# Towards Automation of Dynamic-Gaze Video Analysis Taking Functional Upper-Limb Tasks as a Case Study

Musa Alyaman[1*], Mohammad Sobuh[2], Alaa Abu Zaid[1], Laurence Kenney[3], Adam J Galpin[3], Majid A. Al-Taee[4]

[1]Mechatronics Engineering Department, School of Engineering, The University of Jordan, Amman, 11942, Jordan
[2]Department of Orthotics & Prosthetics, School of Rehabilitation Sciences. The University of Jordan, Amman, 11942, Jordan
[3]School of Health and Society, University of Salford, Manchester M5 4WT, UK
[4]School of Electrical Engineering, Electronics and Computer Science, University of Liverpool, Liverpool L69 3BX, UK

**ABSTRACT**

**Background and objective**: Previous studies in motor control have yielded clear evidence that gaze behavior (where someone looks) quantifies the attention paid to perform actions. However, eliciting clinically meaningful results from the gaze data has been done manually, rendering it incredibly tedious, time-consuming, and highly subjective. This paper aims to study the feasibility of automating the coding process of the gaze data taking functional upper-limb tasks as a case study.

**Methods**: This is achieved by developing a new algorithm capable of coding the collected gaze data through three main stages; data preparation, data processing, and output generation. The input data in the form of a crosshair and a gaze video are converted into a 25Hz frame rate sequence. Keyframes and non-key frames are then obtained and processed using a combination of image processing techniques and a fuzzy logic controller. In each trial, the location and duration of gaze fixation at the areas of interest (AOIs) are obtained. Once the gaze data is coded, it can be presented in different forms and formats, including the stacked color bar.

**Results**: The obtained results showed that the developed coding algorithm highly agrees with the manual coding method but significantly faster and less prone to unsystematic errors. Statistical analysis showed that Cohen's Kappa ranges from 0.705 to 1.0. Moreover, based on the intra-class correlation coefficient (ICC), the agreement index between computerized and manual coding methods is found to be (i) 0.908 with 95% confidence intervals (0.867, 0.937) for the anatomical hand and (ii) 0.923 with 95% confidence intervals (0.888, 0.948) for the prosthetic hand. A Bland-Altman plot also showed that all data points are closely scattered around the mean. These findings confirm the validity and effectiveness of the developed coding algorithm.

**Conclusion**: The developed algorithm demonstrated that it is feasible to automate the coding of the gaze data, reduce the coding time, and improve the coding process's reliability.

**Keywords**

Video analysis, image processing, fuzzy logic, gaze tracking, fixation duration, upper limb tasks.

---

*Corresponding Author, email: m.alyaman@ju.edu.jo

## 1. Introduction

Studying gaze behavior is a growing area of research in motor control that has led to a better understanding of how humans learn to efficiently use the surrounding environment's information during task performance [1, 2]. For instance, during manual tasks that involve completing some sub-actions, humans tend to plan the sub-actions [3]. This is supported by the fact that we "look ahead" in time onto the object related to the forthcoming sub-action even, sometimes, before we finish the prior sub-action [3, 4]. Moreover, during manual task performance, the gaze is used in a way in which it is intimately linked to specific requirements of the task [5-8]. The gaze is predominantly fixated on objects/locations relevant to the manual task.

In contrast, irrelevant objects/locations are rarely fixated, reflecting the selective nature of humans' visual perception during manual task performance [9]. The ability of using the gaze only when and where it is needed during manual performance suggests that gaze behavior provides a proxy for the amount and location of attention required to perform a motor task (i.e., reach to grasp an object). The amount of attention itself seems to be a function of the level of experience/practice. Therefore, gaze, presumably, can indicate the degree of motor learning [10-14]. The total number of fixations, gaze duration, scan paths, and the number of transitions between areas of interest (AOI) are classic examples of gaze-related parameters. These parameters have been repeatedly used in previous studies [10-15] to indicate learning.

Upper limb prostheses are medical devices designed to restore the shape and function of the missing limb segment(s) following upper limb amputation or for people born without an upper limb. There has been a tremendous investment in research focusing on upper limb prosthetic technology, which led to the introduction of sophisticated multi-articulated "smart" hands. These hands offer the amputee the ability to actively grasp objects using various grasping patterns selected based on control signals from the residuum muscles. This development may improve the functionality (i.e., ability to perform tasks) and/or support more natural approaches to grasping. However, there are limited reports on the extent to which these assumptions are correct [16], and there are very few publications on real-world use (or non-use) of such devices [17]. Perhaps, one of the factors that may relate to this observation is the visual attention required to control prosthetic devices. Various approaches, including EEG analysis and gaze analysis, have been used to characterize attentional demands during the performance of tasks using myoelectric controlled prosthetic hands [18]. Findings from studies such as these consistently show distinct differences between behaviors seen during upper limb task performance with the anatomical hand compared to the prosthetic hand, reflecting increased cognitive load [15]. This is likely due to the open-loop nature of commercial prosthetic hand controllers and uncertainty in response to commands introduced at the socket-limb interface [19].

In the case of gaze behavior analysis methods, recent studies have focused on the determination of AOI. Muthumanickam *et al.* [20] introduced a method to automatically identify spatial AOIs changing over time through a combination of clustering and cluster merging in the temporal domain. Their work determines the AOI over long durations, though it can be applied to other domains such as monitoring complex systems. Chukoskie *et al.* [21] reported a neural network method to determine the AOI. Mohseni et al. [22] reported a classification method for five complex functional upper-limb movements using pre-movement planning and preparation recording of EEG data. However, the feature extraction model learning and classification were carried out offline in this study.

Other studies attempted to understand the differences between the anatomical and prosthetic gaze behaviors during manual task performance [15, 23]. In these studies, the gaze analysis involved defining at which objects the gaze fixation took place in the visual scene ahead of the user while performing the task and the gaze fixations' duration at each object. For this purpose, the objects (and sometimes multiple areas within objects) are defined virtually in the scene, and then a rater went through the recorded videos, frame by frame, to code the gaze. This process is too time consuming and highly subjective and is inevitably prone to unsystematic error [24]. Therefore, there is a need for technology-assisted methods to overcome these limitations. To identify more promising methods for evaluating prostheses that could be used in early-stage studies of novel designs, it is first necessary to better understand what factors are closely associated with the ease of control of a prosthesis [25, 26]. Secondly, there is a need to identify the extent to which studying prosthesis control with anatomically intact subjects is a valid approach.

An interesting attempt to automatically define gaze data is reported by Lavoie et al. [27] and reused by Williams et al. [28, 29] and Hebert et al. [30]. The reported approach relies on measuring the distance between the objects (i.e., AOIs) defined by marker data and a gaze vector defined by the eye tracker. In this method, a gaze fixation is considered when the distance between the gaze vector and AOI is under a particular minimum distance value. The velocity from gaze vector to AOI is also adequately low (0.5 m/s). The minimum distance values seem to be specific to the experimental setup used in these studies. This neat approach also requires using an infrared-based motion capture system to define the AOIs. Such costly systems may not be available for use in many research centers and defiantly unavailable in most clinical sites, limiting this approach's usability. A motion capture system also limits the ability to explore gaze behaviors in a more realistic environment (i.e., therapeutic apartments) as it requires specialized room settings.

The work presented in this paper utilizes an existing gaze dataset reported in Sobuh *et al.* [15]. It aims to study the feasibility of automating gaze video analysis taking functional upper limb tasks as a case study. This is achieved by developing a new algorithm based on image processing techniques and a cascaded fuzzy-logic controller. The proposed algorithm can automate the coding process to elicit clinically meaningful results from the gaze data, thus saving time and minimizing potential human errors.

## 2. Methods

### *2.1 Study subjects*

Four unilateral trans-radial amputees, three males and one female with a mean age of 49.25 years (range: 35 – 56 years), were given written consent and participated in the study. All the participants use a myoelectric upper-limb prosthesis. The study was approved by the University of Salford Ethics Committee (Ref# REP11/028) and Northwest 10 NHS Research Ethics Committee (Ref: 11/NW/0060).

### *2.2 Experimental considerations*

The experimental setup is fully detailed in the study by Sobuh *et al.* [15]. In essence, the subject's gaze behaviors were recorded while performing ten repeats (trials) of a functional task, namely water pouring, first with the anatomically intact hand, then with the contra-lateral prosthetic arm. The task was performed from a sitting position, and both hands are rested on a predefined position on a table, as seen in Figure 1. The subjects were instructed to reach for a (9.5 x 7 x 23 cm) squeezable juice carton (filled with 200 ml of water) placed on a predefined location on the table, pick it up, then pour all of the water from the carton to a glass. Finally, the subject was required to place the carton back at its starting point, release the carton, and return the hand to its starting point.

Before starting each attempt at the task, the subject was instructed to focus on a marked "gaze reference point" (GRP) in the center of the table (approximately 10 cm from the distal edge of the table) to prevent subjects from fixating the carton before task onset. During task performance, subjects were allowed to move their eyes freely. Furthermore, head movements during task performance were unconstrained. At the end of each trial, subjects were instructed to return their gaze to the GRP.

The experimental work was carried out in the morning under the same testing settings in a gait laboratory with the same lighting conditions. Gaze behaviors were recorded using iView X™ HED 2 (SenseMotoric Instruments GmbH, Tellow, Germany) eye-tracking system. The recorded gaze data represented a video file that displayed a scene ahead of the subject with an embedded crosshair in each trial. The anatomical and prosthetic hands were tested in each session, and the marker and gaze data were collected. The marker data (not discussed in this paper) was used to calculate the upper limb's kinematic characteristics (with and without a prosthesis) while completing the functional task. Each session lasted for about an hour, distributed as follows:

1. Obtain written consent, including reading the participant information sheet, explaining the study protocol, demonstrating the task completion, and answering any relevant questions (about 20 minutes).
2. Attach the markers to the body and calculate the center of rotation of the shoulder joint (about 20 minutes).

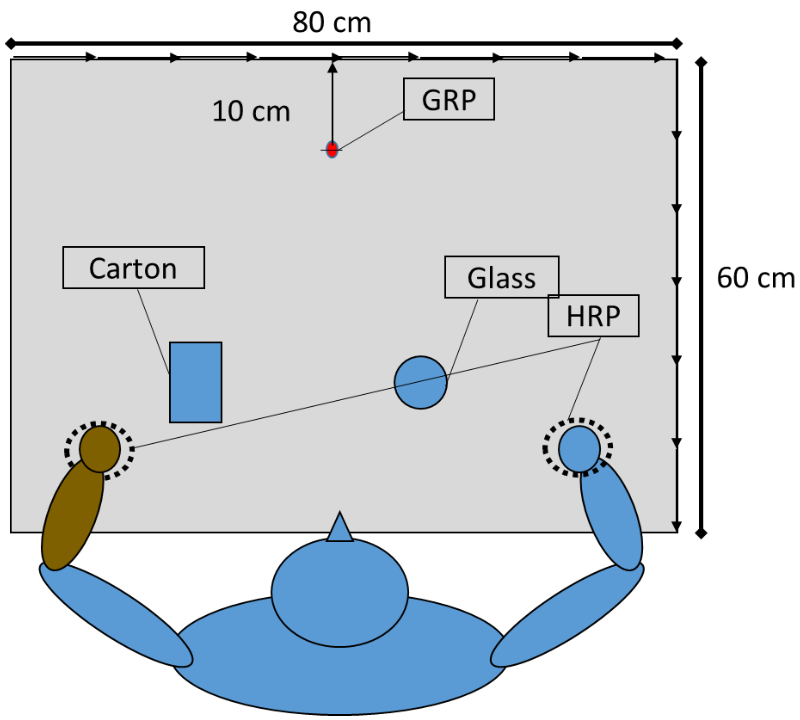



**Figure 1**. Experimental setup for the performed tasks, Sobuh et al. [15]

3. Data collection according to the following order:
    (a) Collecting gaze and marker data for the intact limb (Condition 1).
    (b) Quick inspection for the collected data (about 10 minutes).
    (c) Collecting gaze and marker data for the prosthetic limb (Condition 2).

Each participant completed the task 10 times, and each trial (attempt to complete the task) took at most 15 seconds for the anatomical hand and 25 seconds for the prosthetic hand. The total data collection duration for both conditions was less than 8 minutes (about 3 minutes for Condition 1 and 5 minutes for condition 2). As the actual testing (i.e., data collection) period for both conditions was less than 10 mins, the effect of the subject's fatigue on the results is considered negligible [31, 32].

In this study, two test conditions, using anatomical and prosthetic hands, are considered for coding each of the subjects involved. Each test is based on five trials data, resulting in 40 coded trials of gaze data. The Gaze data is coded twice: once manually, where a rater goes through gaze data frame by frame to identify which AOIs are fixated throughout the gaze trail. The data is then exported to the algorithm designed for this study to determine which AOIs are fixated throughout the gaze trial, then label the gaze data.

The proposed coding algorithm consists of three main stages: data preparation, data processing, and output generation. Each stage performs specific tasks towards obtaining a stacked color bar representing the location and duration of gaze fixation at AOIs in each trial.

### *2.3 Data preparation*

The data preparation stage consists of two stages: preprocessing and keyframes detection. The main tasks performed in each of these stages are outlined as follows.

#### *2.3.1 Preprocessing*

At this stage, the eye tracker collects data as a crosshair image and a gaze video clip. The crosshair image has a fixed feature (i.e., shape and color); thus, it can be detected directly using its preloaded image. The gaze video can be optionally cropped to remove unnecessary parts of the video (e.g., irrelevant initial recording, if any). This reduces the processing time by focusing on a particular area of interest. The gaze video is then converted into frames sequence and used to detect the AOIs of the non-key frames.

Unlike the crosshair image, the AOIs feature of the video frames/images, captured from the head-mounted scene camera, changes from one frame to another, as shown in Figure 2. The AOIs, therefore, cannot be detected from a single image/frame, as in the case of the crosshair. Also, some AOIs may appear after the video clip's onset; for instance, the hand only appears sometime after the first frame in Figure 2. These challenges are addressed by detecting keyframes, which help manual identification of the AOIs by a specialist.

#### *2.3.2 Keyframes detection*

When a prostatic user video is watched with a frame rate of 25 frames per second, the frame changes can be noticed every few seconds. This means that the first three seconds might be very similar, with a little difference between the 75 frames. Then frame number 76 may have a big difference, such as the appearance of new AOI (hand), so to reduce the required processing time, we introduce a new idea called a Keyframe.

Keyframes (KFs) are the gaze-data frames that exhibit significant feature differences from one to another. The first frame of the gaze video is, therefore, always considered a keyframe. Other KFs are detected by a fuzzy-logic controller based on three inputs, shown in Figure 3. Each input represents a measure for the level of change between two consecutive frames (e.g., Frame 1 and Frame 2, then Frame 2 and Frame 3, and so on). The first input that represents the absolute difference between the current frame/image and the previous one of the red, green, and blue color components is calculated as follows:

$$\Delta I_{(rgb)} = I_{n(rgb)} - I_{n-1(rgb)} \tag{1}$$

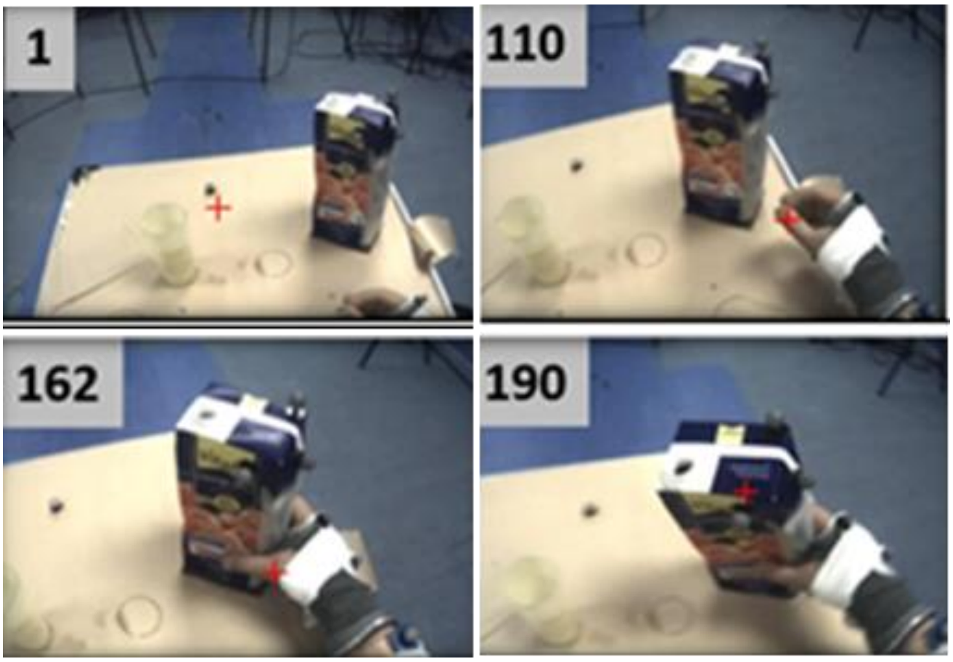



**Figure 2**. Example showing how the carton object changes in orientation and shape across different frames (labeled 1, 110, 162, and 190) while the shape of the crosshair (red cross) remains fixed

The main challenge in using a color difference is its sensitivity to the lighting conditions. To eliminate this effect, an anisotropic diffusion filter [33] is used to smooth the image and preserve the leading edges. The absolute difference between the images is calculated in (2) and used as a second input to the controller.

$$\Delta I_{(Anisotropic)} = I_{n(Anisotropic)} - I_{n-1(Anisotropic)} \tag{2}$$

The last input is the energy-absolute difference between the current image and the previous one. To calculate the energy of an image, a gray level co-occurrence matrix [34] is calculated.

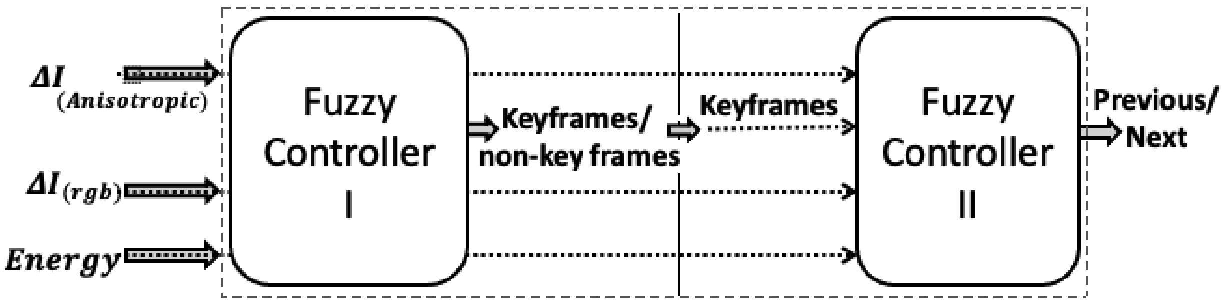



**Figure 3**. Simplified block diagram of the cascaded fuzzy logic controllers

The co-occurrence matrix size of the grayscale image in gaze data videos is 256x256. This size is scaled down to an 8x8 matrix to reduce the calculation time. The energy of the image is then calculated as follows:

$$Energy = \sum_{i=0}^{7} \sum_{j=0}^{7} G^2(i.j) \tag{3}$$

where G is the value of the gray level co-occurrence at index (i, j); i is a row number, and j is a column number.

The three inputs of the fuzzy controller are then normalized between -1 and 1, as follows:

$$I_{Norm} = 2\left(\frac{I-I_{Min}}{I_{Max}-I_{Min}}\right) - 1 \tag{4}$$

Given the normalized inputs of the controller, a given frame is considered as a keyframe if:

- The high weight change in (2) is detected since the difference becomes large when there are actual changes in the image because the lighting conditions are eliminated after applying an anisotropic diffusion filter.
- A high weight change in energy and color difference is detected since any small change in lighting conditions significantly affects the result.

The fuzzy output value that is determined in the aggregation stage is converted to a crisp value (0 or 1) by calculating the area's center. For example, if the output value is higher than a certain threshold, the frame is considered a keyframe (membership 1; otherwise, it is considered a non-key frame (membership 2). In this study, a threshold value of 0.7 is found appropriate to obtain the best results based on a trial-and-error method.

The fuzzy controller iterates through all frames (except for the first one), generating a group of keyframes. The keyframes can therefore be detected even in slow-moving objects. However, the limitation here is that a keyframe with an order $F$ will be considered closer to frame $F$−1 than to the previous keyframe (e.g., if frames 1 and 100 are considered keyframes, then frame 99 is considered closer to keyframe 100 than to keyframe 1). To address this limitation, a second fuzzy controller is cascaded to the first one. The keyframes detected by the first controller are now fed as a fourth input to the second one, as shown in Figure 3. The second controller iterates through all keyframes and eventually generates an output with two membership functions: previous and next. The previous output designates that the current frame is similar to the previous keyframe, while the next output designates that the current frame is similar to the next keyframe.

The keyframe identification efforts performed by the single and cascaded controllers are tested using a sample video of 603 frames. The number of keyframes detected in this video is 11 (1.8%). These keyframes require manual intervention to select the AOIs, while all other frames (98.2%) are processed automatically by the developed algorithm. Table 1 shows the numbers of the identified keyframes and the corresponding frame ranges (start and end frame numbers) using both the single and cascaded controllers. It is worth noting that the cascaded controller also improves the AOIs detection in non-key frames as the template matching (explained later in Section 2.4.2) will be low in the frame immediately before the next keyframe if the second fuzzy controller is excluded.

**Table 1.** Comparison between Keyframes detection using single and cascaded fuzzy controller

| Keyframe # | First fuzzy controller | | Cascaded fuzzy controllers | |
|---|---|---|---|---|
| | Start frame # | End frame # | Start frame # | End frame # |
| **1** | 1 | 99 | 1 | 97 |
| **100** | 100 | 111 | 98 | 105 |
| **112** | 112 | 198 | 106 | 189 |
| **199** | 199 | 267 | 190 | 216 |
| **268** | 268 | 451 | 217 | 447 |
| **452** | 452 | 462 | 448 | 457 |
| **463** | 463 | 526 | 458 | 522 |
| **527** | 527 | 543 | 523 | 540 |
| **544** | 544 | 550 | 541 | 549 |
| **551** | 551 | 562 | 550 | 556 |
| **563** | 563 | 603 | 557 | 603 |

Figure 4 shows an example for keyframes 1 and 100 and the keyframes detected by the second controller. As illustrated, frames 98 and 99 are in the range of keyframe 100 instead of keyframe 1. Therefore, these two frames are considered closer to keyframe 100, especially in prostheses hand shape, since its larger part appears in those two frames compared to keyframe 1.

The selection of the AOIs is performed in terms of a group of polygon vertices, shown in Figure 5(a). After defining the vertices, the polygon is closed by moving the mouse pointer and clicking on the initially selected point. Once the detection process completes, the identified keyframes are displayed to help the specialist select the corresponding AOIs. A user-friendly graphical user interface (GUI) shown in Figure 5(b) is developed to facilitate the manual selection of the AOIs. Initially, the specialist can select an AOI either from an existing dropdown menu or provide its name in a textbox. Then, s/he can either use the 'Next AOI' button to select another AOI or use the 'Restart Selection' button if an error exists.  Once the AOIs' selection in a particular keyframe is completed, the specialist continues the selection process in other keyframes in a similar manner.

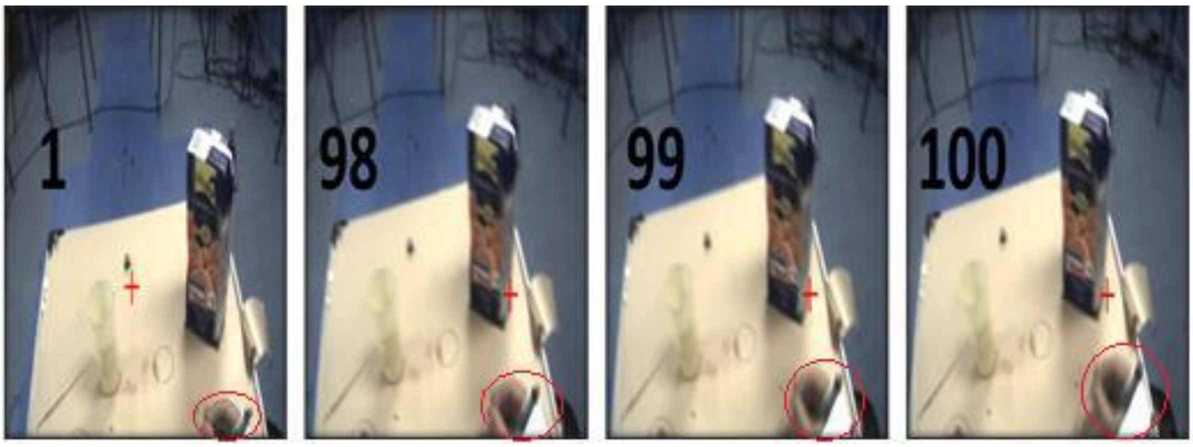
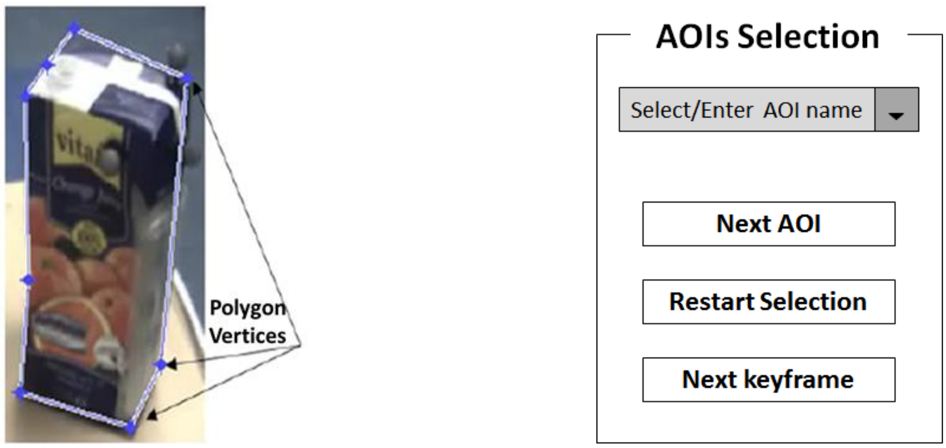



Figure 4. Example results obtained from the second fuzzy controller demonstrating how the frames 98 and 99 are closer to keyframe 100 than keyframe 1.




(a) Object's polygon vertices (b) Graphical user interface

**Figure 5.** An example demonstrating how the carton object is selected as an AOI, using polygon vertices

The algorithm uses the order in which the specialist selects the AOIs to prioritize the AOIs (i.e., the first AOI has the highest priority and the last one has the lowest priority). If AOIs overlap in a specific frame and the crosshair appears over the overlapped section, the frame is automatically labeled after the first AOI.

*2.4 Data processing*

At this stage, the crosshair and AOIs in the remaining non-key frames are detected by the algorithm, considering that the shape of AOIs may change slightly in these frames. The detection process is explained as follows.

*2.4.1 Crosshair detection*

The eye tracker generates a gaze position and projects it onto the scene ahead in a crosshair shape, as discussed earlier in the experimental considerations section. The user can select a distinctive crosshair color to ensure clarity in the scene. The color selection is, therefore, done before recording the gaze data by the eye tracker. In this experiment, a red crosshair is considered an appropriate choice. The crosshair detection is performed by converting the frame and crosshair images from RGB color space to YCbCr color space [35]. In the present work, the YCbCr is used to: (i) enhance the contrast between the preloaded crosshair image and its background and (ii) eliminate the effect of any changes to the background, as follows:

$$\begin{bmatrix} Y \\ C_B \\ C_R \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \frac{1}{256}\begin{bmatrix} 65.738 & 129.057 & 25.064 \\ 37.945 & -74.494 & 112.439 \\ 112.439 & -94.154 & -18.285 \end{bmatrix} \cdot \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{5}$$

where

R, G, and B: Red, Green, and Blue are color intensity values

Y: Luminance component

$C_B$: Blue difference chroma component

$C_R$: Red difference chroma component

The stages of the crosshair detection process are shown in Figure 6. The red diffidence chrominance component (CR), which represents the third layer in YCBCR color space, is initially obtained for both the frame and crosshair images (Figure 6a). Next, the normalized cross-correlation is applied, as shown in Figure 6b, and the obtained result is used to determine whether the crosshair exists in the frame or not.

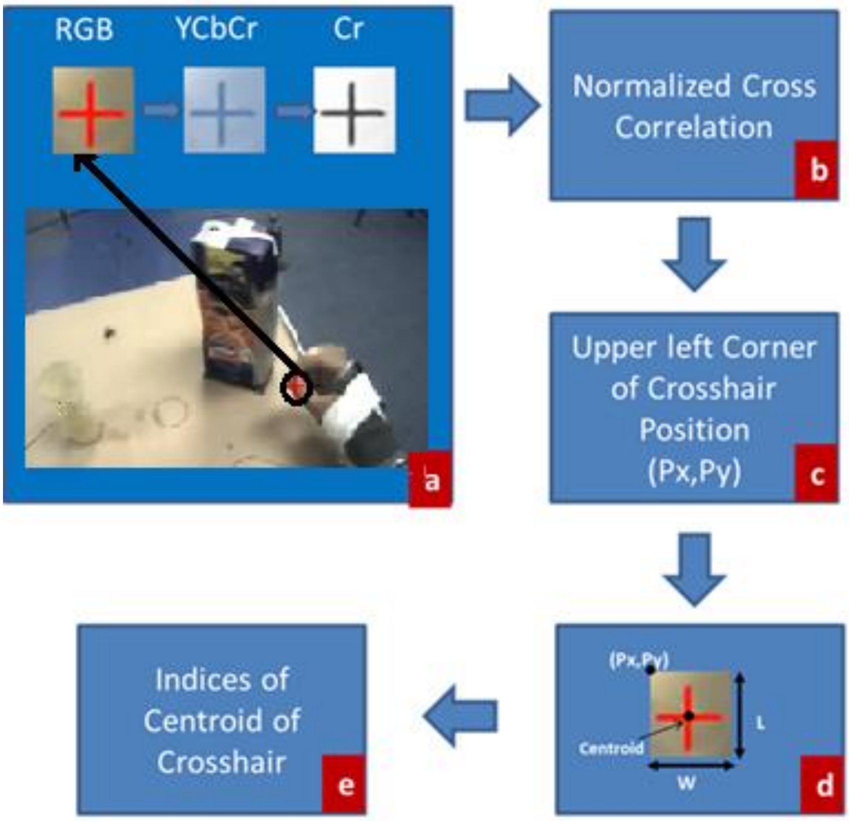



**Figure 6**. Stages of the crosshair detection process

In the present experiment, a threshold of 0.6 (obtained by trial and error) is found appropriate to decide the crosshair existence. When the normalized cross-correlation result is greater than 0.6, the extracting indices are used as the crosshair's centroid. Otherwise, the crosshair is considered missing; this can happen due to eye blinking or saccade. However, if the detection process fails in a particular frame due to an eye blink or saccade, it is automatically labeled "Missing data." Otherwise, indices of the crosshair centroid are obtained, as follows:

$$Centroid = [P_x . P_y] + [\frac{L}{2} . \frac{W}{2}] \tag{6}$$

where Px, Py represents the upper left corner of the crosshair position (see Figure 6c), and L, W are the length and width of the smallest rectangle enclosing the crosshair (see Figure 6d). It is worth mentioning here that the detected crosshair indices and the number of "Missing data" in each video are compared with that of the eye tracker to validate the crosshair detection process's correctness. The obtained results showed that they are 99% identical, validating both the detection algorithm and the chosen threshold. Although it is available in the eye tracker, crosshair detection is used to generalize the developed algorithm for any gaze dataset regardless of its structure or format. Thus, no additional preprocessing or format conversion is required to handle data collected from different eye trackers.

*2.4.2 AOIs Detection in non-key frames*

Unlike AOIs of the keyframes, which are defined manually by a specialist, AOIs in the non-key frames are detected automatically by the algorithm using a simple template matching technique. The matching process moves the template image to all possible positions in a larger input frame's image. This process is demonstrated in the block diagram of Figure 7. As illustrated, the matching process moves the template image, starting from the upper left corner, with a fixed step within each frame to cover all possible positions in the frame's image. It computes a numerical difference that detects how well the template matches the image in that position. The results of these movements are stored in an output array from which the minimum value that represents the closest position is obtained. In this study, a movement step of 4 pixels is considered an acceptable compromise between accuracy and time complexity (i.e., execution time) of the developed algorithm.

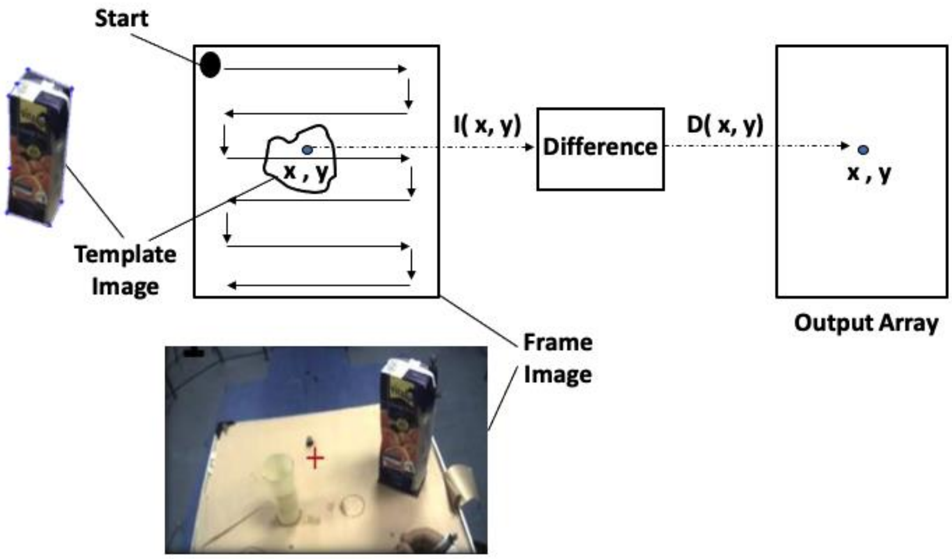



**Figure 7**. Schematic of the templet matching operation

*2.5 Output generation*

Gaze coding involves labeling each frame of the gaze data with the name of a prespecified AOI depending on the gaze crosshair's location. Each AOI typically represents the objects in the scene of the gaze data. The frames can be labeled, after specifying an AOI, only if the crosshair is located within the specified AOI at that particular frame. In this part of the algorithm, the crosshair's relative position to the selected AOI at each frame is determined and counted. Postprocessing of the input data, the total fixation time for each AOI, including "Other" and "Missing data", are determined by the algorithm. This is achieved by iterating along all frames to obtain the crosshair's relative position to the AOIs. The instant and duration of obtaining the existing crosshair position is obtained for each AOI. As illustrated in the flowchart of Figure 8, the search process starts by examining whether a crosshair is detected in the frame or not. If not, the frame is labeled as "Missing data." Otherwise, the search continues for intersections crosshair's centroid and all AOIs. Now, if no intersection exists, the frame is labeled as "Other." Otherwise, the frame is either labeled for the intersected AOI (in case of a single intersection) or labeled for the highest priority AOI to detect multiple intersections. The highest priority AOI is the one that is firstly selected by the specialist.

Once the relative location of crosshair in each frame is determined, the total fixation duration ($F_d$) for each AOI, including "Other" and "Missing data," is calculated by mapping the total number of frames in each area into time, as follows:

$$F_d\,(s) = \frac{Number\ of\ Frames}{Frame\ Rate\ (Hz)} \tag{7}$$

In this study, the gaze video is converted to a group of frames with a 25Hz frame rate. The output is a vector that has the fixation duration (the time spent looking) for each area, "Other" and "Missing data." Therefore, the total length of the output vector is represented by the AOIs selected by the specialist plus the "Missing data" and the "Other" conditions. Finally, the algorithm generates the output in the form of a stacked color bar representing the location and duration of gaze fixation at the AOIs in each trial.

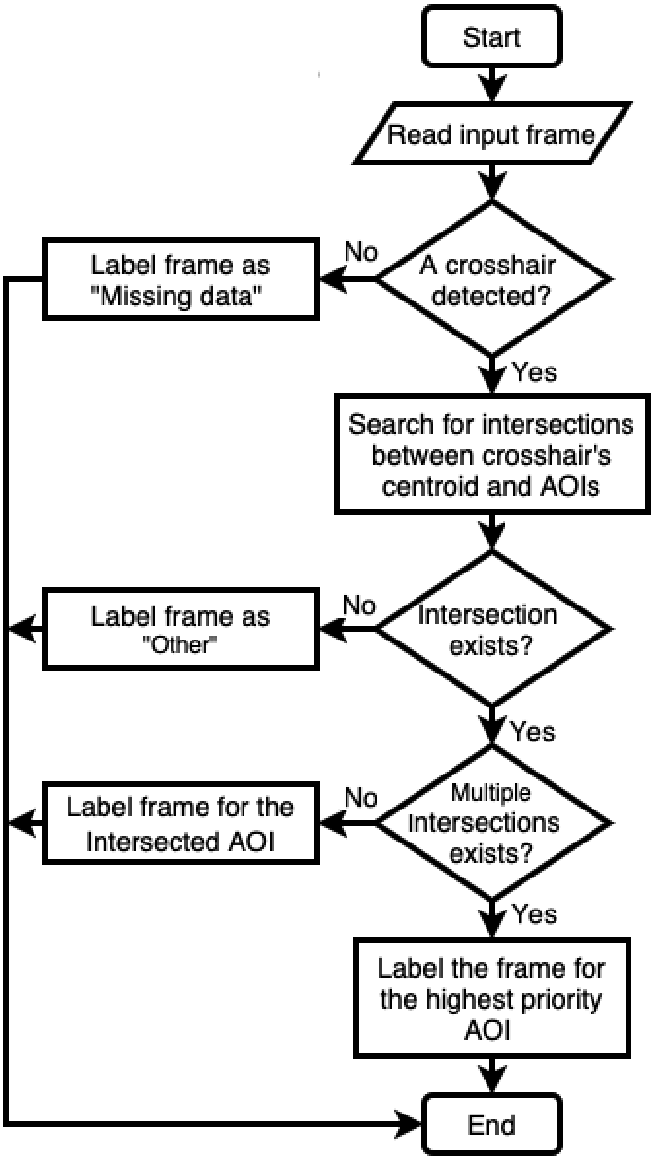



**Figure 8.** A flowchart for frame labeling process

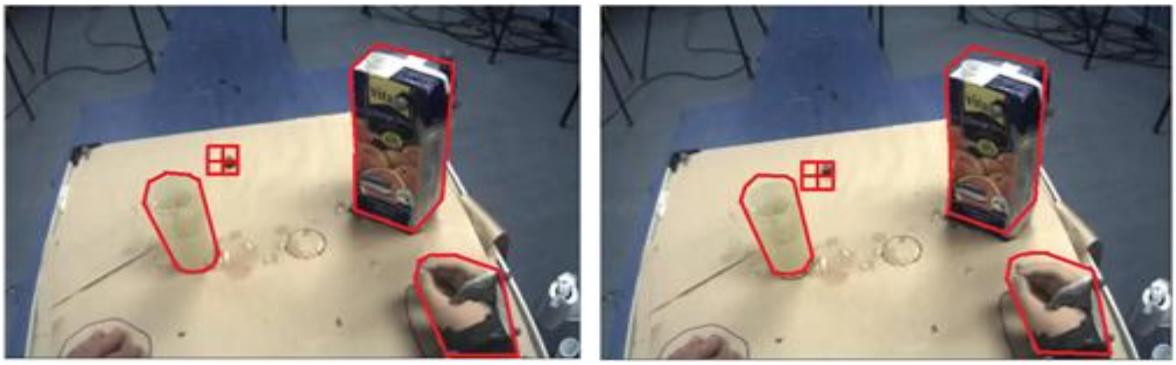
## 3. Results

### *3.1 Experimental results*

As detailed earlier in the previous section, once the crosshair image and gaze video clip are loaded to the developed coding algorithm, the keyframes are detected using a fuzzy-logic controller. Examples of detected keyframes and their order in the gaze video are shown in Figure 9. As illustrated, the shape and color of objects in the scene are changing significantly. These changes cannot be recognized and tracked using image processing techniques alone; thus, the utilization of more intelligent algorithms becomes crucial to address this challenge.


**Figure 9.** Example of detected keyframes and their order in the gaze video

Post manual labeling of the AOIs in keyframes, the algorithm detects the AOIs and crosshair in non-key frames. Figure 10a shows an example of several AOIs (Carton, Glass, and Hand) in a keyframe. Figure 10b shows how the algorithm detects these AOIs in a non-key frame that comes later (i.e., several frames after the keyframe of Figure 10a). For each AOI, the total-fixation duration is calculated and exported as a vector in an Excel sheet.



(a) (b)

**Figure 10**. Examples of labeling/detection of AOIs and crosshair; (a) AOIs and crosshair in a keyframe, (b) AOIs and crosshair in a non-key frame

The primary added value of the developed algorithm is automating the coding process. Once the gaze data is coded, it can be easily presented in different forms and formats, including the stacked bar commonly called a scanpath in behavioral psychology. It is used to demonstrate the gaze sequence and how it differs under different testing conditions [36]. In the present work, we used the stacked color bar to illustrate each frame's fixation position to the hand. This helps specialists to analyze the prosthesis user behavior and elicit clinically meaningful results. Figure 11 shows an example stack bar which demonstrates the relative gaze position to the AOI during the trial time. At the beginning of the trial, the prosthesis user moves his/her gaze around for 2 seconds then starts looking at the cartoon.

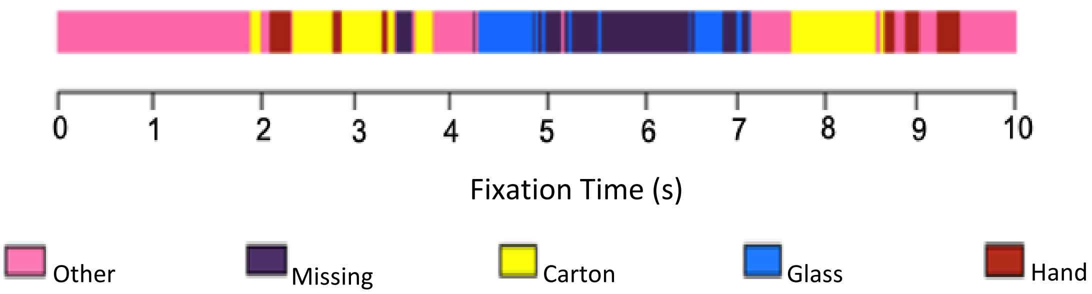



**Figure 11.** Example of the generated stacked bar showing the fixation position at each frame to the hand.

*3.2 Validity analysis*

The validity of the coding results obtained by the developed algorithm is assessed by comparing them to those obtained manually using Cohen's Kappa statistical measure. To avoid exaggerating the sample size, Cohen's Kappa is calculated for each trial's coding results separately, then the average of the obtained results from each trial is considered. The obtained results for each trial's anatomical and prosthetic hands are respectively shown in Tables 2 and 3. It can be noticed that Kappa values in all trials ranged from 0.705 and 1, but, on average, they are mostly higher than 0.8, which can be statistically considered almost perfect agreement [37].

**Table 2.** Cohen's Kappa analysis results of the anatomical hand

| Trial No. | 1 | 2 | 3 | 4 | 5 | Mean |
|---|---|---|---|---|---|---|
| **Subject 1** | 1 | 0.934 | 0.962 | 0.984 | 0.972 | 0.970 |
| **Subject 2** | 0.770 | 0.866 | 0.815 | 0.788 | 0.914 | 0.831 |
| **Subject 3** | 0.934 | 1 | 0.753 | 0.964 | 0.995 | 0.929 |
| **Subject 4** | 0.936 | 0.986 | 1 | 0.969 | 1 | 0.978 |

Overall mean: 0.927

**Table 3.** Cohen's Kappa analysis results of the prosthetic hand

| Trial No. | 1 | 2 | 3 | 4 | 5 | Mean |
|---|---|---|---|---|---|---|
| **Subject 1** | 0.884 | 0.785 | 0.838 | 0.724 | 1 | 0.846 |
| **Subject 2** | 0.741 | 0.792 | 0.790 | 0.760 | 0.793 | 0.775 |
| **Subject 3** | 0.930 | 0.923 | 0.947 | 0.928 | 0.871 | 0.920 |
| **Subject 4** | 0.929 | 0.837 | 0.705 | 0.839 | 0.709 | 0.804 |

Overall mean: 0.836

The confusion matrices that pinpoint the miscoding results between "Carton" and "Glass," and between "Hand" and "Carton" for all trails, using both the anatomical and prosthetic hands, are respectively shown in Tables 4 and 5. The accuracy, precision, false-negative rate (FNR), and false-positive rate (FPR) are calculated from the confusion matrices in these tables and are summarized in Table 6. For these calculations, the manual analysis is considered the actual data, and the computerized analysis is the predicted data. The FNR and FPR represent type-1 and type-2 errors, respectively.

**Table 4.** Miscoding results of the anatomical hand

| | | Computerized Analysis | | | | | |
|---|---|---|---|---|---|---|---|
| | | **Carton** | **Glass** | **Hand** | **Missing** | **Other** | **Total** |
| **Manual Analysis** | **Carton** | 471 | 25 | 0 | 1 | 56 | 553 |
| | **Glass** | 63 | 2662 | 0 | 0 | 121 | 2846 |
| | **Hand** | 0 | 0 | 19 | 0 | 19 | 38 |
| | **Missing** | 0 | 0 | 0 | 614 | 0 | 614 |
| | **Other** | 15 | 31 | 5 | 6 | 3837 | 3894 |
| **Total** | | 549 | 2718 | 24 | 621 | 4033 | 7945 |

**Table 5**. Miscoding results of the prosthetic hand

| | | Computerized Analysis | | | | | Total |
|---|---|---|---|---|---|---|---|
| | | Carton | Glass | Hand | Missing | Other | |
| Manual Analysis | Carton | 1844 | 17 | 64 | 0 | 369 | 2294 |
| | Glass | 63 | 1054 | 0 | 0 | 205 | 1322 |
| | Hand | 143 | 0 | 292 | 0 | 135 | 570 |
| | Missing | 0 | 0 | 1 | 1512 | 0 | 1513 |
| | Other | 115 | 46 | 61 | 4 | 4057 | 4283 |
| Total | | 2165 | 1117 | 418 | 1516 | 4766 | 9982 |

**Table 6**. Accuracy, precision, FNR and FPR metrics of the miscoding results

| | Accuracy (%) | Precision (%) | FNR (%) | FPR (%) |
|---|---|---|---|---|
| **Anatomical hand** | 95.70 | 91.33 | 04.3 | 1.43 |
| **Prosthetic hand** | 87.75 | 86.85 | 12.25 | 3.75 |

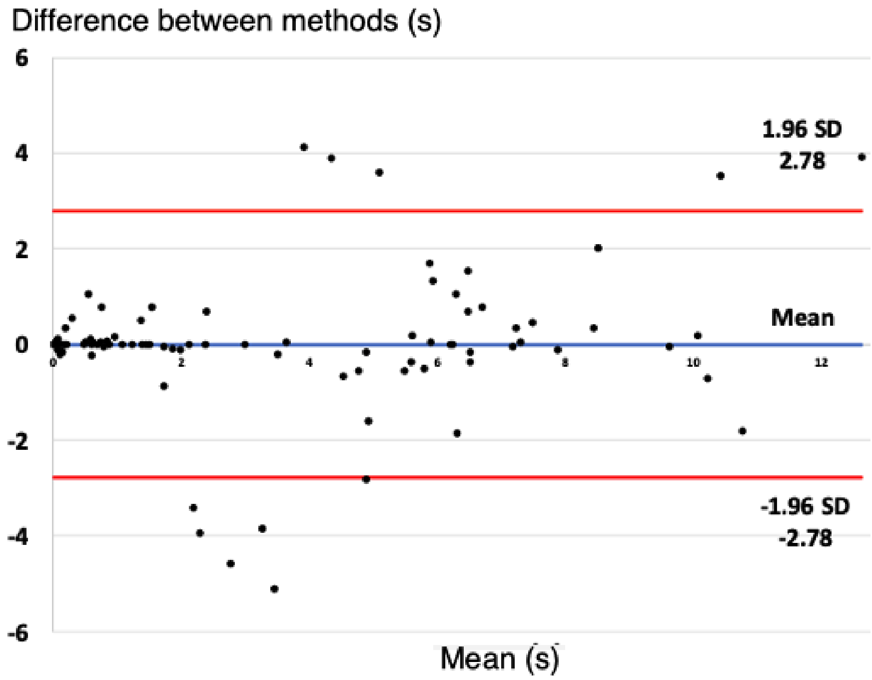
The agreement between the manual and computerized rating methods is assessed using the intra-class correlation coefficient (ICC). The ICC estimate for each AOI's total fixation duration and its 95% confidence intervals are calculated using SPSS statistical package version 23 (SPSS Inc, Chicago, IL). A two-way random-effect model based on single ratings and absolute agreement is used. The estimated agreement index is found to be 0.908 with 95% confidence intervals (0.867, 0.937) for the anatomical hand and 0.923 with 95% confidence intervals (0.888, 0.948) for the prosthetic hand. The Bland-Altman plots for both the anatomical and prosthetic hands are shown in Figure 12. As illustrated, the difference between fixation duration is plotted against the mean fixation duration at each AOI across all trials, as identified by the rating methods.

*3.3 Coding-process efficiency*

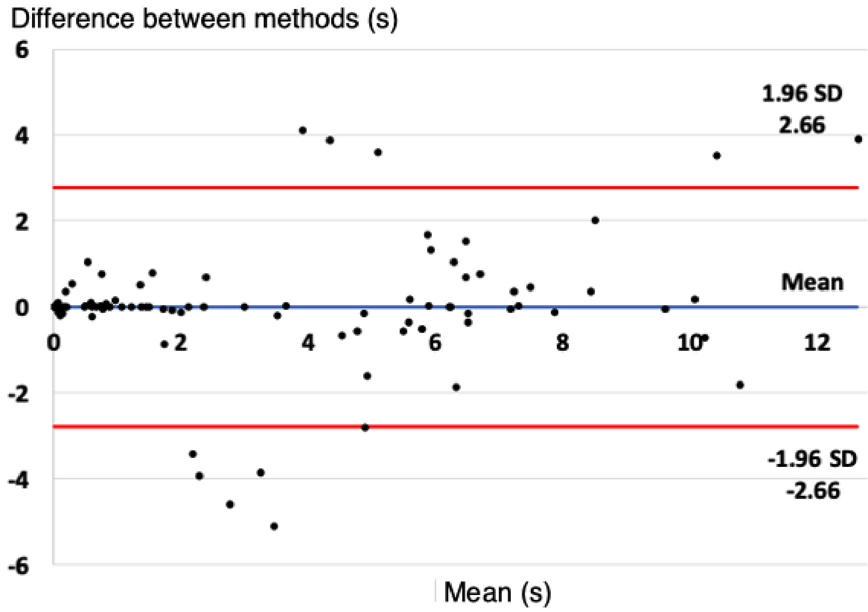
The time complexity analyses of the developed algorithms representing both the compiled time and execution time are calculated, as suggested in [38]. However, as the compiled time is not involved in the algorithm's real-time operation, the time estimation is limited to the algorithm's execution time. Estimating the execution time is performed by running the algorithm under test through a specific number of loop iterations. Timestamps of the start ($T_{start}$) and end ($T_{end}$) instants of the loop are recorded, and the execution time ($T_{exec}$) is then calculated as suggested in [39]:

$$T_{exec} = \frac{T_{end} - T_{start}}{n} \quad (8)$$

where n is the number of loop iteration. In the present analysis, n = 100,000 is considered adequate to estimate the average $T_{exec}$ with acceptable accuracy. Finally, the computerized coding process's total time is obtained by adding the specialist's time to obtain the keyframes and AOIs to the execution time estimated in (8). Estimation of $T_{exec}$ is carried out using a laptop with Intel(R) Core (TM) i7-4770 M CPU @ 3.4 GHz, 4.0 GB RAM, and 64-bit Windows 10 operating system, and the code is run on MATLAB with real-time priority mode. Further reduction in the $T_{exec}$ is possible using a more time-efficient programming language such as C/C++ or assembly programming compared to the MATLAB.




(a) Anatomical hand




(b) Prosthetic hand

**Figure 12.** Bland-Altman plots of the total fixation duration on AOI for the manual and computerized methods

A comparison between the computerized and manual coding efficiency is shown in Table 7. The timeframes reported in this table represent the average times taken by two specialists to perform the given tasks (i.e., watching the video and counting the number of crosshair appearance on each AOI). As illustrated, a significant timesaving is achieved by the developed algorithm as compared to the traditional manual coding. The central part of the time saving is reflected by reducing the number of keyframes that require a specialist's intervention towards generating the stacked-bar output. This development reduces the data-preparation time and allows the specialist to focus on the analysis rather than extracting the fixation data.

**Table 7.** Time efficiency comparison between the proposed system and the manual coding

| Coding activity | Coding time | | |
|---|---|---|---|
| | Manual (min) | Computerized (min) | |
| | | Keyframe Selection | Program Execution |
| Video 1: Duration: 16s, number of frames: 400 | 11 | 0.5 | 5.5 |
| Video 2: Duration: 18s, number of frames: 450 | 12 | 0.6 | 6.4 |

## 4. Discussion

In this work, a new algorithm capable of detecting gaze fixation at predefined areas of interest in given gaze data has been designed, developed, and tested successfully. Two test conditions, using anatomical and prosthetic hands, are considered for coding in this study. Each test is based on five trials' data, resulting in 40 coded trials of gaze data. We used an intelligent algorithm based on a fuzzy controller to automate fixation position detection on trials. This development has improved the process of obtaining clinically meaningful findings from the gaze data by saving time and improving reliability by obtaining consistent results. The main findings and limitations of the work presented in this paper are discussed as follows.

### *4.1 Main findings*

The developed algorithm is considered an essential step towards fully automating the dynamic-gaze video analysis. It contributed to automating the process of finding AOIs in non-key frames and generating the required stacked bar that the specialists require. AOIs of the keyframes are still defined manually with the help of a specialist. Such a human intervention is considered necessary to deal with the alterations in the shape of the AOIs. These alterations are mainly caused by the dynamic changes in the location and orientation of the AOIs relative to the tracking camera, see Figure 9. On average, about ten frames in each gaze video required user intervention, representing only a small portion (2.6%) of the entire manual analysis process.

The developed algorithm is tested by analyzing the gaze data collected during completing a simple Activities of Daily Living (ADL) task under two testing conditions using both a prosthesis and anatomically intact hands. The gaze data described in the methods (Section 2) is collected to characterize prosthetic users' gaze behavior to understand prosthetic control's underlying process [15]. The difficulty the team has faced while analyzing the gaze data to quantify the gaze fixation patterns on the scene is the actual rationale behind developing the proposed algorithm. Such a difficulty has resulted from the fact that the scene comprises many AOIs that can be fixated. In particular, the interest is to count the gaze duration at each AOI and the sequence of this gaze fixation (known as the scan-path) during task completion.

Comparable results are obtained when comparing the algorithm's coding results with those obtained manually. The developed algorithm showed a comparable accuracy of coding with high precision, especially when the anatomical hand is used. Tables 1 and 2 indicated high agreement between the algorithm coding and the manual coding for all testing conditions as

Cohen’s Kappa values ranged from 0.705 to 1.0. However, Cohen's Kappa is slightly higher for the anatomical hand than that of the prosthetic hand. This difference may be due to the relatively short task duration of the anatomical hand usage and fewer keyframes. Besides, when the anatomical hand is used, the gaze involves fewer transitions between AOIs; therefore, the possibility of mislabeling/disagreement is reduced. The ICC results also demonstrated high agreement between the two methods, as illustrated in the Bland-Altman plot of Figure 12. It can be noticed that the data points for both hands are closely scattered around the mean.

### *4.2 Challenges and limitations*

#### 4.2.1 *Challenges*

Design and development of the proposed coding algorithm have dealt with and addressed some technical challenges, including the following:

a) Dynamic changes in the location and orientation of the AOIs - this challenge is addressed by using fuzzy logic to support the algorithm in making decisions similar to human thinking. The small weight given to the color change helped minimize the lighting effect on the performance, unless it had sufficient change (energy).

b) Identification of accurate crosshair borders - the algorithm defines the crosshair as a square that may cause bias in its center, especially when the crosshair intersects with two AOIs. In this case, the algorithm considers the fixation at the AOI listed first by the specialist.

c) The color contrast between the AOIs and the background of a similar color. For example, the "Glass" increases the difficulty of visually distinguishing it from the surrounding area. This problem is tackled by using an anisotropic diffusion filter, which helped remove the small details (e.g., texture and internal edges) without affecting the main object's edges.

#### *4.2.2 Limitations*

The developed algorithm represents a significant step towards the automation of dynamic-gaze video analysis for upper-limb prosthesis users. However, for full automation of the coding analysis process, which is not the objective of this study, there is a need for further improvements. For example, obtaining the AOIs in the keyframes still requires the intervention of a specialist. The mislabeling may occur between the "Other" backgrounds, and the AOIs that need further improvements. A miscoding between “Carton” and “Glass”, and between “Hand” and “Carton” can also happen with the prosthesis during the task completion. These limitations can be avoided by prioritizing the intersected AOIs.

The developed algorithm deals with the physical AOIs border, but the nearby areas may also contain important information beyond the scope of the present study. Such information can be of a particular importance when the gaze fixation is located within the vicinity of the object(s) of interest rather than at the object itself. For instance, perhaps, it is reasonable to assume that the participant is looking at the carton and/or glass to check onto the pouring action. Alternatively, gaze might be kept in an area close to both areas in order to achieve the same intended function (i.e., check onto the pouring action). To address this limitation, defining a “functional AOI” sometimes relates to an action (i.e., pouring water). This functional AOI can comprise several objects that need to complete the action (i.e., carton, glass, and their vicinity). These limitations and others are currently part of the ongoing work of the authors.

## 5. Conclusion

In this paper, a bespoke algorithm for detecting gaze fixation in given gaze data has been designed, developed, and tested successfully. The results obtained using the developed algorithm agree with those obtained manually but found to be significantly faster and less prone to human errors when compared to the manual coding. Statistical analysis showed that Cohen's Kappa ranges from 0.705 to 1.0. Moreover, based on the ICC, the agreement index between computerized and manual coding methods is found to be (i) 0.908 with 95% confidence intervals (0.867, 0.937) for the anatomical hand and (ii) 0.923 with 95% confidence intervals (0.888, 0.948) for the prosthetic hand. A Bland-Altman plot also showed that all data points are closely scattered around the mean. These findings confirm the validity and effectiveness of the developed coding algorithm. This confirms the validity of the developed coding algorithm.

The developed algorithm demonstrated a significant step forward for full automation of the dynamic-gaze video analysis process; thus, reliable, accurate and clinically meaningful findings can be obtained in a short period as compared to the existing tedious and time-consuming manual analysis. However, further investigations are still required to improve the developed coding algorithm's structure and performance by conducting a more comprehensive clinical study. A more intelligent machine learning approach, such as neural networks, can also be adopted to automate the coding process of keyframes in the gaze data. Other factors that may affect the collection and analysis of gaze data such as gamification of eye exercises for evaluating the eye fatigue, gender, age, color, position, and stress can be considered in future experiments. These potential experiments and others are currently part of the authors' ongoing work and will be the subject of a future publication(s).

## Declaration of competing interest

There is no conflict of interest.

## References

[1] C. M. Huang, S. Andrist, A. Sauppé, B. Mutlu, Using gaze patterns to predict task intent in collaboration. Front Psychol. vol. 6:1049, 2015, pp. 1-12.

[2] A. Shumway-Cook, M. H. Woollacott, Motor control: translating research into clinical practice 3ed. Philadelphia, Pennsylvania LWW, 2007.

[3] E. Lehtonen, O. Lappi, H. Otkanen, H. Summala, Look-ahead fixations in curve driving. Ergonomics. vol. 56(1), 2013, pp. 34-44.

[4] S. Tuhkanen, J. Pekkanen, E. Lehtonen, O. Lappi, Effects of an active visuomotor steering task on covert attention , Journal of Eye Movement Research, vol. 12, 2019, pp. 1-19.

[5] A. Kővári, J. Katona, C. Pop. Quantitative analysis of relationship between visual attention and eye-hand coordination. Acta Polytechnica Hungarica. vol. 17, 2020, pp. 77-95.

[6] C. A. Rothkopf, D. H. Ballard, and M. M. Hayhoe, Task and context determine where you look, J. Of Vision, vol. 7, 2007, pp. 1-20.

[7] A. M. Brouwer, V. H. Franz, and K. R. Gegenfurtner, Differences in fixations between grasping and viewing objects, J. of Vision, vol. 9, 2009, pp. 1-24.

[8] L. Desanghere, J. J. Marotta, Graspability of objects affects gaze patterns during perception and action tasks, J. Exp Brain Res, vol. 212, 2011, pp. 177-187.

[9] C. C. Williams, M.S. Castelhano, The Changing Landscape: High-Level Influences on Eye Movement Guidance in Scenes. Vision, vol. 3, 2019, pp. 1-20.

[10] J. S. Cuellar, G. Smit, P. Breedveld, A. Zadpoor, D. Plettenburg, Functional evaluation of a non-assembly 3D-printed hand prosthesis, Proc. of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, vol. 233(11), 2019, pp. 1122–1131.

[11] H. Wang, Z. Pi, W. Hu. The instructor's gaze guidance in video lectures improves learning. Journal of Computer Assisted Learning, 35(1), 2019, pp. 42–50.

[12] D. J. Harris, G. Buckingham, M. R. Wilson, J. Brookes, F. Mushtaq, M. Mon-Williams, S. J. Vine, The effect of a virtual reality environment on gaze behaviour and motor skill learning, Psychology of Sport and Exercise, vol 50, 2020, pp. 441-457.

[13] H. Ashraf, MH. Sodergren, N. Merali, G. Mylonas, H. Signh, H. Darzi, Eye-tracking technology in medical education: a systematic review, Med Teach, vol. 40, 2018, pp. 62–69.

[14] F. Sharafian; M. Shahbazi; S. T. Boroujeni, Effect of technical and quiet eye training on the gaze behavior and long-term learning of volleyball serve reception in 10 to 12-year-old female, Int. Journal of School Health, 6, 3, 2019, pp. 1-6.

[15] M. M. Sobuh, L. P. Kenney, A. J. Galpin, S. B. Thies, J. McLaughlin, J. Kulkarni, and P. Kyberd, Visuomotor behaviours when using a myoelectric prosthesis, J. Neuroeng Rehabil, vol. 11, 2014, pp. 1-11.

[16] O. V. Otr, H. A. Reinders-Messelink, R. M. Bongers, H. Bouwsema and C. K. Van Der Sluis, The i-limb hand and the dmc plus hand compared: A case report, J. Prosthet Orthot Int. 34(2), 2010, pp. 227-230.

[17] A. Chadwell, L. Diment, M. Micó-Amigo, D. Ramírez, A. Dickinson, M. Granat, L. Kenney, S. Kheng, M. Sobuh, R. Ssekitoleko, P. Worsley, Technology for monitoring everyday prosthesis use: a systematic review. J. NeuroEngineering Rehabil, 17(93), 2020, pp. 17-33.

[18] J. V. Parr, S.J. Vine, M.R. Wilson, N. R. Harrison, G. Wood, Visual attention, EEG alpha power and T7-Fz connectivity are implicated in prosthetic hand control and can be optimized through gaze training". J NeuroEngineering Rehabil, 16(52), 2019, pp. 16-36.

[19] A. Chadwell, L. Kenney, S. Thies, A. Galpin, J. Head, The reality of myoelectric prostheses: understanding what makes these devices difficult for some users to control, Front Neurorobot. vol. 10, 2016, pp 1-21.

[20] P. K. Muthumanickam, K. Vrotsou, A. Nordman, J. Johansson and M. Cooper, Identification of temporally varying areas of interest in long-duration eye-tracking datasets, IEEE Transactions on Visualization and Computer Graphics, vol. 25, no. 1,2019, pp. 87-97.

[21] L. Chukoskie et. al., Quantifying Gaze Behavior During Real-World Interactions Using Automated Object, Face, and Fixation Detection, IEEE Transactions on Cognitive and Developmental Systems, vol. 10, no. 4, 2018, pp. 1143-1152.

[22] M. Mohseni, V. Shalchyan, M. Jochumsen, I. K. Niazi, Upper limb complex movements decoding from pre-movement EEG signals using wavelet common spatial patterns, J. Computer Methods and Programs in Biomedicine, vol. 183, 2020, pp 1-10.

[23] H. Bouwsema, P. Kyberd, W. Hill, C. P. Van der Schans, and R. Bongers, Determining skill level in myoelectric prosthesis use with multiple outcome measures, J. Rehabil Res Dev., vol. 49, 2012, pp. 1331-1348.

[24] A. T. Duchowski, Eye tracking methodology: theory and practice. New York: Springer, 2007.

[25] P. J. Kyberd, P.H. Chappell, The Southampton hand: an intelligent myoelectric prosthesis. J. Rehabil Res Dev 1994, pp. 326-334.

[26] J. V. V. Parr, S. J. Vine, N. R. Harrison, G. Wood, Examining the spatiotemporal disruption to gaze when using a myoelectric prosthetic hand, J. Motor Behavior, 50:4, (2018) pp. 416-425.

[27] E. B. Lavoie, A. M. Valevicius, Q. A. Boser, O. Kovic, A. H. Vette, P. M. Pilarski, J. S. Hebert, C. S. Chapman, Using synchronized eye and motion tracking to determine high-precision eye-movement patterns during object-interaction tasks. J. Vision (2018); pp.18-24.

[28] H. E. Williams, C. S. Chapman, P. M. Pilarski, A. H. Vette, J. S. Hebert, Gaze and movement assessment (GaMA): inter-site validation of a visuomotor upper limb functional protocol, PLOS ONE, 2019, pp. 14-26.

[29] H. E. Williams, Q. A. Boser, P. M. Pilarski, C. S. Chapman, A. H. Vette and J. S. Hebert, "Hand function kinematics when using a simulated myoelectric prosthesis, IEEE 16th Int. Conf. on Rehabilitation Robotics (ICORR), Toronto, ON, Canada, 2019, pp. 169-174.

[30] J. S. Hebert, Q. A. Boser, A. M. Valevicius, H. Tanikawa, E. B. Lavoie, A. H. Vette, P. M. Pilarski, C. S. Chapman, Quantitative eye gaze and movement differences in visuomotor adaptations to varying task demands among upper-extremity prosthesis users. JAMA Netw Open, 2019, pp. 2-11.

[31] M. Vasiljevas, T. Gedminas, A. Ševčenko, M. Jančiukas, T. Blažauskas and R. Damaševičius, Modelling eye fatigue in gaze spelling task, 2016 IEEE 12th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, 2016, pp. 95-102.

[32] Y. Yamada, M. Kobayashi, Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults, Artificial Intelligence in Medicine, 2018, pp. 39-48.

[33] P. PERONA and J. MALIK, Scale-space and edge detection using anisotropic diffusion," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12(7), 1990, pp. 629-639.

[34] A. Baraldi F and F. Parmiggiani, Aninvestigation of the textural characteristics associated with gray level cooccurrence matrix statistical parameters," IEEE Trans. Geosci. Remote Sensing, 1995, pp. 293-304.

[35] M.S. Iraji, and A. Yavari, Skin color segmentation in fuzzy ycbcr color space with the mamdani inference, American Journal of Scientific Research, 2011, pp. 131-137.

[36] N. C Anderson, F. Anderson, A. Kingstone, W. F. Bischof, “A comparison of scanpath comparison methods, J. Behavior Research Methods, vol. 47, 2015, pp. 1377–1392.

[37] M. L. McHugh, Interrater reliability: the kappa statistic, Biochemiamedica, vol. 22, 2012. pp. 276-282.

[38] M. A. Al-Taee, T. Abu-Jabarah, I. Sultan, A. Al-Karmi, Real-time capturing and visualization of human hand pose using low-cost web cameras, J. Dirasat Engineering Science, vol. 32 (1), 2010, pp. 28-42.

[39] Q. Al-Jubouri, R. J. Al-Azawi, M. Al-Taee, I Young, Efficient individual identification of zebrafish using Hue/Saturation/Value color model, The Egyptian Journal of Aquatic Research, vol. 44 (4), 2018, pp. 271-277.