# Body-part Attention Probability for Measuring Gaze During Impression Word Evaluation

Ken KINOSHITA[1], Michiko INOUE[1],
Masashi NISHIYAMA[1 [0000−0002−5964−3209]], and Yoshio IWAI[1]

Graduate School of Engineering, Tottori University
101 Minami 4-chome, Koyama-cho, Tottori, 680-8550, Japan
nishiyama@tottori-u.ac.jp

**Abstract.** We investigate how to probabilistically describe the distribution of gaze with respect to body parts when an observer evaluates impression words for an individual in an image. In the field of cognitive science, analytical studies have reported how observers view a person in an image and form impressions about him or her. However, a probabilistic representation of their gaze distributions has not yet been discussed. Here, we represent the gaze distribution as a conditional probability according to each body part. To do this, we measured the gaze distribution of observers performing a task that consists of assessing an impression word. We then evaluated whether these distributions change with respect to the impression word and body part specified in the task. Experimental results show that the divergences between the conditional probabilities of gaze distributions are large when the impression words or body parts of the task are changed.

**Keywords:** Impression words · body parts · observers · gaze distribution · probability

## 1 Introduction

A person's impression is important in large, formal occasions, such as weddings and parties. For example, when we attend a wedding ceremony, we take care to make an impression on the attendees that is appropriate for the occasion. In this paper, we analyze the impressions made by individuals in images photographed in formal scenes. We assume that observers are seeing these individuals for the first time.

We consider several words that describe the impressions made by people photographed in formal scenes, such as "beautiful," "cute," "clean," "elegant," and "friendly." When observers are asked about these impression words while looking at an image of a person, many observers are likely to provide the same responses. For example, if we ask observers whether the person shown in Fig. 1(a) is beautiful, we assume that many observers will respond that she is beautiful and some will respond that she is not. Suppose that we ask whether the person in the image shown in Fig. 1(b) is clean. "Clean," like "beautiful," is an impression word, so we should obtain similar results.

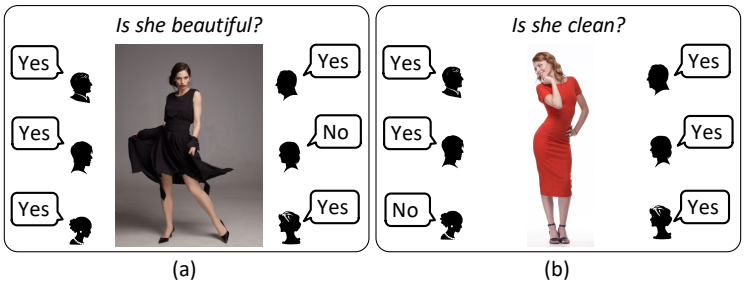


**Fig. 1.** Examples of observers evaluating whether an impression word describes a person in an image

Here, we focus on a technique that automatically predicts the impression words that describe individuals in images. To do this, machine-learning and deep-learning techniques are usually applied. In recent years, methods that incorporate the gaze distribution of image observers into machine- and deep-learning techniques have emerged [4, 8, 12, 11, 5, 10, 6]. These existing methods should improve the prediction accuracy in practice even when a large number of training samples cannot be collected. However, these existing methods do not consider how to handle the gaze distribution in terms of impression words in machine- and deep-learning techniques.

In cognitive science, several analytical studies [7, 1, 2, 9] have reported how observers view an image of a person and form impressions of him or her. In these analytical studies, observers were given the task of evaluating their impression words with respect to the image of a person, and their gaze locations were measured. The results showed that the observers spatially located their gaze on various body parts, mainly the face. The results also showed that gaze locations change depending on the impression words given in the task. However, these analytical studies did not discuss how to incorporate the gaze distribution measured from observers into practical applications that perform impression evaluation tasks using machine- and deep-learning techniques. The gaze distribution should be represented probabilistically when applying these techniques.

To achieve this, we propose a method for probabilistically representing the gaze distribution that indicates which body parts are frequently viewed when observers evaluate an impression word for a body part in an image of a person. We also investigate whether there are differences in the probability distributions of impression words by measuring the gaze locations of observers.

## 2 Measurement of the gaze locations of observers

To investigate whether the probabilistic representation of gaze distribution would be of value if incorporated into machine- and deep-learning techniques, an impression predictor should be constructed and its prediction accuracy should be

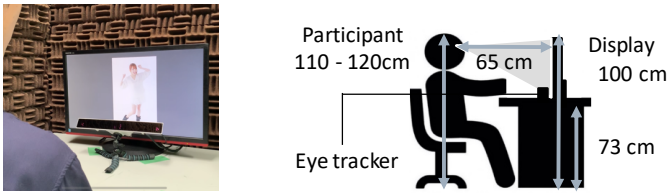


**Fig. 2.** Settings for measuring the gaze distribution.

evaluated. However, the prediction accuracy will not be improved if the gaze distribution does not change depending on the impression word. In this paper, we first investigate whether the gaze distribution varies by evaluating a probabilistic representation. For our investigation, we hypothesize the following:

– Gaze distribution changes when tasks with different impression words and different body parts are given to the participants.

In our investigation, we assigned observers six tasks ($\mathcal{T} = \{t_1, ..., t_6\}$) in which they evaluated impression words for formal scenes to measure the observers' gaze locations. The tasks are as follows:

– $t_1$: Do you feel the person's hands are beautiful?
– $t_2$: Do you feel the person's hands are clean?
– $t_3$: Do you feel the person's hands are cute?
– $t_4$: Do you feel the person's feet are beautiful?
– $t_5$: Do you feel the person's feet are clean?
– $t_6$: Do you feel the person's feet are cute?

The participants answered yes or no for each stimulus image in each task. Twenty-four participants (12 males and 12 females, average age 22.4 years, Japanese students) participated in the study.

Figure 2 shows the settings for measuring the gaze distribution of the participants. We used a measurement device (Gazepoint GP3 HD) and a 24-inch display for gaze location recording. Figure 3 shows a subset of the 96 stimulus images of people used in our experiments. In all stimulus images, the whole body is contained and the posture of the person is unrestricted. Our method represents the frequency of gaze locations in terms of each body part $b$ of the subject in the stimulus image as a conditional probability of each task $t$. Figure 4 shows the body parts $b_1, \ldots, b_{12}$ used for computing the conditional probability.

## 3 Probabilistic representation of gaze distribution

### 3.1 Pixel-attention probability for each stimulus image

We consider a probability that represents how many gaze locations are concentrated on each pixels of a stimulus image when a task $t$ is given to a participant.

**Fig. 3.** Example stimulus images $\mathcal{X}_x$ of people.


**Fig. 4.** Body parts $b_1, \ldots, b_{12}$ used for computing the body-part attention probability.

In this study, a stimulus image is represented as a set of pixels $\mathcal{X}_x$. Suppose that a gaze location is measured at location (pixel) $\boldsymbol{x}_f$ from participant $i$ when a certain frame $f$ is shown. We represent the probability as follows:

$$p(\boldsymbol{x}_j|t, i, \mathcal{X}_x, f) = \mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_f, \boldsymbol{\Sigma}), \tag{1}$$

where $\mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_f, \boldsymbol{\Sigma})$ is a bivariate normal distribution with mean vector $\boldsymbol{x}_f$ and covariance matrix $\boldsymbol{\Sigma}$. Note that this representation satisfies as follows: $\sum_{\boldsymbol{x}_j \in \mathcal{X}_x} p(\boldsymbol{x}_j|t, i, \mathcal{X}_x, f) = 1$.

When gaze is measured, gaze locations will not be observed for some frames because of eye blinks or noise. As a result, the total number of frames varies in each gaze location recording. We use $\mathcal{F}_{ti}$ to denote the set of frames in which gaze locations are measured from a participant $i$ performing task $t$. Here, we marginalize the probability using all frames $\mathcal{F}_{ti}$ as follows:

$$p(\boldsymbol{x}_j|t, i, \mathcal{X}_x) = \sum_{f \in \mathcal{F}_{ti}} p(\boldsymbol{x}_j|t, i, \mathcal{X}_x, f)p(f). \tag{2}$$

We approximate $p(f)$ using a uniform distribution $1/F_{ti}$ as follows:

$$p(\boldsymbol{x}_j|t, i, \mathcal{X}_x) = \frac{1}{F_{ti}} \sum_{f \in \mathcal{F}_{ti}} p(\boldsymbol{x}_j|t, i, \mathcal{X}_x, f), \tag{3}$$

where $F_{ti}$ is the number of frames of set $\mathcal{F}_{ti}$. Note that this probability satisfies $\sum_{\boldsymbol{x}_j \in \mathcal{X}_x} p(\boldsymbol{x}_j|t, i, \mathcal{X}_x) = 1$.

Next, we marginalize probability $p(\boldsymbol{x}_j|t, i, \mathcal{X}_x)$ according to the set of participants $\mathcal{I}_t$ performing task $t$ as follows:

$$p(\boldsymbol{x}_j|t, \mathcal{X}_x) = \sum_{i \in \mathcal{I}_t} p(\boldsymbol{x}_j|t, i, \mathcal{X}_x)p(i). \tag{4}$$

We approximate $p(i)$ by uniform distribution $1/I_t$ as follows:

$$p(\boldsymbol{x}_j|t,\mathcal{X}_x) = \frac{1}{I_t}\sum_{i\in\mathcal{I}_t} p(\boldsymbol{x}_j|t,i,\mathcal{X}_x), \tag{5}$$

where $I_t$ is the number of participants in set $\mathcal{I}_t$. Note that this equation satisfies $\sum_{\boldsymbol{x}_j\in\mathcal{X}_x} p(\boldsymbol{x}_j|t,\mathcal{X}_x) = 1$. We call $p(\boldsymbol{x}_j|t,\mathcal{X}_x)$ the pixel-attention probability.

Note that the people in the stimulus images have different postures, which means that their body regions are not aligned. Thus, pixel-attention probability can only handle a single stimulus image $\mathcal{X}_x$. In the next section, we describe a probabilistic representation that focuses on body parts to compare the probabilities among various stimulus images.

### 3.2 Body-part attention probability

We define a probability that represents how many gaze locations are concentrated on body part $b$ when participants observe stimulus image $\mathcal{X}_x$ in task $t$ as follows:

$$p(b|t,\mathcal{X}_x) = \sum_{\boldsymbol{x}_j\in\mathcal{X}_x} p(b|\boldsymbol{x}_j,t,\mathcal{X}_x)p(\boldsymbol{x}_j|t,\mathcal{X}_x). \tag{6}$$

However, it is difficult to obtain this probability directly from the outputs of gaze measurements. We assume that a smaller distance between the measured gaze location and the body-part location means the probability that the gaze is located on that body part is higher. We assume $p(b|\boldsymbol{x}_j,t,\mathcal{X}_x) \propto \mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_b,\boldsymbol{\Sigma})$ with the following:

$$p(b|t,\mathcal{X}_x) = \sum_{\boldsymbol{x}_j\in\mathcal{X}_x} \mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_b,\boldsymbol{\Sigma})p(\boldsymbol{x}_j|t,\mathcal{X}_x), \tag{7}$$

where $\boldsymbol{x}_b$ is the location for body part $b$, $\mathcal{N}(\boldsymbol{x}_j|\boldsymbol{x}_b, and\boldsymbol{\Sigma})$ is a bivariate normal distribution with mean vector $\boldsymbol{x}_b$ and covariance matrix $\boldsymbol{\Sigma}$. Note that this equation satisfies $\sum_{b\in\mathcal{B}} p(b|t,\mathcal{X}_x) = 1$, where $\mathcal{B} = \{b_1,\ldots,b_{12}\}$ is the set of body parts.

Next, we calculate the probability $p(b|t)$ that gaze is located on body part $b$ for a given task $t$. We marginalize the probability using a set of stimulus images $\mathcal{X}$ as follows:

$$p(b|t) = \sum_{\mathcal{X}_x\in\mathcal{X}} p(b|t,\mathcal{X}_x)p(\mathcal{X}_x). \tag{8}$$

We approximate $p(\mathcal{X}_x)$ using uniform distribution $1/X$ as follows:

$$p(b|t) = \frac{1}{X}\sum_{\mathcal{X}_x\in\mathcal{X}} p(b|t,\mathcal{X}_x), \tag{9}$$

where $X$ is the number of stimulus images. We call $p(b|t)$ the body-part attention probability.

**Table 1.** Body-part attention probabilities of body parts in the tasks for hands or feet.

| Body part | Body-part attention probability (%) | |
|---|---|---|
| | Hands $p_{t_{1,2,3}}$ | Feet $p_{t_{4,5,6}}$ |
| Nose | 16.73 | 12.73 |
| Right shoulder | 9.42 | 3.06 |
| Left shoulder | 10.42 | 3.16 |
| Right elbow | 8.35 | 2.55 |
| Left elbow | 9.46 | 2.24 |
| Right wrist | 17.00 | 3.98 |
| Left wrist | 16.61 | 4.66 |
| Waist | 7.93 | 13.54 |
| Right knee | 1.60 | 16.04 |
| Left knee | 1.48 | 18.90 |
| Right toes | 0.49 | 8.95 |
| Left toes | 0.51 | 10.19 |

## 4 Experimental results

We first computed the body-part attention probability for the hands with $p_{t_{1,2,3}} = (p(b|t_1) + p(b|t_2) + p(b|t_3))/3$ using tasks $t_1$, $t_2$, and $t_3$. We also computed it for the feet with $p_{t_{4,5,6}} = (p(b|t_4) + p(b|t_5) + p(b|t_6))/3$ using tasks $t_4$, $t_5$, and $t_6$. Table 1 shows the body-part attention probabilities of the body parts under these two conditions (hands or feet). For $p_{t_{1,2,3}}$, the right and left wrists, which are adjacent to the hands, have higher probabilities than all other parts of the body except for the nose. For $p_{t_{4,5,6}}$, the lower body parts (waist, knees, and toes) have higher probabilities than the upper body parts (shoulders, elbows, and wrists). We believe that when a body part is included in the task, it is more likely that the participant's gaze is drawn to the body part itself in the task. We also believe that even if the face (nose) is not explicitly included in the task, the participants frequently look at the face.

Next, we computed the body-part attention probability for "beautiful" with $p_{t_{1,4}} = (p(b|t_1) + p(b|t_4))/2$ using tasks $t_1$ and $t_4$. We computed it for "clean" with $p_{t_{2,5}} = (p(b|t_2) + p(b|t_5))/2$ using tasks $t_2$ and $t_5$. We also computed it for "cute" with $p_{t_{3,6}} = (p(b|t_3) + p(b|t_6))/2$ using tasks $t_3$ and $t_6$. Table 2 shows the probabilities with respect to the impression words in the tasks ("beautiful," "clean," or "cute"). In $p_{t_{1,4}}$, the gaze is likely to be drawn to the wrists, waist, and knees when "beautiful" is evaluated. In $p_{t_{2,5}}$, the gaze is strongly drawn to the nose when "clean" is evaluated. Additionally, in $p_{t_{3,6}}$, the gaze is drawn to the nose and wrists when "cute" is evaluated. We confirmed that there is a tendency for the frequencies of the body parts viewed by the participants to change if the impression word in the task is changed.

Next, we compared how similar the gaze distributions for different impression words were in $p_{t_{1,4}}$, $p_{t_{2,5}}$, and $p_{t_{3,6}}$. To achieve this, we computed the

**Table 2.** Body-part attention probabilities for the impression words "beautiful," "clean," and "cute."

| Body part | Body-part attention probability (%) | | |
|---|---|---|---|
| | Beautiful $p_{t_{1,4}}$ | Clean $p_{t_{2,5}}$ | Cute $p_{t_{3,6}}$ |
| Nose | 9.46 | 21.55 | 13.18 |
| Right shoulder | 4.99 | 7.74 | 5.98 |
| Left shoulder | 5.42 | 8.39 | 6.56 |
| Right elbow | 4.77 | 6.00 | 5.58 |
| Left elbow | 5.86 | 5.84 | 5.85 |
| Right wrist | 11.44 | 9.10 | 10.93 |
| Left wrist | 11.20 | 9.07 | 11.65 |
| Waist | 12.77 | 10.43 | 9.00 |
| Right knee | 11.60 | 5.96 | 8.90 |
| Left knee | 12.85 | 8.04 | 9.69 |
| Right toes | 4.45 | 3.86 | 5.85 |
| Left toes | 5.19 | 4.02 | 6.83 |

distances between the conditional probability distributions of the tasks using the Jensen–Shannon (JS) divergence [3]. The JS divergence between "beautiful" $p_{t_{1,4}}$ and "clean" $p_{t_{2,5}}$ was $D_{JS}(p_{t_{1,4}}||p_{t_{2,5}}) = 2.52$. The JS divergence between "clean" $p_{t_{2,5}}$ and "cute" $p_{t_{3,6}}$ was $D_{JS}(p_{t_{2,5}}||p_{t_{3,6}}) = 1.25$. The JS divergence between "cute" $p_{t_{3,6}}$ and "beautiful" $p_{t_{1,4}}$ was $D_{JS}(p_{t_{3,6}}||p_{t_{1,4}}) = 0.69$. To understand the meaning of these divergence values, we calculated the differences in the gaze distributions of male and female participants. We obtained $D_{JS}(p_{men}||p_{women}) = 0.43$. We found that the differences in the body-part attention probabilities with respect to impression words were larger than the differences caused by the gender of the participants. We believe that there is a tendency for the conditional probabilities of gaze on body parts to differ with respect to impression words in formal scenes. Although we have not yet reached the level of incorporating gaze distribution into practical applications, our main contribution is the development of a probabilistic representation of the gaze distribution, which is a fundamental technique that is essential for these applications.

## 5 Conclusions

We proposed a method for representing a probability that indicates which body parts are frequently viewed when observers evaluate impression words in stimulus images. The experimental results show that the combinations of impression words and body parts contained in the tasks change the values of the JS divergence between conditional probabilities, which means the observers focus their gaze on different body parts of an individual in an image for each task. In future work, we intend to develop a method for predicting impression words using

machine- and deep-learning techniques using our probabilistic gaze representation.

## Acknowledgment

This work was partially supported by JSPS KAKENHI Grant No. JP20K11864.

## References

1. Bareket, O., Shnabel, N., Abeles, D., Gervais, S., Yuval-Greenberg, S.: Evidence for an association between men's spontaneous objectifying gazing behavior and their endorsement of objectifying attitudes toward women. Sex Roles pp. 245–256 (2018)
2. Dixson, B., Grimshaw, G., Linklater, W., Dixson, A.: Eye-tracking of men's preferences for waist-to-hip ratio and breast size of women. Archives of sexual behavior **40**, 43–50 (2009)
3. Fuglede, B., Topsoe, F.: Jensen-shannon divergence and hilbert space embedding. In: Proceedings of the IEEE International Symposium onInformation Theory. p. 31 (2004)
4. Murrugarra-Llerena, N., Kovashka, A.: Learning attributes from human gaze. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision. pp. 510–519 (2017)
5. Murrugarra-Llerena, N., Kovashka, A.: Learning attributes from human gaze. In: Proceedings of IEEE Winter Conference on Applications of Computer Vision. pp. 510–519 (2017)
6. Nishiyama, M., Matsumoto, R., Yoshimura, H., Iwai, Y.: Extracting discriminative features using task-oriented gaze maps measured from observers for personal attribute classification. Pattern Recognition Letters **112**, 241–248 (2018)
7. Philippe, B., Gervais, S.J., Holland, A.M., Dodd, M.D.: When do people "check out" male bodies? appearance-focus increases the objectifying gaze toward men. Psychology of Men and Masculinity **19**(3), 484–489 (2018)
8. Qiao, T., Dong, J., Xu, D.: Exploring human-like attention supervision in visual question answering. In: Proceedings of the 32nd AAAI Conference on Artificial Intelligence. pp. 7300–7307 (2018)
9. Riemer, A.R., Haikalis, M., Franz, M.R., Dodd, M.D., Dilillo, D., Gervais, S.J.: Beauty is in the eye of the beer holder: An initial investigation of the effects of alcohol, attractiveness, warmth, and competence on the objectifying gaze in men. Sex Roles **79**, 449–463 (2018)
10. Sattar, H., Bulling, A., Fritz, M.: Predicting the category and attributes of visual search targets using deep gaze pooling. In: Proceedings of IEEE International Conference on Computer Vision Workshops. pp. 2740–2748 (2017)
11. Sugano, Y., Ozaki, Y., Kasai, H., Ogaki, K., Sato, Y.: Image preference estimation with a data-driven approach: A comparative study between gaze and image features. Journal of Eye Movement Research **7**(3) (2014)
12. Wu, J., Zhong, S., Ma, Z., Heinen, S.J., Jiang, J.: Gaze aware deep learning model for video summarization. In: Proceedings of the Pacific Rim Conference on Multimedia. pp. 285–295 (2018)