times, online video conference has proven to work better in establishing communication people in remote locations, which would otherwise would not have been possible. However, multiple unforeseen technical events can occur that can reduce the experience of the conference. Some of the issues can be alluded to the fact that the camera position is not fixed and standardized, as the position may vary upon any given device. These sorts of issues can cause a lack of eye contact between the participants of the users, essentially, putting us in a position where we are still not able to express our thoughts to their fullest extent in a conversation. Many plans had been developed to solve this issue, mostly depending upon the reliance of AI. But most of them had failed, due to the unpredictable position a camera has on a device. To overcome this situation, our proposed solution is to use a ML-based framework Media Pipe, that will track face visibility alongside iris movement and guide the under into maintaining steady eye contact. During the video conference, the user will feel they are experiencing a conversation that feels more authentic. Users on all ends will not feel uncomfortable, the constant eye contact that is maintained produces an easier conversation to carry out. Thus, eye contact is necessary for a conversation and in our proposed we suggest that we suggest that we need an application or protocol that maintains eye contact, which results in the process of reducing the issues that can occur in a video conference.

II. PROBLEM STATEMENT

Maintaining eye contact in a conference is necessary for a conversation. However, when people engage in a conference, it never feels genuine. This is due to the uncertain position of the camera and the face. Thus, we have to create an application or protocol to support eye contact maintenance.

III. SCOPE OF THE PROJECT

The different locations of video camera and display makes eye contact between users near impossible. A combined lack of life-sized video image and mutual gaze makes the quality of video conferencing an artificial experience(i.e.) having a lack of decreased communication quality. Although solutions have been suggested in the past, to overcome the obstacles of mutual gaze and a lack of life-sized video image, our system tracks the center of the eyes of the user and gives instructions to improve the subjective feeling of eye contact. This is done through Machine Learning based-framework Media pipe. Through the help of this framework, eye contact correction is carried out, which is needed to work when a user is engaged in the conversation, rather than when they naturally take their eyes off the screen

IV. LITERATURE SURVEY

[1]Leanne Bohannon al, "Effects of video-conferencing on gaze behavior and communication Effects of video-conferencing on gaze behavior and communication" 2010 Through the use of eye-tracking and conversation analysis this study examined the impact of video-conferencing on communication. Paired participants performed a collaborative task over four communication media: face-to-face; desktop video-conferencing with eye contact; and life-size video-conferencing with and without eye contact. The system allows attentive compression by reducing resolution of video users that are not being looked at Participants more frequently checked

the information their partner verbally relayed when communicating face-to-face and over life-size video-conferencing with eye contact.

[2]Romaine Belmonte, et al, "Video-Based Face Alignment With Local Motion Modeling", 2019 Face alignment remains difficult under uncontrolled conditions due to many variations that may considerably impact facial appearance. Recently, video-based approaches have been proposed, which take advantage of temporal coherence to improve robustness. These new approaches suffer from limited temporal connectivity. We show that early, direct pixel connectivity enables the detection of local motion patterns and the learning of a hierarchy of motion features.

[3]Grayson& Monk, et al, "Implementing Eye-to Eye Contact in Life-Sized Videoconferencing", 2003
People are using videoconferencing frequently; they might learn to interpret gaze direction to a very high degree of accuracy if the equipment is configured optimally. This is helpful when addressing objects in the environment, but does not provide the perception of eye-to-eye contact. These factors have big importance that is observer distance, head orientation, visibility of the eyes, and the presence of a second head on the perceived direction and width of the gaze cone in videoconferencing. Also, we are less sensitive to eye contact when people look below our eyes than when they look to the left, right, or above our eyes. That is, we will think that someone is making eye contact with us, unless we are certain that the person is not looking into our eyes. Those aspects help to mitigate the effects of the lack of (well configured) videoconferencing.

[4]László A Jeni, et al "The First 3D Face Alignment in the Wild (3DFAW) Challenge", 2016 2D alignment of face images works well provided images are frontal or nearly so and pitch and yaw remain modest. In spontaneous facial behavior, these constraints often are violated by moderate to large head rotation. 3D alignment from 2D video has been proposed as a solution. A number of approaches have been explored, but comparisons among them have been hampered by the lack of common test data. To enable comparisons among alternative methods, the 3D Face Alignment in the Wild (3DFAW) Challenge, presented for the first time, created an annotated corpus of over 23,000 multi-view images from four sources together with 3D annotation, made training and validation sets available to investigators, and invited them to test their algorithms

[5]Roel Vertegaal, et al, "GAZE-2: Conveying eye contact in group video conferencing using eye-controlled camera direction", 2003 GAZE-2 is a novel group video conferencing system that uses eye-controlled camera direction to ensure parallax-free transmission of eye contact. To convey eye contact, GAZE-2 employs a video tunnel that allows placement of cameras behind participant images on the screen. To avoid parallax, GAZE-2 automatically directs the cameras in this video tunnel using an eye tracker, selecting a single camera closest to where the user is looking for broadcast. Images of users are displayed in a virtual meeting room, and rotated towards the participant each user looks…

[6]Boyi Jiang, et al, "Deep Face Feature for Face Alignment and Reconstruction", 2018 In this paper, we propose a novel face feature extraction method based on deep learning. Using synthesized multi-view face images, we train a deep face feature (DFF) extractor based on the correlation between projections of a face point on images from different views. A feature vector can be extracted for each pixel of the face image based on the trained DFF model, and it is more effective than general purpose feature descriptors for face-related tasks such as alignment, matching, and reconstruction. Based on the DFF, we develop an effective face alignment method with single or multiple face images as input, which iteratively updates landmarks, pose and 3D shape. Experiments demonstrate that our method can achieve state- of-the-art results for face alignment with a single image, and the alignment can be further improved with multi-view face images.

[7]Aleš Jaklič, et al, "User interface for a better eye contact in videoconferencing", 2017 In this era of information explosion, network video conference systems allow for long-distance communication. Accordingly, Yuzhen proposed a subjective feeling scheme to improve eye contact, with the aim of addressing the sense of distance associated with network-based interactions. The network display mode is crucial because it can provide a multimedia webpage with enhanced interactive effects. In this project, there seems to be a lag in face sync.

[8]Andrew Herbert, et al, "Eye contact and video-mediated communication: A review", 2013 It presents a set of algorithms and an associated display system capable of producing correctly rendered eye contact between a three-dimensionally transmitted remote participant and a group of observers in a 3D teleconferencing system. The participant's face is scanned in 3D at 30Hz and transmitted in real time to an auto stereoscopic horizontal-parallax 3D display, displaying him or her over more than a 180◦ field of view observable to multiple observers.

V. EXISTING SYSTEM

Video conferencing has now become mainstream via through PCs and smartphones due to it's ability to establish face to face communication between users in remote areas. Yet, the general method of a video conference system does not fully allow us to communicate ourselves effectively. Multiple attempts have been made to rectify this issue, but none had a favorable result, as they could not simulate the experience a real life meeting would have. A team of researchers in in the Faculty of Computer and Information Science at the University of Ljubljana had proposed address the eye contact issue. The user's video stream will be rotated on the x-axis accordingly to the tilt of the display. When we distance ourselves from the device, the rotation reacts to tilting of the video-image plane, achieving eye-to-eye communication.

VI. PROPOSED SYSTEM

The different locations of video camera and video display make it impossible to make direct eye contact. This issue is known as the lack of mutual gaze. Our idea is to solve these in our modified video calling application. The idea is to use an ML(Machine Learning) based framework, Media Pipe, to track the eye movement of the user by using a face mesh, which is generated by visualizing landmarks around the face of the user. An iris tracking model in the system uses the face mesh to track eye movement. If the user is maintaining proper eye contact, the system remains dormant. Otherwise, the system will notify the user with a screen notification. Thus, the ML powered framework achieves the effect of face alignment in real-time during the video conference. Our solution will track both face visibility and iris movement, which guide the user with both audio and visual alerts to maintain good eye contact throughout while the video conference is being conducted.

There would be three outputs to our project that would be:

·Face and iris tracking both in their ideal position, in this case there would be no alerts as the output is the desired output.

·Face is detected but the iris is moving (i.e., iris is moving left or right and not present in the ideal position to maintain proper eye contact), in this case our system would detect it immediately and alert the user using both audio and text alerts, stating to "see camera".

·When both face and iris is not present within the visibility of the camera, in this case the system will notify the user with both an audio and text alert stating that the "face is not detected".By notifying the user when their eye contact is not correctly maintained, it will prompt the user to correct their posture, resulting in the effect of face alignment.

VII. METHODOLOGY

The idea is to track the eye movement on the camera screen, permitting the user to concentrate on the screen, achieving the result of harmonious eye contact without any obstacles while a video conference is being. When elaborated, we can say that there will be an iris- tracking model, a method used to track the movement of the eyes, which is done with the help of the media pipe framework. the media pipe framework helps us to track the iris movement in any direction similar as right or left or up or down. This is done via a Face Mesh that's created by the method. The Face Mesh is a three dimensional model that's generated by the method to detect the face of the user. This allows for facial landmarks to detected. In this case, the landmarks around the iris of the eyes are detected. We can understand how the varied eye movements or position of iris affect the user's appearance in a video call, through the pictures below.

Fig 7.1: describing the effects of various eye movements

The ideal position of the eye's iris is to see through the camera in order to maintain a proper eye to eye contain with the other user in the video call. The mechanism will constantly watch the activity of the eyes. However, the system will notify the user right away, If any issues occur in the maintenance of consistent eye contact. This process is done through the landmarks around the iris of the eyes that are constantly being watched by the system. The mechanism recognizes the eye movement through these landmarks and is mindful of the correct eye posture wanted for video communication.

The pipeline is applied as a Media Pipe graph that uses a face landmark subgraph from the face landmark module, an iris landmark subgraph from the iris landmark module, and renders using a devoted iris-and- depth renderer subgraph, which uses a face finding subgraph from the face detection module. The first step in the pipeline leverages our former work on 3D Face Meshes, which uses high- dedication facial landmarks to induce a mesh of the approximate face figure. From this mesh, we seclude the eye region in the original image for use in the iris tracking model. The problem is also divided into two region eye figure estimation and iris position

Building on our work on Media Pipe Face Mesh, this model is able to track landmarks involving the iris, pupil and the eye contours using a single RGB camera, in real-time, without the need for specialized hardware. Through use of iris landmarks, the model is also able to determine the metric distance between the subject and the camera with relative error less than 10% without the use of depth sensor. .

4.5.1 INSTRUMENTATION

- All dataset images were captured on a diverse set of smartphone cameras, both front- and back-facing.
- All images were captured in real-world environments with different light, noise and motion conditions via an AR application.

4.5.2 ATTRIBUTES

The model is intended to be used primarily in the tracking mode that guarantees certain accuracy of the eye location, scale and rotation (see specification in "Attributes").

- Eye region cropped from the captured frame should contain a single left eye with eyebrow placed in the center of the image.
- There should be a margin around the eye region calculated as 25% of eye region size.

Margin should be applied to a minimal proportional bounding box enclosing eye and eyebrow.

- Image must be rotated in a way that a horizontal line can connect the two corners of the eye.
- Model is tolerant to certain level of input inaccuracy:
    1. 10% shift and scale (taking eye region width/height as 100% for corresponding axis)
    2. 8° roll

4.5.3 INPUTS

Videos should be captured in" selfie" mode. As alike, it's not suitable for detection when

1. Face is directed away from the camera (further than 60 ° when eye is no longer visible),
2. Face is inclined from the vertical aspect (further than 8 °),
3. Eye is only halfway visible (lower than 50) or no iris (eye is closed),
4. Eye is located too far off from the camera .

4.5.4 ENVIRONMENT

When demeaning the environment conditions (very dark or noisy camera video, video with a lot of movement or significant eye overlap) one can anticipate declination of quality and increase of " jittering" (although we cover similar cases during training with real- world samples and augmentations). Model is trained on images with varied lighting, noise and movement conditions and with different augmentations. Still, its quality can degrade in extreme conditions (specified in" Limitations Environment"). This may lead to increased " jittering" (inter-frame predicting noise).

4.6 The alerts:

If the user properly looks at the screen, then they will not be notified by any messages from the system, as there are no faults in the video communication. However, if the user is not giving proper eye contact towards the video communication, they shall be alerted immediately of their error. Thus, when the user is informed of their situation, they can immediately rectify their position, hence, solving the complication of any miscommunication in the video conference. Consequently, this allows for an uninterrupted communication while in a video conference.

There would be three outputs to our project that would be

- Face and iris tracking both in their ideal position in this case there would be no alerts as the output Is the desired output.
- Face is detected but the iris is moving (i.e. iris is moving left, right and not present in the ideal position to maintain proper eye contact) in this case our system would detect it immediately and alert the user using both audio and text alerts, stating to **"see camera".**
- When both face and iris is not present within the visibility of the camera, in this case the system will notify the user with both an audio and text alert stating that the **"face is not detected**

VIII. RESULTS AND DISCUSSION

Our project based on the iris tracking framework, has been successfully implemented and we have obtained the desired outputs.The following are the results obtained on implementing our project "eye to eye contact maintenance in video conferences".

8.1 Face detection alerts

Figure 8.1 shows how our system detects whether a face is visible or not within the visibility of the camera /screen.
When our system detects that there is no face visible in the camera or screen it displays an alert via both text and audio
" no face detected".

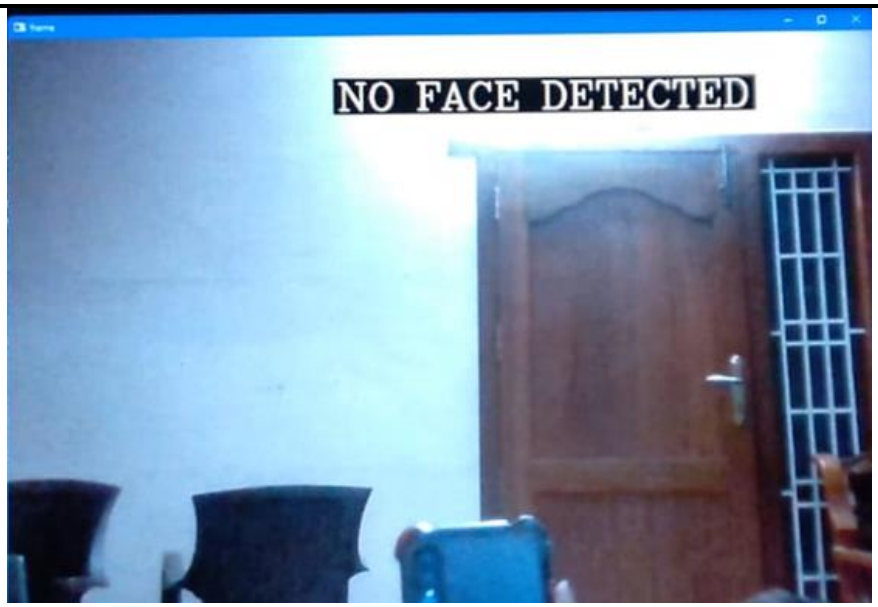

Fig 8.1: no face detected alert

8.2 Eye position alerts

Figure 8.2.1 and 8.2.2 shows how our system detects whether the user's iris is in the correct position or not inorder to maintain the correct eye contact with the user on the other end in the video call. It displays an alert through audio and text as "see camera" when the eye tracking mechanism tracks the eye's iris is not in the correct position, which in turn helps the user to rectify the error .

Fig 8.2.1: Iris is facing right side - SEE CAMERA alert

Fig 8.2.2: Iris is facing left side - SEE CAMERA alert

8.3 Ideal eye to eye contact

Figure 8.3 shows how our system detects whether the user's iris is in the correct position or not in order to maintain the eye contact with the user on the other end-user in the video call .When the system detects the face and the iris is in the correct position, it detects that the process is running successfully and no alerts are displayed.

Fig 8.3: Ideal eye to eye contact

IX.CONCLUSION AND FUTURE WORKS

In this article we propose a simple solution which is based on tracking the center of eyes using Media Pipe which tracks eye movement on screen, allowing the user to improve the subjective feeling of eye contact. The idea is to track the eye movement on the camera screen, permitting the user to focus on the screen, achieving the result of consistent eye contact without any obstacles while a video conference is occurring. When elaborated, we can say that there will be an iris-tracking model, a mechanism used to track the motion of the eyes. If any issues occur in the maintenance of consistent eye contact, the system will notify the user immediately. This allows for a persistent and continual maintenance of eye contact in the

progression of the video conference. In future we hope to develop it into a cleaner, even more user friendly and also to build a video conferencing app of our own embedded with the features of our eye contact maintenance system or at least embed our system to pre-existing video calling apps like Google meet, Webex and much more. Since our new features will be very useful in terms of a good video call experience by maintaining a constant eye contact between the various users attending the video call. We also have plans to make it into a software and hardware project.

REFERENCES

[1]Cootes, T., Cooper, D., Taylor, C., Graham, J.: Active shape models – their training and application. Computer Vision and Image Understanding 61, 38-59 (1995) Google Scholar

[2]Cootes, T., Adwards, G., Taylor, C.: Active appearance model. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(6), 681-685 (2001) Google Scholar

[3]Lowe, D.G.: Distinctive image features from scale-invariant key points. International Journal of Computer Vision 60(2) (2004) Google Scholar

[4]Slot, K., Kim, H.: Key points derivation for object class detection with sift algorithm. In: Rutkowski, L., Tadeusiewicz, R., Zadeh, L.A., Zurada, J.M. (eds.) ICAISC 2006. LNCS, vol. 4029, pp. 850-859. Springer, Heidelberg (2006) Google Scholar

[5]Cameiro, G., Jepson, A.D.: Pruning local feature correspondences using shape context. In: Proceedings of IEEE International Conference on Image Processing (2004) Google Scholar

[6]Bayliss, A. P., & Tipper, S. P. (2006). Predictive gaze cues and personality judgments: Should eye trust you? Psychological Science, 17(6), 514-Bekkering, E., and Shim, J. P. (2006). i2i Trust in Videoconferencing. Communications of the ACM, 49(7), 103-107.

[7]Buxton, W. (1992). Telepresence: integrating shared task and person spaces. Proceedings of Graphics Interface '92, 123-129.

[8]Chen, M. (2002). Leveraging the Asymmetric Sensitivity of Eye Contact for Videoconference. Paper presented at the Conference on Human Factors in Computing Systems, Minneapolis, Minnesota, USA.

[9]Detenber, B. and Reeves, B. (1996). A bio-informational theory of emotion: Motion and image size effects on viewers. Journal of Communication, 46(3), 66-84.

[10]Ferrán-Urdaneta, C., & Storck, J. (1997). Truth or Deception: The Impact of Videoconferencing for Job Interviews. Proceedings of the Eighteenth International Conference on Information Systems. Atlanta, Georgia, United States. 183-196.

[11]Are You Looking at Me? Measuring the Cone of Gaze. Journal of Experimental Psychology: Human Perception and Performance, 33(3), 705-715.Garau, M., Slater, M., Bee, S., & Sasse, M. A. (2001).

[12]The Impact of Eye Gaze on Communication using Humanoid Avatars. Proceedings of the SIGCHI conference on Human factors in computing systems. Seattle, Washington, United States, 3(1), 309-316. Gemmell, J., Toyama, K., Zitnick, C. L., Kang, T., & Seitz, S. (2000).

[13]ITU (1999). International Telecommunication Union (ITU-T P.920). Subjective video quality assessment methods for multimedia applications.

[14]Jerald, J., & Daily, M. (2002). Eye gaze correction for videoconferencing. Proceedings of the 2002 symposium on Eye tracking research & applications. New Orleans, Louisiana, USA, 77-81.