



A Validation of Automatically-Generated Areas-of-Interest in Videos of a Face for Eye-Tracking Research

Roy S. Hessels^{1,2*}, Jeroen S. Benjamins^{1,3}, Tim H. W. Cornelissen⁴ and Ignace T. C. Hooge¹

¹ Experimental Psychology, Helmholtz Institute, Utrecht University, Utrecht, Netherlands, ² Developmental Psychology, Utrecht University, Utrecht, Netherlands, ³ Social, Health and Organisational Psychology, Utrecht University, Utrecht, Netherlands,

⁴ Scene Grammar Lab, Department of Cognitive Psychology, Goethe University Frankfurt, Frankfurt, Germany

OPEN ACCESS

Edited by:

Martin Lages,
University of Glasgow,
United Kingdom

Reviewed by:

Mariano Luis Alcañiz Raya,
Universitat Politècnica de València,
Spain
Carl M. Gaspar,
Hangzhou Normal University, China

*Correspondence:

Roy S. Hessels
royhessels@gmail.com

Specialty section:

This article was submitted to
Quantitative Psychology and
Measurement,
a section of the journal
Frontiers in Psychology

Received: 03 April 2018

Accepted: 16 July 2018

Published: 03 August 2018

Citation:

Hessels RS, Benjamins JS,
Cornelissen THW and Hooge ITC
(2018) A Validation of
Automatically-Generated
Areas-of-Interest in Videos of a Face
for Eye-Tracking Research.
Front. Psychol. 9:1367.
doi: 10.3389/fpsyg.2018.01367

When mapping eye-movement behavior to the visual information presented to an observer, Areas of Interest (AOIs) are commonly employed. For static stimuli (screen without moving elements), this requires that one AOI set is constructed for each stimulus, a possibility in most eye-tracker manufacturers' software. For moving stimuli (screens with moving elements), however, it is often a time-consuming process, as AOIs have to be constructed for each video frame. A popular use-case for such moving AOIs is to study gaze behavior to moving faces. Although it is technically possible to construct AOIs automatically, the standard in this field is still manual AOI construction. This is likely due to the fact that automatic AOI-construction methods are (1) technically complex, or (2) not effective enough for empirical research. To aid researchers in this field, we present and validate a method that automatically achieves AOI construction for videos containing a face. The fully-automatic method uses an open-source toolbox for facial landmark detection, and a Voronoi-based AOI-construction method. We compared the position of AOIs obtained using our new method, and the eye-tracking measures derived from it, to a recently published semi-automatic method. The differences between the two methods were negligible. The presented method is therefore both effective (as effective as previous methods), and efficient; no researcher time is needed for AOI construction. The software is freely available from <https://osf.io/zgmch/>.

Keywords: eye tracking, Areas of Interest, faces, automatic, videos

1. INTRODUCTION

In many areas of eye-tracking research, inferences about perception or cognition are drawn by relating eye-movement behavior to the visual stimulus that was presented to the observer. Using this approach, researchers have for example concluded that individuals with social phobia look less at facial features (eyes, nose, mouth) than controls (Horley et al., 2003), that the time to disengage from a centrally presented stimulus decreases with age (Van der Stigchel et al., 2017), and have investigated how information graphics are used (Goldberg and Helfman, 2010). Although data-driven methods exist for this purpose (Caldara and Miellet, 2011), coupling eye-tracking data to the visual stimulus is usually done using so-called Areas of Interest (AOIs) (Holmqvist et al., 2011), especially when there are clear hypotheses to be tested.

There are many ways in which one can construct AOIs, for example by using a grid superimposed on the visual stimulus. Grid-cells may subsequently be combined into AOIs and labeled; e.g., cells X and Y belong to the “nose” AOI in a face (Hunnius and Geuze, 2004). AOIs may also be drawn manually (Chawarska and Shic, 2009)—as is possible with most eye-tracker manufacturer software—or constructed using state-of-the-art computer-assisted techniques (Hessels et al., 2016). When visual stimuli are static (e.g., pictures or schematic drawings), constructing AOIs need only be done once for each stimulus, and this can be achieved with most eye-tracker manufacturer software. However, when visual stimuli are moving (e.g., videos or animated elements on a screen; often referred to as *dynamic* stimuli), constructing AOIs becomes problematic. In this case, AOIs have to be manually constructed for each frame in the video (or each *n*th frame when there is little movement) (Falck-Ytter, 2008; Tenenbaum et al., 2013), which is a time-consuming process. Alternatively, computer software can be programmed to facilitate constructing AOIs. However, this may be beyond the technical skills of many researchers. Here we present software for automatically constructing AOIs in videos containing a face. This fully-automatic technique provides AOI coordinates for the features of a face (left eye, right eye, nose, and mouth) for each frame in the video.

The current state-of-the-art in eye-tracking studies on face-scanning and face-processing exemplifies that a reliable automatic AOI-construction method for videos of a face is needed. There is a large body of literature where videos of faces are shown to participants, to investigate what facial features are looked at (e.g., Vö et al., 2012; Gobel et al., 2015; Pons et al., 2015; Rutherford et al., 2015; Senju et al., 2015; Gobel et al., 2017). In these recent studies, (dynamic) AOIs were manually constructed by the researcher for each individual video, costing valuable researcher time. Additionally, there is a surge in studies investigating gaze behavior during so-called “naturalistic social interactions,” conducted using head-mounted eye-tracking glasses (Ho et al., 2015; Birmingham et al., 2017) and two-way video setups combined with remote eye trackers (Hessels et al., 2017, 2018). Here, each visual stimulus is a unique video containing a moving face (unlike studies where the same videos are used for each participant). In this field, mapping gaze to a video is generally done manually, and it is therefore an even more time-consuming process than manually constructing AOIs for one set of videos. Automatic AOI construction could significantly reduce the time invested by researchers.

In recent work, researchers have attempted to reduce the time spent mapping gaze to a moving stimulus; in this case a video of a face. In their study, Hessels et al. (2017) used a two-way video setup combined with two remote eye trackers to investigate gaze behavior during dyadic interaction (the interaction of two people). A video of the face of one participant was streamed through a live video setup to the other participants (and vice versa). Instead of constructing AOIs on a frame-by-frame basis, they applied a computer-assisted AOI-construction method to the videos. First, the centers of the left eye, right eye, nose and mouth were manually selected in the first frame of the video. Hereafter, the upper portion of the face was selected, in which

high-contrast points were detected. The software then tracked these high-contrast points throughout the video. The centers of the left eye, right eye, nose, and mouth were derived from the location of the high-contrast points. In essence, this method is ignorant to the visual stimulus being tracked: it doesn’t need a face, but can track high-contrast points in any video. The initial detection of, and the choice for, the facial landmarks (eyes, nose, and mouth) was done by the researcher. The researchers could intervene and manually correct the centers of the left eye, right eye, nose, and mouth if needed. From these centers, AOIs were constructed using the Limited-Radius Voronoi-tessellation method (LRVT) (Hessels et al., 2016). Although Hessels et al. (2017) state that their semi-automatic method drastically reduces time used for AOI construction, it still required action on the researcher’s part (1) at the start of the video, (2) whenever there were not enough high-contrast points left to track, (3) whenever there was too much movement such that the face-tracking software lost track, and (4) whenever the researchers deemed the AOI cell centers to have moved off the intended location.

A semi-automatic AOI-construction method (as outlined above) is already an improvement over manual AOI construction methods in terms of the time involved. One may wonder, however, why automatic methods for AOI construction aren’t the standard yet in eye-tracking research. With such methods, AOIs could be constructed objectively and in less time than when the researcher constructs the AOIs manually. In fact, face-detection and facial landmark detection methods (detecting locations such as the boundaries of the eyes, nose, and mouth) have been around for a long time (see e.g., Wang et al., 2018, for a comprehensive survey). In many cases, robust detection of facial landmarks can be achieved (see e.g., Baltrušaitis et al., 2013; Li et al., 2018). There are, however, only few examples where such face-detection or facial landmark detection methods have been explicitly applied in eye-tracking research and made available to the public. Examples of face-detection methods used for constructing AOIs are scarce (see e.g., de Beugher et al., 2014; Bennett et al., 2016).

Why are automatic AOI-construction methods not yet commonplace in applied eye-tracking research? We believe that there may be three reasons.

1. The methods that are available may not be reliable enough. In one eye-tracking study in which an automatic AOI-construction method was used (Bennett et al., 2016), the authors stated that “... detection of the eyes was prone to error” and that eyes and mouth were missed in up to 30% of the frames. This is unacceptable for most eye-movement researchers, who are already dealing with eye-tracking data loss due to e.g., difficult participants (e.g., infants; Hessels et al., 2015, school children, or certain patient groups; Birmingham et al., 2017). Empirical researchers are not necessarily interested in automatic methods that are efficient (i.e., run automatically, and don’t cost any researcher time), but not as effective (i.e., the AOIs are not adequately constructed) as manual methods, even though these manual methods may be highly inefficient. However, given that the reliability of facial-landmark detection methods is rapidly

increasing (Wang et al., 2018) and overall very high (Li et al., 2018), this cannot be the whole story. We believe that other reasons may be more important.

2. The average researcher in experimental psychology may not have the technical expertise to build, adapt or implement a face-detection method for usage with videos of a face.
3. There are no AOI-construction methods available that have been specifically validated against other AOI-construction methods using eye-tracking data.

Our goal in this paper is not to improve or revolutionize existing face-detection or facial landmark detection methods. We see this as the topic of research for computer vision and computer science. Our goal as eye-tracking researchers investigating gaze behavior to faces is to provide an out-of-the-box solution for AOI-construction for videos containing a face, and—most importantly—validate the method. In doing so, we make use of a particularly promising automatic method for detecting facial landmarks: the recently-released OpenFace toolbox (Baltrušaitis et al., 2016). We chose this toolbox, as it is freely available, and appeared to us to work well out-of-the-box. OpenFace excels at facial landmark detection under varying lighting conditions and facial poses (Baltrušaitis et al., 2013). Here we present a fully automatic AOI-construction method using OpenFace (Baltrušaitis et al., 2013, 2016) and the LRVT AOI method (Hessels et al., 2016), and validate it by comparing it to the semi-automatic AOI-construction method previously applied by Hessels et al. (2018) to nearly a hundred videos. We limit the application of the method to videos of frontal recordings of one face, as they are commonly used stimuli in the face-processing literature. By focusing on a small range of video-types, we can optimize on the effectivity of the method. The method presented here combines existing validated techniques (OpenFace for face-detection and facial landmark detection, and LRVT for AOI-research using face stimuli) and is easy to implement by researchers, who might not have the necessary technical skills to develop their own method.

2. METHODS

2.1. Facial Landmark Detection

Ninety-six videos of participants who were engaged in dyadic interaction through a two-way video setup were taken from a recently published study (Hessels et al., 2018). These videos each contained the frontal view of the face of one person that was in interaction with another person. Overall, there were periods of talking, laughter, making faces, movement of the head, etc. Facial landmark detection was done on these videos using the OpenFace command line binaries, which were retrieved from <https://github.com/TadasBaltrušaitis/OpenFace/>. These were applied to all videos on a computer running Windows 7.

2.2. Area of Interest Construction

OpenFace detects 68 facial landmarks, which correspond to fixed locations on the face¹. These facial landmarks were stored in a

¹see <https://github.com/TadasBaltrušaitis/OpenFace/wiki/Output-Format> for the location of each landmark.

text file containing horizontal and vertical pixel coordinates for each video frame. As not all 68 landmarks are relevant for the facial AOIs to be constructed, only a subset were used. **Table 1** details which specific landmarks were used for deriving the AOI cell centers for the left eye, right eye, nose, and mouth. Using MATLAB R2013a, AOI cell centers were obtained by averaging the coordinates of the OpenFace landmarks for each AOI, and were subsequently stored in a text file with a set of coordinates for each video frame.

From the AOI cell centers, AOIs were constructed using the LRVT method (Hessels et al., 2016). This method assigns each gaze coordinate to one of four facial features (left eye, right, nose, and mouth) based on which AOI cell center is closest to the gaze coordinate, provided that the distance does not exceed a maximum radius. Any gaze coordinate not assigned to one of the four facial AOIs is assigned to the “non” AOI. Although facial landmarks can be used to construct AOIs in any form desired, the LRVT method was chosen for two reasons. First, previous research has shown that large AOIs (such as LRVT AOIs with large radii) are to be preferred in sparse stimuli (such as faces) (Hessels et al., 2016; Orquin et al., 2016). Second, AOIs in faces created using the LRVT method with a large radius have been shown to be most noise-robust compared with other researcher-defined AOIs (Hessels et al., 2016). The LRVT radius was set to 4°. An example video frame with corresponding AOIs can be seen in **Figure 1**. As can also be seen, the videos roughly contained the face and upper torso of the participants on a uniform dark-gray background.

3. RESULTS

3.1. AOI Coordinates

In order to validate the automatic AOI-construction method presented here, we compared coordinates of the AOIs as derived from the fully automatic method against a semi-automatic method previously described in the literature. The AOI coordinates for the left eye, right eye, nose and mouth were obtained from the study of Hessels et al. (2018). We assessed (1) whether the AOI coordinates differed systematically between the two AOI-construction methods, and (2) whether the AOI coordinates were more variable over time within one method compared to the other.

The systematic difference between the two methods was assessed by calculating the mean absolute difference in AOI positions between the fully automatic and semi-automatic method. As can be seen in **Figure 2**, the mean difference across

TABLE 1 | OpenFace landmarks used for deriving Area of Interest cell centers.

AOI	OpenFace landmark numbers
Left eye	37–42
Right eye	43–48
Nose	31
Mouth	63, 67

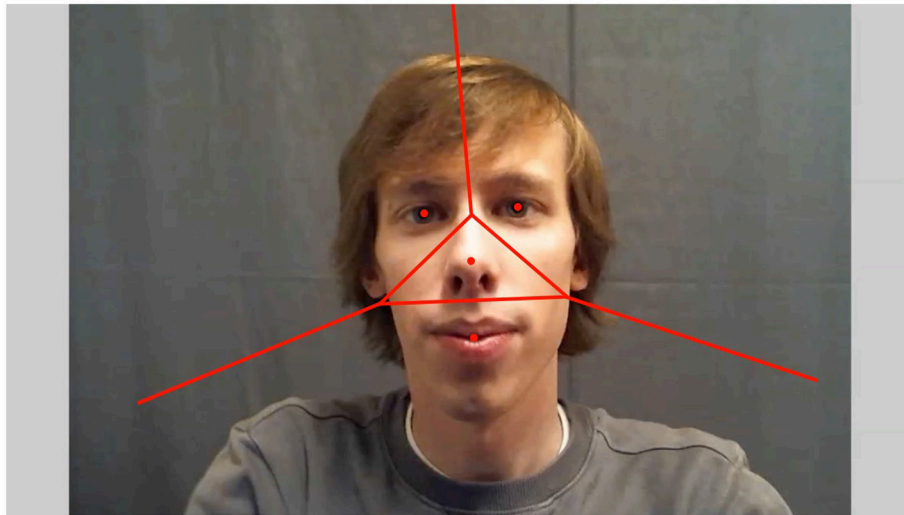


FIGURE 1 | One example frame of a video of the first author recorded with the two-way video setup (Hessels et al., 2017). Area of Interest (AOI) cell centers for the left eye, right eye, nose, and mouth are noted with red dots. Red lines indicate the borders between the AOIs derived from the Voronoi-tessellation method. The ends of these borders are arbitrary; they extend into infinity.

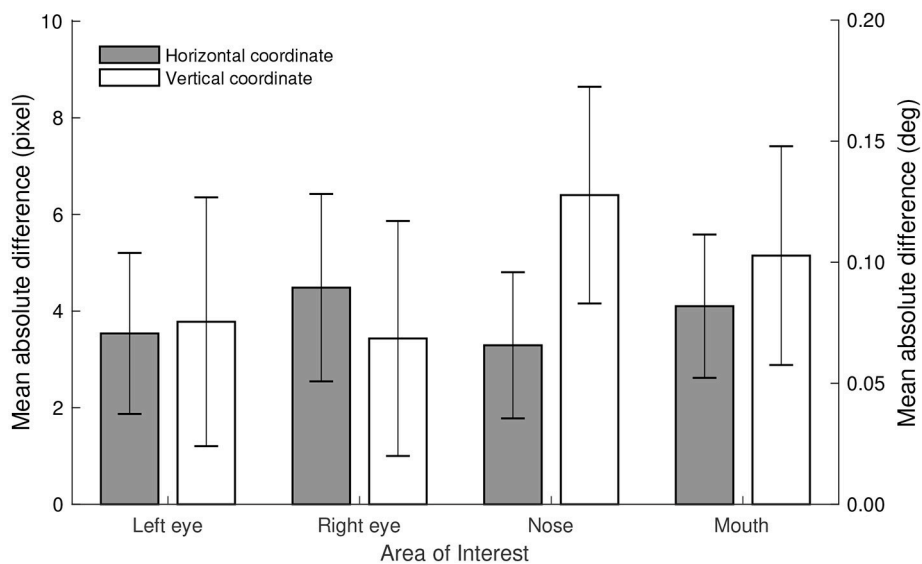


FIGURE 2 | Mean absolute differences in pixels and degrees of visual angle between the coordinates of the Area of Interest cell centers (left eye, right eye, nose, and mouth) as determined by the fully automatic (using OpenFace) and semi-automatic methods. Error bars depict standard deviation calculated across participants.

the entire video did not exceed 6.5 pixels or 0.13° of visual angle²; the largest difference being observed for the vertical coordinate of the nose AOI. These differences are much smaller than the accuracy³ values obtained with most eye trackers (around 0.5° of visual angle) (Holmqvist et al., 2011). Moreover, these values are

²Visual angles are reported under the assumption that participants remained at 81 cm from the screen and in the center of the camera image.

³Accuracy refers to the systematic offset between the gaze position as reported by the eye tracker, and the actual (or instructed) gaze position of the participant.

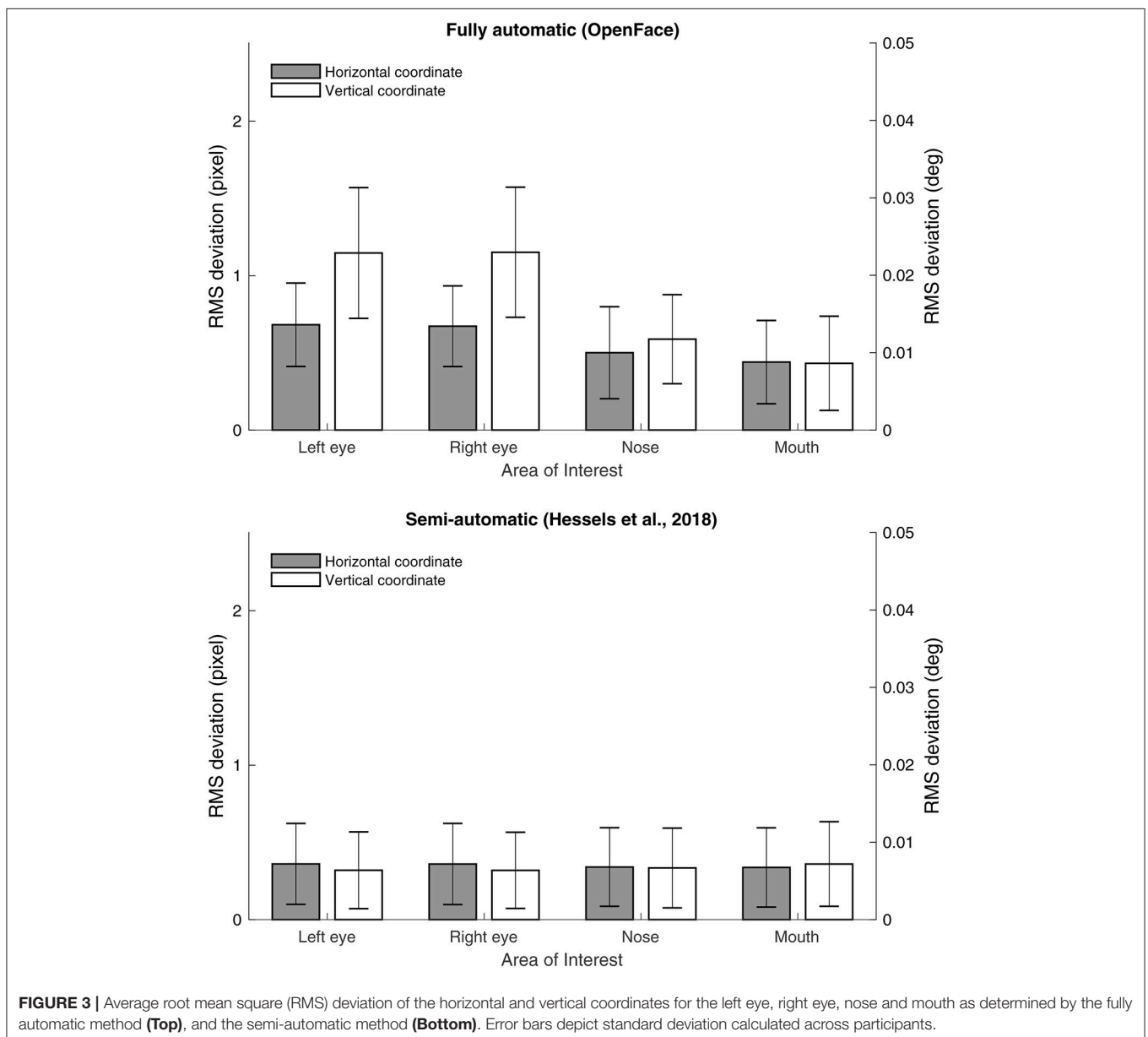
well below the AOI span of 1.9° (the mean distance from each AOI to its closest neighbor Hessels et al., 2016), and merely a fraction of the screen size (22 inch screen of 1680 by 1050 pixels, 47.38 by 29.61 cm, 32.61 by 20.72 $^\circ$). As such, we conclude that the systematic difference in the position of the AOIs between the fully automatic (using OpenFace) and semi-automatic methods is negligible.

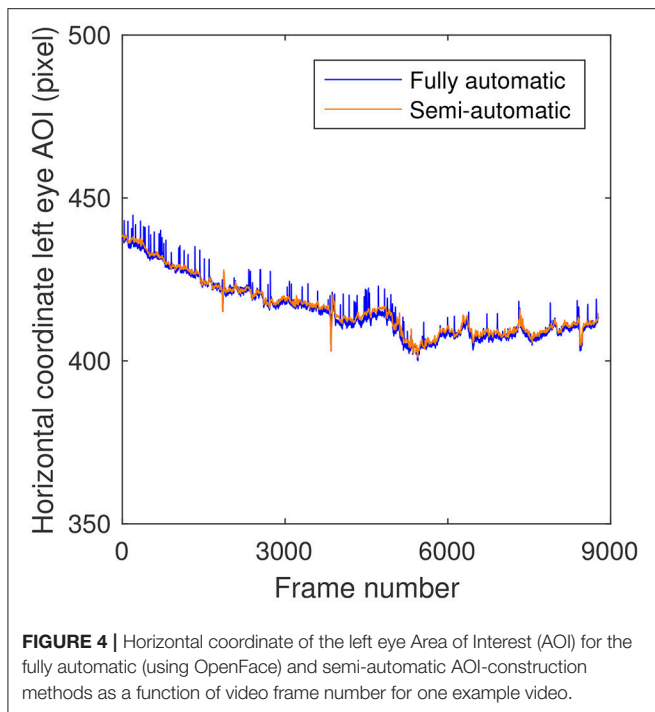
While the systematic difference between the AOI coordinates as determined by the two methods was small, it may be that the AOI coordinates of one method are more variable over time. This

can occur, for example, if the facial landmark detection is unstable over time. To examine whether this was the case, we calculated the root mean square (RMS) deviation of the horizontal and vertical coordinates for the cell centers of the four AOIs (left eye, right eye, nose, and mouth) for each method. If the fully automatic method is less stable over time, this should be visible as a larger RMS deviation in the AOI coordinates. As can be seen in **Figure 3**, the RMS values were all below 2 pixels (or 0.04° of visual angle), well below the optimal accuracy in most eye trackers as well as the AOI span. However, the RMS values were somewhat higher for the fully automatic than for the semi-automatic method. The reason for this is at least 2-fold: (1) the semi-automatic method derives AOI position from a large number of high-contrast points, making it less susceptible to

changes in pixel intensities at the location of the facial feature (2) the semi-automatic method does not continue tracking the face when there is too much movement, whereas the fully automatic method does.

Upon closer examination, another likely factor contributing to higher RMS values for the fully-automatic method was uncovered. **Figure 4** depicts the horizontal coordinate of the left eye AOI as acquired by the fully automatic and semi-automatic methods. Although the two signals almost completely overlap, there are several instances in the fully-automatic method where there is a small upward spike in the horizontal coordinate. These spikes were observed for the horizontal and vertical coordinates of the two eye AOIs, yet not for the nose and mouth. On all occasions, the spikes were downward and inward





(toward the center) in the video frame. Upon inspection of the videos, these spikes seemed to correspond to the eye blinks made by the participants. In order to ascertain the size of these spikes, the maximum absolute difference in the coordinates of the left eye and right eye between the two methods were determined for each participant and subsequently averaged across participants. Note that the value per participant corresponds to a single sample, not an average across samples. The values thus obtained were below 20 pixels or 0.4 degrees of visual angle. The largest value obtained for one participant overall was just below 45 pixels or 0.9 degrees of visual angle. It should be noted that this value includes any systematic difference in AOI coordinates between the two methods as well—the difference may not solely be caused by the spike due to the eye blink.

3.2. Eye-Tracking Measures

Although both the systematic and variable differences between the fully-automatic method and the semi-automatic method were below the optimal accuracy values of most eye trackers (i.e., AOI position is more precise and accurate than the eye-tracking signal), the negligibility of the differences can best be shown by comparing the eye-tracking measures obtained using both methods. We compared total dwell times, average dwell times, and number of dwells to the left eye, right eye, eyes (left and right eye AOIs combined), nose, mouth, and non AOI as derived using the two methods. Moreover, the original study (Hessels et al., 2018) featured two participants looking at a live stream of each other. As such, we could also compare measures on the frequency, average duration, and total duration of the paired gaze states that participants could engage in. These

paired gaze states corresponded to periods when participants were looking at each other's eyes simultaneously (two-way eye gaze), when only one was looking at the eyes while the other was not (one-way eye gaze), and when both were not looking at the eyes (no eye gaze). If the method used to determine AOI cell centers affects eye-tracking measures, measures of paired gaze states should be most affected, as they are derived from a combination of AOI data of two participants. As visible from **Table 2**, differences between the eye-tracking measures as derived from both methods were small ($\leq 2.6\%$). This indicates that eye-tracking measures were not affected by which AOI method was used (either the fully-automatic or the semi-automatic AOI method). Hence, the differences in the coordinates of the AOI cell centers between the two methods are evidently negligible.

4. DISCUSSION

Here we have presented and validated an automatic AOI-construction technique based on a combination of recent computer vision and AOI techniques that automatically achieves AOI construction for videos of a face. This technique consists of OpenFace (Baltrušaitis et al., 2013, 2016) for facial landmark detection and the Limited-Radius Voronoi-tessellation AOI method (Hessels et al., 2016). The technique was validated against semi-automatic AOI construction in 96 videos of videos of faces acquired from a study on gaze behavior in social interaction using a two-way video setup (Hessels et al., 2018). The systematic difference between the AOI coordinates as determined by the semi-automatic, and the fully automatic (using OpenFace) methods was far below the average accuracy of modern eye trackers (Holmqvist et al., 2011). This was also the case for the variable difference assessed using the root mean square (RMS) deviation of the AOI positions as derived from both methods. Moreover, eye-tracking measures as derived from the two methods differed $\leq 2.6\%$ from each other. Therefore, the systematic differences between the two methods are negligible.

Upon inspection of the AOI coordinates of the fully automatic method, we found small spikes in the AOI coordinates when participants blinked. When a participant blinked, the AOI coordinates of the eyes moved slightly downward and inward (i.e., toward the nose). Although the magnitude of these spikes was on average below the accuracy achieved by most eye trackers, we will briefly consider the consequences for data analysis. If the eye blink of a person in the video is registered as a change in the coordinate of the eye AOI cell center, it may be that the gaze position of an observer who looks around the border between two AOIs (e.g., the left eye and the nose) is briefly assigned to a different AOI. One might argue that a moving, talking and blinking face is inherently dynamic, which means that the AOIs move in the video, both together (e.g., a translation of the face) or with respect to each other (e.g., when different facial expressions are made). However, if one considers these spikes in AOI coordinate due to blinks to be a problem, it is possible to detect them and filter them

TABLE 2 | Total dwell time (TDT) and total time of paired gaze states as derived from the fully-automatic and semi-automatic methods, and the difference between the two methods in seconds and percentage.

Measure	Full automatic	Semi-automatic	Difference (s)	Absolute difference (%)
TDT left eye (s)	69.79 (<i>sd</i> = 47.05)	70.81 (<i>sd</i> = 47.6)	-1.021 (<i>sd</i> = 4.142)	1.4
TDT right eye (s)	65.95 (<i>sd</i> = 51.89)	64.46 (<i>sd</i> = 51.87)	1.498 (<i>sd</i> = 3.375)	2.3
TDT eyes (s)	135.7 (<i>sd</i> = 70.22)	135.3 (<i>sd</i> = 71.27)	0.4764 (<i>sd</i> = 5.562)	0.4
TDT nose (s)	52.17 (<i>sd</i> = 41.21)	53.56 (<i>sd</i> = 42.24)	-1.397 (<i>sd</i> = 5.611)	2.6
TDT mouth (s)	33.99 (<i>sd</i> = 37.69)	33.39 (<i>sd</i> = 37.09)	0.5945 (<i>sd</i> = 2.863)	1.8
TDT non (s)	20.12 (<i>sd</i> = 30.88)	20.13 (<i>sd</i> = 30.85)	-0.00525 (<i>sd</i> = 0.6876)	0.0
Total time two-way (s)	71.22 (<i>sd</i> = 51.24)	71.16 (<i>sd</i> = 52.22)	0.05629 (<i>sd</i> = 4.149)	0.1
Total time one-way (s)	133.8 (<i>sd</i> = 37.69)	133.3 (<i>sd</i> = 38.27)	0.5348 (<i>sd</i> = 4.139)	0.4
Total time no eye gaze (s)	67.8 (<i>sd</i> = 52.05)	68.92 (<i>sd</i> = 53.41)	-1.122 (<i>sd</i> = 4.497)	1.6

Total dwell time was calculated for the left eye, right eye, eyes (both eye AOs combined), nose, mouth, and non AOs. Total duration was calculated for the two-way eye gaze, one-way eye gaze, and no eye gaze paired gaze states. Percentage absolute difference is calculated as the absolute difference between the methods in seconds divided by the value for the semi-automatic method.

from the signal. Here, we found no evidence that eye-tracking measures were affected by which AOI method was used. To sum, the fully automatic AOI-construction method presented here produces AOI coordinates as reliably as previously used semi-automatic approaches. As it does not need researcher input, however, it greatly improves over manual or semi-automatic AOI-construction methods.

As stated in the introduction, there is a large interest in gaze behavior to facial features, both to video and during “naturalistic social interactions.” Mapping eye-tracking data to the visual stimulus requires time-intensive manual coding, which calls for an automatic method. Even though automatic techniques have previously been proposed (e.g., Bennett et al., 2016), the de facto standard for videos of a face in eye-tracking research is still manual AOI-construction. As we have stated, this might be due to the fact the available methods for such purposes specifically are not reliable enough. Moreover, it may be that researchers in experimental psychology do not have the technical expertise to build or implement existing face-detection and/or facial landmark detection methods for use in eye-tracking research. Finally, no automatic AOI-methods have been extensively validated against other AOI-construction methods using eye-tracking data. Here we have presented a first fully automatic AOI-construction method for videos containing a face in eye-tracking research. We have validated the method and shown that is more *efficient* and at least as *effective* as a previously published semi-automatic method. We therefore believe it has a high potential utility in the applied eye-tracking fields of face scanning, face processing, etc.

We have taken a state-of-the-art toolbox for facial-landmark detection (OpenFace) to construct AOIs for one specific problem in eye-movement research, namely investigating one’s gaze behavior when looking at the moving and deforming face of another. OpenFace is, however, capable of tracking more than one face in a video, as are other techniques in the literature (e.g., Farfade et al., 2015). In the future, automatic AOI-construction methods that are built upon those techniques can be validated for use as an automatic

AOI-construction methods using eye-tracking data. Such validation studies are important to ascertain the robustness and reliability of automatic AOI-construction methods and may be a great push forward for the field of face perception. There is much to be gained by incorporating computer vision/computer science techniques in applied eye-tracking research. Using such a technique we have tackled one specific problem.

5. CONCLUSIONS

In order to map eye-movement behavior to the visual information presented to an observer, Areas of Interest (AOIs) are commonly employed. Constructing AOIs for static stimuli requires that one AOI set is constructed for each individual stimulus, and this is possible with most eye-tracker manufacturer software. For moving stimuli, however, it is often a time-consuming process, as AOIs have to be constructed for each frame of the video. We’ve validated a fully automatic AOI-construction method for videos of faces based on OpenFace facial landmark detection and Voronoi-tessellation. The difference between the AOI coordinates derived from this method and a semi-automatic method was negligible, as were the eye-tracking measures derived from them. This means that the method is at least as effective as manual or semi-automatic AOI construction. Moreover, as the AOI-construction method is fully automatic it is highly efficient, and can save valuable researcher time. Given the effectivity and efficiency of the present method, we believe it could become the new standard in applied eye-tracking fields where videos of faces are used. The software is freely available from <https://osf.io/zgmch/>. Technical details, requirements and instructions are given in **Supplementary Data Sheet 1**.

AUTHOR CONTRIBUTIONS

RH and IH conceived the study. RH, JB, and TC wrote the MATLAB code implementing the AOI-construction method on

the pre-existing OpenFace toolbox. RH performed the data analysis and drafted the manuscript. All authors commented on and helped finalize the manuscript.

FUNDING

RH was supported by the Consortium on Individual Development (CID). CID is funded through the Gravitation

program of the Dutch Ministry of Education, Culture, and Science and the NWO (Grant No. 024.001.003).

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fpsyg.2018.01367/full#supplementary-material>

REFERENCES

- Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2013). "Constrained Local Neural Fields for Robust Facial Landmark Detection in the Wild," in *2013 IEEE International Conference on Computer Vision Workshops (ICCVW)* (Sydney, NSW: IEEE), 354–361.
- Baltrušaitis, T., Robinson, P., and Morency, L.-P. (2016). "OpenFace: an open source facial behavior analysis toolkit," in *IEEE Winter Conference on Applications of Computer Vision* (Lake Placid, NY) 1–10.
- Bennett, J. K., Sridharan, S., John, B., and Bailey, R. (2016). "Looking at faces: autonomous perspective invariant facial gaze analysis," in *SAP '16 Proceedings of the ACM Symposium on Applied Perception* (New York, NY: ACM Press), 105–112.
- Birmingham, E., Smith Johnston, K. H., and Iarocci, G. (2017). Spontaneous gaze selection and following during naturalistic social interactions in school-aged children and adolescents with autism spectrum disorder. *Can. J. Exp. Psychol.* 71, 243–257. doi: 10.1037/cep0000131
- Caldara, R., and Miellet, S. (2011). iMap: a novel method for statistical fixation mapping of eye movement data. *Behav. Res. Methods* 43, 864–878. doi: 10.3758/s13428-011-0092-x
- Chawarska, K., and Shic, F. (2009). Looking but not seeing: atypical visual scanning and recognition of faces in 2 and 4-year-old children with autism spectrum disorder. *J. Autism Dev. Disord.* 39, 1663–1672. doi: 10.1007/s10803-009-0803-7
- de Beugher, S., Brône, G., and Goedemé, T. (2014). "Automatic analysis of in-the-wild mobile eye-tracking experiments using object, face and person detection," in *International Conference on Computer Vision Theory and Applications VISAPP*, Vol. 1 (Lisbon), 625–633.
- Falck-Ytter, T. (2008). Face inversion effects in autism: a combined looking time and pupillometric study. *Autism Res.* 1, 297–306. doi: 10.1002/aur.45
- Farfade, S. S., Saberian, M. J., and Li, L.-J. (2015). "Multi-view face detection using deep convolutional neural networks," in *the 5th ACM* (New York, NY: ACM Press), 643–650.
- Gobel, M. S., Chen, A., and Richardson, D. C. (2017). How different cultures look at faces depends on the interpersonal context. *Can. J. Exp. Psychol.* 71, 258–264. doi: 10.1037/cep0000119
- Gobel, M. S., Kim, H. S., and Richardson, D. C. (2015). The dual function of social gaze. *Cognition* 136, 359–364. doi: 10.1016/j.cognition.2014.11.040
- Goldberg, J. H., and Helfman, J. I. (2010). "Comparing information graphics: a critical look at eye tracking," in *BELIV'10* (Atlanta, GA), 71–78.
- Hessels, R. S., Andersson, R., Hooge, I. T. C., Nyström, M., and Kemner, C. (2015). Consequences of eye color, positioning, and head movement for eye-tracking data quality in infant research. *Infancy* 20, 601–633. doi: 10.1111/infa.12093
- Hessels, R. S., Cornelissen, T. H. W., Hooge, I. T. C., and Kemner, C. (2017). Gaze behavior to faces during dyadic interaction. *Can. J. Exp. Psychol.* 71, 226–242. doi: 10.1037/cep0000113
- Hessels, R. S., Holleman, G. A., Cornelissen, T. H. W., Hooge, I. T. C., and Kemner, C. (2018). Eye contact takes two – autistic and social anxiety traits predict gaze behavior in dyadic interaction. *J. Exp. Psychopathol.* 9, 1–17. doi: 10.5127/jep.062917
- Hessels, R. S., Kemner, C., van den Boomen, C., and Hooge, I. T. C. (2016). The area-of-interest problem in eyetracking research: a noise-robust solution for face and sparse stimuli. *Behav. Res. Methods* 48, 1694–1712. doi: 10.3758/s13428-015-0676-y
- Ho, S., Foulsham, T., and Kingstone, A. (2015). Speaking and listening with the eyes: gaze signaling during dyadic interactions. *PLoS ONE* 10:e0136905. doi: 10.1371/journal.pone.0136905
- Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., and van de Weijjer, J. (2011). *Eye Tracking: A Comprehensive Guide to Methods and Measures*. Oxford, UK: Oxford University Press.
- Horley, K., Williams, L. M., Gonsalvez, C., and Gordon, E. (2003). Social phobics do not see eye to eye: a visual scanpath study of emotional expression processing. *J. Anxiety Disord.* 17, 33–44. doi: 10.1016/S0887-6185(02)00180-9
- Hunnius, S., and Geuze, R. H. (2004). Developmental changes in visual scanning of dynamic faces and abstract stimuli in infants: a longitudinal study. *Infancy* 6, 231–255. doi: 10.1207/s15327078in0602_5
- Li, M., Jeni, L., and Ramanan, D. (2018). Brute-force facial landmark analysis with a 140,000-way classifier. *arXiv:1802.01777* [preprint]. 1–9. Available online at: <https://arxiv.org/abs/1802.01777>
- Orquin, J. L., Ashby, N. J. S., and Clarke, A. D. F. (2016). Areas of interest as a signal detection problem in behavioral eye-tracking research. *J. Behav. Decis. Making* 29, 103–115. doi: 10.1002/bdm.1867
- Pons, F., Bosch, L., and Lewkowicz, D. J. (2015). Bilingualism modulates infants' selective attention to the mouth of a talking face. *Psychol. Sci.* 26, 490–498. doi: 10.1177/0956797614568320
- Rutherford, M. D., Walsh, J. A., and Lee, V. (2015). Brief report: infants developing with ASD show a unique developmental pattern of facial feature scanning. *J. Autism Dev. Disord.* 45, 2618–2623. doi: 10.1007/s10803-015-2396-7
- Senju, A., Vermetti, A., Ganea, N., Hudry, K., Tucker, L., Charman, T., et al. (2015). Early social experience affects the development of eye gaze processing. *Curr. Biol.* 25, 3086–3091. doi: 10.1016/j.cub.2015.10.019
- Tenenbaum, E. J., Shah, R. J., Sobel, D. M., Malle, B. F., and Morgan, J. L. (2013). Increased focus on the mouth among infants in the first year of life: a longitudinal eye-tracking study. *Infancy* 18, 534–553. doi: 10.1111/j.1532-7078.2012.00135.x
- Van der Stigchel, S., Hessels, R. S., van Elst, J. C., and Kemner, C. (2017). The disengagement of visual attention in the gap paradigm across adolescence. *Exp. Brain Res.* 235, 3585–3592. doi: 10.1007/s00221-017-5085-2
- Võ, M. L. H., Smith, T. J., Mital, P. K., and Henderson, J. M. (2012). Do the eyes really have it? Dynamic allocation of attention when viewing moving faces. *J. Vis.* 12:3. doi: 10.1167/12.13.3
- Wang, N., Gao, X., Tao, D., Yang, H., and Li, X. (2018). Facial feature point detection: a comprehensive survey. *Neurocomputing* 275, 50–65. doi: 10.1016/j.neucom.2017.05.013

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Hessels, Benjamins, Cornelissen and Hooge. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.