Using 3D heatmaps to visualize the gaze distributions of observers watching a moving subject

Fuyuko Iwasaki, Shouta Hioki, Shunsuke Yoneda, Michiko Inoue and Masashi Nishiyama[1]

***Abstract*— We propose a method to visualize the measured gaze distribution of observers asked to perceive the dynamism of a subject's movements in a sports video. This visualization method uses a three-dimensional heatmap on the surface of a human body model. An existing method generates the heatmap using gaze measurements on a body surface in a still image. However, this method does not handle changes over time in a subject's posture in a video. Furthermore, this method does not visualize gaze in the region surrounding the subject's body. Our method calculates the angle between the gaze direction and vertex position to visualize the gaze distribution on the body surface and surrounding regions. Experimental results confirmed that our method visualizes not only the gaze distribution on the surface region but also that in the surrounding region. We also confirmed that it is possible to visualize the gaze distribution over a subject's movements in a video without depending on changes in posture using a standard human body model.**

I. Introduction

In recent years, sports video content on a virtual large display, for example, a head-mounted display (HMD), has been eagerly anticipated. We consider a situation in which an audience watches a video of an athlete performing sports actions using an HMD. Figure 1 shows an example of such a situation. When the audience is affected by the dynamism in an athlete's movement, their gaze is focused on the part of the athlete's body that provides the cue for this dynamism. To determine which body part is the focus of gaze in various sports videos, an analyst can measure the gaze distribution of the audience. In the following, we refer to the athlete in the video as the subject and the audience who watches the athlete's movements in the video as the observer.

When an analyst is studying gaze distribution, there is a need to visualize in an easy-to-understand manner which body part of a subject an observer's gaze is focused on. A two-dimensional (2D) heatmap superimposed on a still image is commonly used to visualize the gaze distribution. For example, a 2D heatmap has been used in existing analytical studies [1], [2], [3] to investigate the gaze distribution when a subject in a still image is used as the stimulus. However, the posture of a subject performing a sports action continuously changes when a subject in a video is used as the stimulus. Specifically, to study gaze distribution using a conventional 2D heatmap, differences in the subject's posture at different temporal points in the video must be taken into account.

Hence, a visualization method is needed so that analysts can instantly understand the gaze distribution without taking into account differences in posture, even when the subject's posture changes over time. Here, we consider a visualization method proposed by Inoue et al. [4]. In this method, a standard human body model representing a three-dimensional (3D) human shape with a typical posture is estimated from the subject in a still image, and the gaze distribution on the surface of the body model is visualized using a 3D heatmap. This method enables the gaze distribution in still images to be evaluated without the need to consider differences in subject posture. However, the existing method considers only subjects in still images, not subjects in videos, who continuously change their posture. In addition, there are significant challenges with the existing method, which does not visualize the gaze measured in the surrounding regions because it visualizes only the gaze measured in the subject's surface region. The "surface region" here refers to a set of pixels corresponding to the subject, as estimated in the video, and the "surrounding region" refers to the set of pixels in the vicinity of the subject. Figure 2 shows examples of these regions.

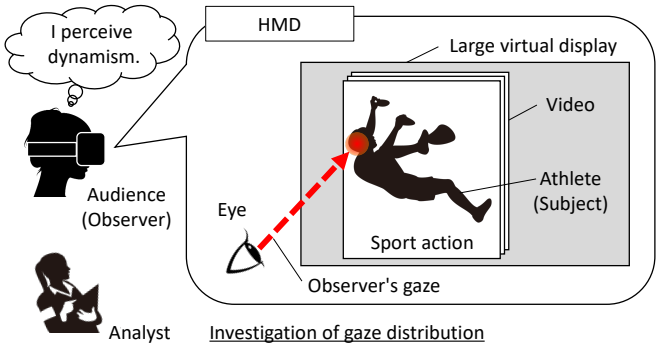

Fig. 1. When watching a sports video on an HMD, an observer is affected by the dynamism in the subject's movement. The analyst investigates where the observer's gaze is focused on the subject's body.

In this paper, we propose a method to visualize the gaze distribution measured in a subject's surface and surrounding regions using a 3D heatmap in which the subject's posture is standardized over time. In the experiments, we asked observers to assess the dynamism of a subject's movements in a sports video. We measured and then visualized the gaze distribution of the observers while they watched the video. Experimental results show that, in contrast to the existing method, which only visualizes the measured gaze distribution on the surface region, our method visualizes this distribution in both the surface and surrounding regions as a 3D heatmap. Furthermore, we confirmed that our method can visualize the

[1]The authors are with the Faculty of Engineering, Tottori University, Japan. nishiyama@tottori-u.ac.jp

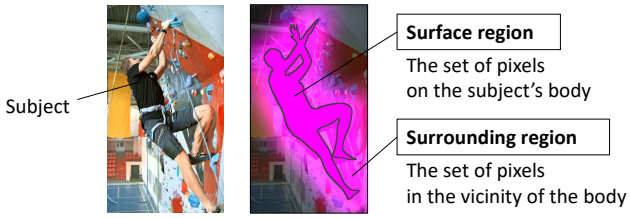

Fig. 2. Surface and surrounding regions of a subject's body.

measured gaze distribution throughout the entire duration of the video, in contrast to the existing method, which cannot cope with changes in the subject's posture over time.

II. Visualizing Measured Gaze Distribution in the Surface and Surrounding Regions

A. Overview of the existing method

The existing method [4], as described in Section I, has two limitations: it cannot handle a sports video, and it does not visualize the measured gaze distribution in the region surrounding the human body. We first resolve the latter limitation, which arises at each time point, i.e., at each video frame, and then address the former limitation in the temporal direction.

To clarify the limitation, we present an overview of the existing method [4], which consists of the following two steps. In step 1, the eye tracker measures the gaze direction vector $\boldsymbol{g}(t,o)$ of the observer o looking at a still image at time t. Here, $\boldsymbol{g}(t,o)$ is a unit vector whose starting point is the center between the observer's eyes. In step 2, the level of attention is calculated. This indicates how much the gaze is focused on each vertex $\boldsymbol{v}(t)$ of the human body model estimated from the still image using the gaze direction vector $\alpha\boldsymbol{g}(t,o)$ multiplied by constant $\alpha(>0)$. Note that SMPL [5] is used as the human body model, which is represented by posture and body shape parameters. The human body model consists of a set $\mathcal{V}(t)$ of 3D vertex position vectors $\boldsymbol{v}(t)$ on the body surface and the adjacency relations among the vertices. Vector $\alpha\boldsymbol{g}(t,o)$ is determined by the intersection with the surface of the human body model. The existing method computes the attention $d(\boldsymbol{v}(t),\boldsymbol{g}(t,o))$ based on the distance ϵ as

$$d(\boldsymbol{v}(t),\boldsymbol{g}(t,o)) = \exp(-\frac{\epsilon^2}{2\sigma^2}), \tag{1}$$

where $\epsilon = \|\boldsymbol{v}(t) - \alpha\boldsymbol{g}(t,o)\|_2$. Smaller values of distance ϵ, indicate higher attention with respect to $\boldsymbol{v}(t)$, and the larger the distance, the smaller the amount of attention degree. The existing method [4] is based on distance because the visualization of gaze distribution in a conventional 2D heatmap is based on the distance from the position of the measured gaze on the image plane.

B. How the existing method handles the gaze direction

This section discusses why the existing method [4] cannot handle the gaze distribution measured in the surrounding region. In the existing method, the gaze direction vector $\alpha\boldsymbol{g}(t,o)$ is calculated under the strong constraint that it must intersect with the surface of the human body model estimated from the still image. Figure 3(a) shows a case that satisfies this constraint. When intersection with the human body model occurs, the attention $d(\boldsymbol{g}(t,o),\boldsymbol{v}(t))$ can be computed because the distance ϵ in Eq. (1) between gaze direction vector $\alpha\boldsymbol{g}(t,o)$ and the vertex position vector $\boldsymbol{v}(t)$ of the human body model can be calculated. By contrast, Figure 3(b) shows a case that does not satisfy this constraint. When there is no intersection with the human body model, the attention $d(\boldsymbol{g}(t,o),\boldsymbol{v}(t))$ cannot be computed because distance ϵ cannot be defined. Hence, when gaze is measured in the surrounding region, the existing method's 3D heatmap does not adequately visualize the gaze distribution.

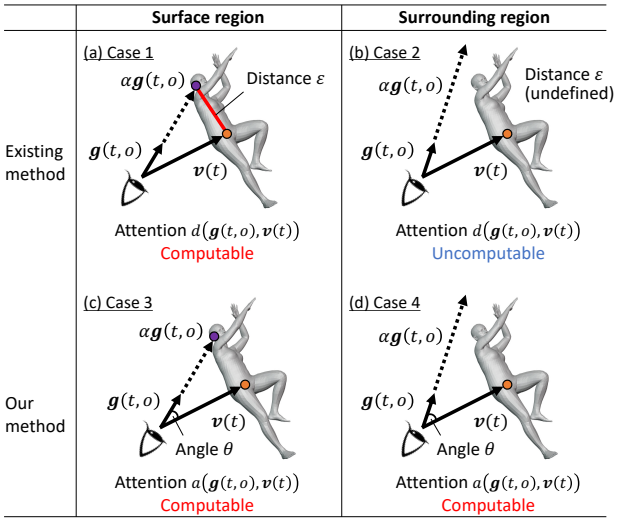

Fig. 3. Examples of calculating the level of attention in the surface and surrounding regions using the existing method and our method.

We consider two possible situations that occur when the observer's gaze direction vector $\alpha\boldsymbol{g}(t,o)$ passes through the surrounding region. The first situation occurs when the observer looks at a region in the background near the subject's contour. The second situation occurs when the observer views an edge region, such as the subject's hair or clothing, that is not represented by the SMPL human body model. When these situations occur, the observer simultaneously views the human body in the nearby surface region and its surrounding regions. Specifically, the observer is not just looking in the gaze direction $\boldsymbol{g}(t,o)$ but also at the vertex $\boldsymbol{v}(t)$ of the human body model that exists in the vicinity of $\boldsymbol{g}(t,o)$. Therefore, when visualizing the gaze distribution using a 3D heatmap, it is necessary to consider both the surface and surrounding regions simultaneously.

C. Attention to the surface and surrounding regions in our method

Instead of using distance ϵ, our method uses the angle θ between the gaze direction vector $\boldsymbol{g}(t,o)$ and vertex position vector $\boldsymbol{v}(t)$ to calculate the level of attention $a(\boldsymbol{v}(t),\boldsymbol{g}(t,o))$,

which indicates how often the gaze is focused on the subject's body. Figure 3(c) shows a case where the measured gaze distribution falls within a surface region. We can calculate $a(\boldsymbol{g}(t,o),\boldsymbol{v}(t))$ since the angle θ between gaze direction vector $\boldsymbol{g}(t,o)$ and vertex position vector $\boldsymbol{v}(t)$ has been obtained. Furthermore, Figure 3(d) shows a case where the measured gaze distribution falls within the surrounding region. We can also calculate $a(\boldsymbol{g}(t,o),\boldsymbol{v}(t))$ in this case because angle θ has been defined. Thus, our angle-based method can simultaneously handle surface and surrounding regions when visualizing the gaze distribution using a 3D heatmap.

D. Calculating the level of attention at each time point

Our method calculates the attention $a(\boldsymbol{v}(t),\boldsymbol{g}(t,o))$ of observer o at time t using the angle θ between gaze direction vector $\boldsymbol{g}(t,o)$ and vertex position vector $\boldsymbol{v}(t)$ as follows:

$$a(\boldsymbol{v}(t),\boldsymbol{g}(t,o)) = \exp(-\frac{\theta^2}{2\sigma^2}). \tag{2}$$

Angle θ is represented using the inner product as

$$\cos\theta = \frac{(\boldsymbol{v}(t),\boldsymbol{g}(t,o))}{\|\boldsymbol{v}(t)\|_2\|\boldsymbol{g}(t,o)\|_2}. \tag{3}$$

The smaller the angle θ, the greater the attention with respect to $\boldsymbol{v}(t)$; the larger the angle θ, the smaller the attention. The level of attention is set to 0 for $\boldsymbol{v}(t)$ on the other side of the human body model, which cannot be seen directly from the observer's viewpoint.

Our method calculates the vertex attention probability $p(\boldsymbol{v}(t)|\mathcal{S}(t))$, which represents how much the gaze tends to be focused on vertex $\boldsymbol{v}(t)$ of the human body model when viewing a frame $\mathcal{S}(t)$ at time t in a video. The vertex attention probability $p(\boldsymbol{v}(t)|\mathcal{S}(t))$ is defined using the level of attention $a(\boldsymbol{v}(t),\boldsymbol{g}(t,o))$ as follows:

$$p(\boldsymbol{v}(t)|\mathcal{S}(t)) = \frac{1}{W}\sum_{o\in\mathcal{O}} a(\boldsymbol{v}(t),\boldsymbol{g}(t,o)), \tag{4}$$

where $\mathcal{O}$ is a set of observers consisting of O individuals. Coefficient W is expressed as

$$W = \sum_{\boldsymbol{v}(t)\in\mathcal{V}(t)}\sum_{o\in\mathcal{O}} a(\boldsymbol{v}(t),\boldsymbol{g}(t,o)). \tag{5}$$

Moreover, $p(\boldsymbol{v}(t)|\mathcal{S}(t))$ satisfies the following equation:

$$\sum_{\boldsymbol{v}(t)\in\mathcal{V}(t)} p(\boldsymbol{v}(t)|\mathcal{S}(t)) = 1. \tag{6}$$

E. Marginalization of vertex attention probability over the duration of the video

We discuss the remaining limitation of the existing method [4], which is that it is not suitable for sports videos. The subject's posture changes continuously over time in such videos, as described in Section I. The analyst needs to instantly grasp where on the subject's body the observers were looking throughout the entire duration of the video. Our method standardizes the different postures over time using a standard human body model with a common posture at time t, and the vertex attention probabilities are marginalized over the duration of the video. This marginalization allows the analyst to investigate the gaze distribution throughout the entire duration of the video by instantly viewing a single 3D heatmap visualized on a standard human body model.

The following describes the details of this marginalization. The human body model estimated from frame $\mathcal{S}(t)$ at time t consists of a set $\mathcal{V}(t)$ of vertices $\boldsymbol{v}(t)$. The standard human body model consists of a set $\tilde{\mathcal{V}}$ of vertices $\tilde{\boldsymbol{v}}$ under the condition that the posture is constant over time. The adjacencies of $\boldsymbol{v}(t)$ on the human body surface are equal to those of $\tilde{\boldsymbol{v}}$ on the standard human body model, although the 3D position vectors are different. Therefore, the vertex attention probability $p(\boldsymbol{v}(t)|\mathcal{S}(t))$ in Eq. (4) at time t can be converted to the vertex attention probability $p(\tilde{\boldsymbol{v}}|\mathcal{S}(t))$ of the standard human body model as

$$p(\tilde{\boldsymbol{v}}|\mathcal{S}(t)) = p(\boldsymbol{v}(t)|\mathcal{S}(t)). \tag{7}$$

Our method uses 4D Humans [6] to estimate the SMPL pose parameters of the human body model at time t. The SMPL posture parameters are set to constant values to generate a standard human body model. The standard posture is upright, with both arms extended horizontally to the sides. The set of time points for the frames of the video is denoted by $\mathcal{T} = \{t_1,\cdots,t_T\}$ and the set of frames in the video is denoted by $\mathcal{S} = \{\mathcal{S}(t_1),\cdots,\mathcal{S}(t_T)\}$. The number of elements in $\mathcal{T}$ is equal to the number of elements in $\mathcal{S}$.

To generate the gaze distribution over the entire duration of the video, the vertex attention probability $p(\tilde{\boldsymbol{v}}|\mathcal{S}(t))$ of Eq. (7) is marginalized using the set of frames $\mathcal{S}$. Specifically, the marginalized vertex attention probability $p(\tilde{\boldsymbol{v}}|\mathcal{S})$ is calculated as

$$p(\tilde{\boldsymbol{v}}|\mathcal{S}) = \frac{1}{n(\mathcal{S})}\sum_{\mathcal{S}(t)\in\mathcal{S}} p(\tilde{\boldsymbol{v}}|\mathcal{S}(t)), \tag{8}$$

where function $n(\ )$ returns the number of elements in the set. In addition, $p(\tilde{\boldsymbol{v}}|\mathcal{S})$ satisfies the following equation:

$$\sum_{\tilde{\boldsymbol{v}}\in\tilde{\mathcal{V}}} p(\tilde{\boldsymbol{v}}|\mathcal{S}) = 1. \tag{9}$$

Finally, our method visualizes where the observers are looking at the subject's body throughout the entire duration of the video so that the analyst can understand it instantly. Specifically, the vertex attention probability $p(\tilde{\boldsymbol{v}}|\mathcal{S})$, which is marginalized over the duration of the video, is represented in a 3D heatmap and superimposed on the standard human body model. We color the surface of the standard human model such that the vertices with a higher probability of focused gaze are closer to red, and those with a lower probability of focused gaze are closer to blue.

F. Overall method

Figure 4 presents an overview of our overall method. Our method acquires the observer's gaze direction vector $\boldsymbol{g}(t,o)$ at time t in step S1 and the vertex position vector $\boldsymbol{v}(t)$ of the human body model in step S2. Next, in step S3 (Section II-D), our method calculates the observer's level

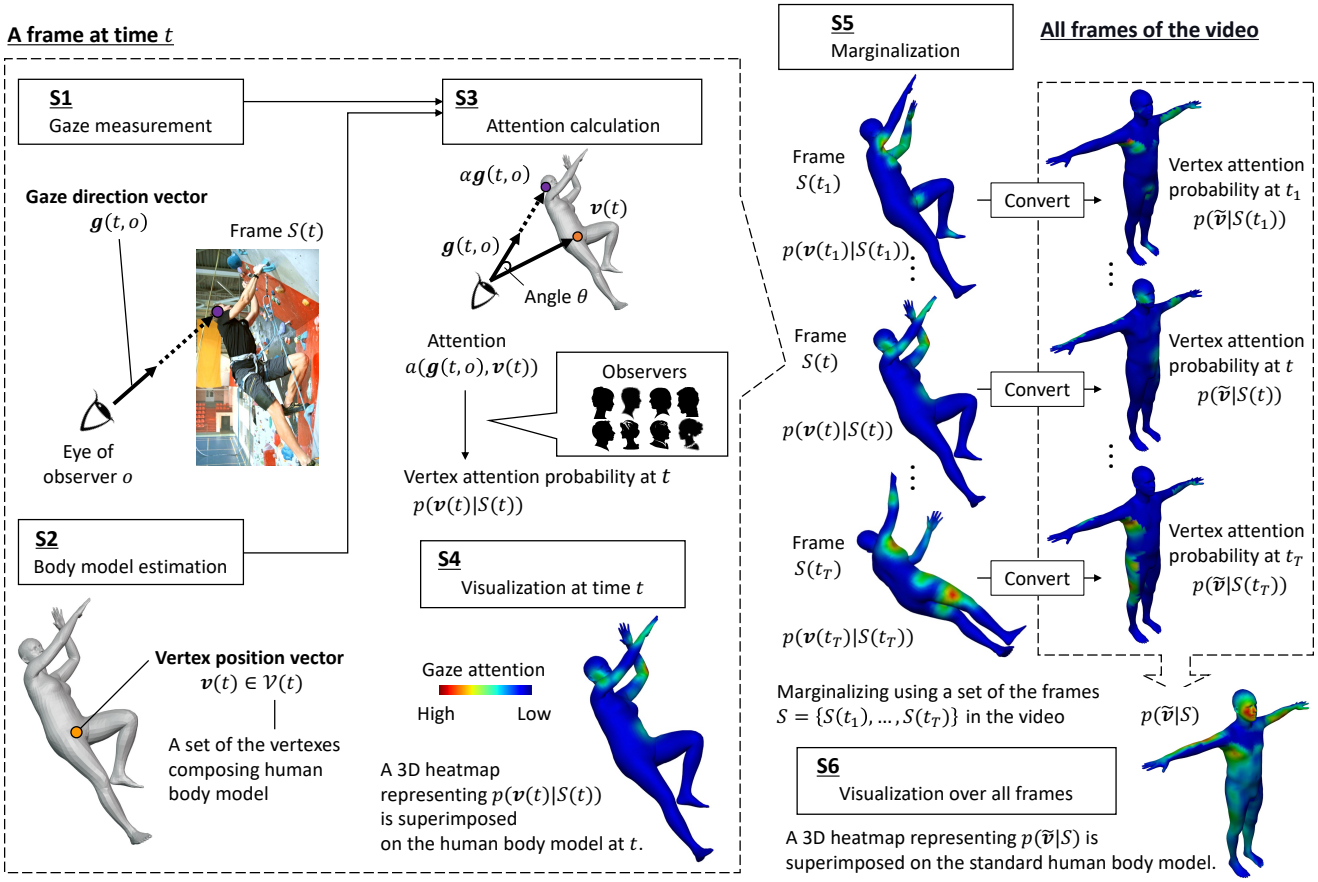

Fig. 4. Overview of our method for visualizing the gaze distribution using a 3D heatmap on the standard human body model.

of attention $a(\boldsymbol{g}(t,o),\boldsymbol{v}(t))$ based on the angle between those vectors, and we obtain the vertex attention probability $p(\boldsymbol{v}(t)|\mathcal{S}(t))$ from the level of attention. Our method visualizes the probability at time t in step S4. Finally, it marginalizes the probability over the duration of the video in step S5 (Section II-E) and visualizes the marginalized probability $p(\tilde{\boldsymbol{v}}|\mathcal{S})$ using a 3D heatmap on a standard human body model in step S6.

III. Experimental Results

A. Setup

Twenty-four observers (12 male, 12 female, mean age 23.6 $\pm$ 2.0 years) participated in the gaze measurement experiments. We used an HMD (VIVE Pro Eye, HTC) with a built-in eye tracker. We showed the observers three stimulus sports videos on a large virtual display and measured their gaze while watching those videos. We prepared stimulus sports videos with three subjects: Subject 1 climbed a wall, Subject 2 hit a ball with a tennis racket, and Subject 3 performed martial arts kicks. The durations of the stimulus videos were 10.4 s for Subject 1, 6.9 s for Subject 2, and 7.3 s for Subject 3. During the gaze measurement, we asked observers whether they were affected by the dynamism in the subject's movements in the stimulus sports videos.

B. Visualization of the gaze distribution measured from the surface and surrounding regions at each time point

We evaluated whether our method can handle the surrounding regions in addition to the surface regions when visualizing gaze distributions. The frames in Figures 5(a), (f), and (k) present the target of the visualization. The existing method [4] used the level of attention $d(\boldsymbol{v}(t),\boldsymbol{g}(t,o))$ based on the distance ϵ in Eq. (1), as described in Section II-A. By contrast, our method used the level of attention $a(\boldsymbol{g}(t,o),\boldsymbol{v}(t))$ based on the angle θ in Eq. (2) described in SectionII-D.

First, Figures 5(b), (g), and (l) show the results of the visualization using the existing method [4] based on distance. For comparison, Figures 5(c), (h), and (m) show the visualization results using our method based on angle for the surface region only. The appearance of the visualized gaze distribution is almost identical in Figures 5(b) and (c). The appearances are also almost identical in Figures 5(g) and (h) as well as in Figures 5(l) and (m). Both methods are equally expressive when visualizing gaze distributions measured on the surface regions.

Next, Figures 5(d), (i), and (n) show the results of visualization using our method in the surrounding regions, which are not considered in the existing method. For Subject 1, the observers' gaze were mainly focused on the left elbow

Fig. 5. Visualization of the gaze distribution measured on the surface region and in the surrounding region at each time point.

in the surrounding region (Figure 5(d)), whereas their gaze was mainly focused on the head and right arm in the surface region (Figure 5(c)). For Subject 2, their gaze was mainly focused on the head in the surrounding region (Figure 5(i)), whereas their gaze was mainly focused on the right arm in the surface region (Figure 5(h)). For Subject 3, their gaze was mainly focused on the head in the surrounding region (Figure 5(n)), whereas their gaze was mainly focused on the left shoulder in the surface region (Figure 5(m)). Based on these results, we believe that the gaze distribution measured in the surrounding region can detect trends overlooked when the gaze distribution is measured in the surface region alone. In the figure, the numbers at the top of the visualization results represent the percentage of gaze samples acquired in the surface region or surrounding region, respectively, when the total number of gaze samples for all 24 observers was set to 100%.

Finally, Figures 5(e), (j), and (o) show the results of the visualization using our method for the gaze distributions measured in both the surface and surrounding regions. These results qualitatively confirm that our method can visualize the gaze distribution measured from the surrounding region, which the existing method does not, in addition to the surface region at each time point. If the observer's gaze often deviates from the surface region, our method, which can simultaneously handle the surrounding region, is important for investigating gaze distributions.

C. Visualization of gaze distribution with marginalized vertex attention probability over the entire duration of the video

We evaluated whether our method can handle measured gaze distribution for sports videos by marginalizing the vertex attention probability over the entire duration of the video. Figures 6(a), (e), and (i) show some frames $\mathcal{S}(t)$ selected from the stimulus sports videos.

Figures 6(b), (f), and (j) show the results of visualizing the gaze distribution at time t using a 2D heatmap, which is a conventional visualization technique. Furthermore, Figures 6(c), (g), and (k) show the results of visualizing the vertex attention probability at time t using the 3D heatmap generated in step S4 of Figure 4, as described in Section II-F. Both the 2D and 3D heatmaps visualize where the observers were looking at the subject's body at each time point. However, when an analyst would like to understand where on the subject's body the observers were looking throughout the entire duration of the video, heatmaps for each time point require a detailed memorization of the body parts the gazes were focused on at each video frame.

Next, Figures 6(d), (h), and (l) show the results of visualizing the vertex attention probability on the standard human model using the 3D heatmap generated in step S6 of Figure 4, as described in Section II-F. In the figure, three viewpoints of the virtual camera were set and rendered when visualizing the heatmap of the standard human body model to check the 3D heatmap from multiple viewpoints simultaneously. The visualization of the 3D heatmap using our method normalizes the posture, allowing the analyst to understand at a glance where the observers were looking at the subject's body throughout the entire video.

We discuss the gaze distributions that were measured when observers were affected by the dynamism of the subject's movements throughout the entire duration of the video. For Subject 1, who was climbing a wall, the observers' gaze mainly focused on the head, arms, and right side (Figure 6(d)). For Subject 2, who was hitting a ball with a racket, the observers' gaze mainly focused on the head throughout (Figure 6(h)). For Subject 3, who was performing martial arts kicks, the observers' gaze mainly focused on the head (Figure 6(l)). These results reveal that the observers' gaze tended to focus on the head, even though the subjects' movements were different in the sports videos. We qualitatively confirmed that our method overcomes time-series changes in the subject's posture in a video by visualizing the vertex attention probability on the standard human body model.

IV. Conclusions

We proposed a method to visualize the gaze distribution of observers asked to assess whether they were affected by the dynamism of a subject's movement in a sports video. In this method, the gaze distribution measured on the subject's surface and in the surrounding regions is visualized using a

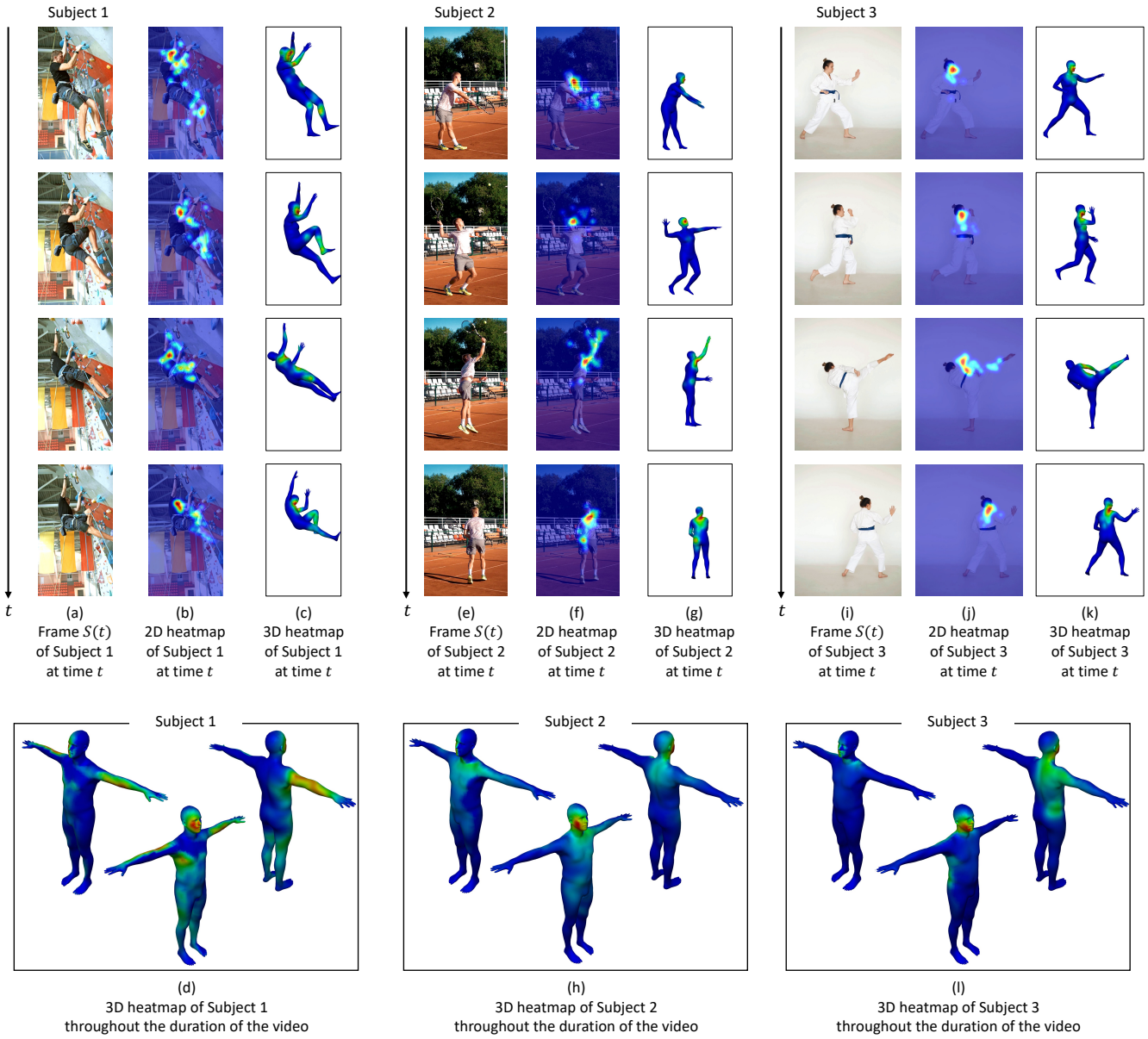

Fig. 6. Visualization of gaze distribution by marginalized vertex attention probability over the entire duration of the video.

3D heatmap in which the subject's posture is standardized over time. In future work, we intend to develop a visualization method for videos with multiple subjects. We also plan to expand our quantitative evaluations through user studies to evaluate the effectiveness of our method. This work was partially supported by JSPS KAKENHI Grant No. JP23K11145.

References

[1] C. Piers, H. Peter, V. Kiviniemi, G. Hannah, and T. Martin, "Patterns of eye movements when male and female observers judge female attractiveness, body fat and waist-to-hip ratio," *Evolution and Human Behavior*, vol. 30, no. 6, pp. 417–428, 2009.

[2] L. Nummenmaa, H. Jari, P. Santtila, and J. Hyönä, "Gender and visibility of sexual cues influence eye movements while viewing faces and bodies," *Archives of Sexual Behavior*, vol. 41, no. 6, pp. 1439–1451, 2012.

[3] K. R. Irvine, K. McCarty, T. V. Pollet, K. K. Cornelissen, M. J. Tovée, and P. L. Cornelissen, "The visual cues that drive the self-assessment of body size: Dissociation between fixation patterns and the key areas of the body for accurate judgement," *Body Image*, vol. 29, pp. 31–46, 2019.

[4] F. Inoue, M.and Iwasaki, M. Nishiyama, and I. Y., "Heatmap overlay using neutral body model for visualizing the measured gaze distributions of observers," in *Asian Conference on Pattern Recognition*, 2023, pp. 102–114.

[5] M. Matthew, N. Mahmood, J. Romero, G. Pons-Moll, and M. Black, "SMPL: A skinned multi-person linear model," *ACM Transactions on Graphics*, vol. 34, no. 6, pp. 1–16, 2015.

[6] S. Goel, G. Pavlakos, J. Rajasegaran, A. Kanazawa, and J. Malik, "Humans in 4D: Reconstructing and tracking humans with transformers," in *International Conference on Computer Vision*, 2023, pp. 1–12.