Age group identification using gaze-guided feature extraction

Michiko Inoue
Graduate School of Engineering
Tottori University, Japan.
mi.inoue@tottori-u.ac.jp

Masashi Nishiyama
Graduate School of Engineering
Tottori University, Japan.
nishiyama@tottori-u.ac.jp

Yoshio Iwai
Graduate School of Engineering
Tottori University, Japan.
iwai@tottori-u.ac.jp

***Abstract*—When an observer identifies the age group of a subject in an image, the observer's gaze focuses on a region containing an informative feature. Our aim is to improve the accuracy of age group identification by extracting features from the regions where an observer's gaze converges. Existing studies have analysed observer gaze on facial images, but not on images containing the subject's whole body. Here, we analysed which regions of the whole-body image an observer's gaze focused on while the observer performed this task. The experimental results revealed that an observer's gaze is drawn to the head of the subject regardless of the subject's age group. They also revealed that the gaze-guided feature extraction in deep learning and machine learning improved the accuracy of age group identification.**

***Index Terms*—Age group, Identification, Observer, Gaze distribution**

I. Introduction

There is a demand for a technique to identify the age groups of people from images acquired from cameras installed in public spaces, e.g., shopping centres and airports. This technique is expected to be used to collect the age groups of people present in public spaces and perform appropriate marketing for the individuals belonging to each age group.

Deep learning and machine learning are commonly used to identify the age group of an individual from images with high accuracy [1]. To further improve the accuracy of age group identification, we consider introducing the gaze distribution measured when human observers identify an age group into the feature extraction of deep learning and machine learning techniques. We assume that the image regions of interest to these observers contain informative features for age group identification, even if they are used for feature extraction in deep learning and machine learning.

Recently, it has been reported that feature extraction techniques based on gaze distribution [2], [3] are effective in several image recognition tasks. Murrugarra-Llerena et al. [2] analysed observer gaze distributions by giving observers the task of predicting the impression of a subject's facial appearance in an image. They also improved the prediction accuracy of facial impressions by introducing the gaze distribution into the feature extraction of machine learning. Nishiyama et al. [3] analysed observer gaze distributions by giving observers the task of gender recognition for images containing the whole bodies of people. They also improved the accuracy of gender recognition by introducing these gaze distributions into the feature extraction of deep learning. However, these existing studies [2], [3] did not analyse observer gaze distributions for the age group identification task. They also were not evaluated for the case where the gaze distribution was used to extract features for age group identification.

To analyse observer gaze distributions in the age group identification task, Daksha et al. [4] revealed the facial parts that observers primarily view when given the task to identify the age group from facial images. Specifically, they reported that observers look at the brow and nose of a face when identifying age. However, they did not analyse the case in which whole-body images of people were used as the stimulus. Thus, it is difficult to use the findings of Daksha et al. [4] directly for age group identification in images of people. They also did not consider using observer gaze distributions for feature extraction in age group identification.

Here, we measure and analyse observer gaze distributions by giving observers the task of identifying the age groups of individuals in images containing whole bodies. We also evaluate whether the accuracy of age group identification can be improved by introducing the gaze distributions into feature extraction in deep learning and machine learning. From the experimental results, we found that observers performing this task focused their gaze on the subject's head. In addition, we suggest that there is the possibility of improving the accuracy of age group identification using gaze-guided feature extraction.

II. Observer gaze distributions in the age group identification task

A. Age groups used in our analysis

We consider the age group identification task given to observers to measure gaze distributions. Generally, age groups are set using various age ranges for different applications. For example, the major age groups in Japanese marketing are the F1 layer (20–34 years old) and F2 layer (35–49 years old). In this paper, we use public datasets with age group labels assigned to the people in the images. We must pay attention to the fact that the age ranges for determining the age groups vary from one dataset to another. Below, we describe the age group labels for two representative datasets: the CUHK dataset, which is part of the PETA dataset [5], and the RAP

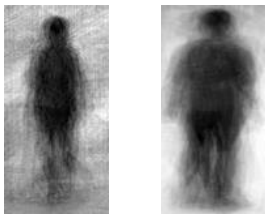
CUHK RAP

(a) A_1 (17-30 years old)

CUHK RAP

(b) A_2 (31-45 years old)

Fig. 1. Examples of the stimulus images.



(a) CUHK (b) RAP

Fig. 2. Average images of the stimulus images in the CUHK and RAP datasets.

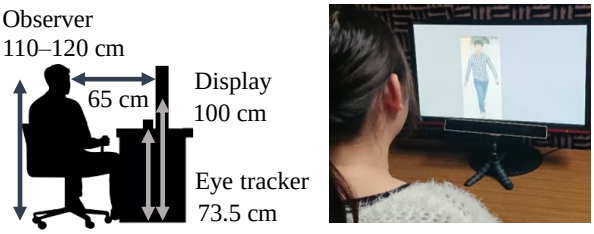


Fig. 3. Experimental setting for gaze measurement.

dataset [6]. Furthermore, we must pay attention to the fact that there is a large bias in the number of images belonging to each age group label in these datasets. When analysing the gaze distribution and evaluating identification accuracy, it is considered advisable to avoid an extremely small number of images of people in a particular age group. Therefore, for our gaze analysis, we used the following two age group labels, which have a large number of images of people belonging to each age group:

A_1: Young age group (17–30 years old),
A_2: Middle age group (31–45 years old).

In the CUHK dataset, the age group labels 'Under 30', 'Under 45', 'Under 60', and 'Over 60' were assigned. Of these, the number of 'Under 30' and 'Under 45' images was 2572 out of 2720.[1] In contrast, the number of 'Under 60' and 'Over 60' images was 148, less than 10% of the total. Thus, we assigned the 'Less than 30' images to A_1 and 'Less than 45' images to A_2. By contrast, in the RAP dataset, the age group labels 'Age under 16', 'Age 17–30', 'Age 31–45', and 'Age over 46' were assigned. Here, we used the images from the camera conditions 'CAM01' and 'CAM25'. The number of 'Age 17–30' and 'Age 31–45' images was 4670 out of 4925. In contrast, the number of 'Age under 16' and 'Age over 46' images was 255, which is less than 10% of the total. Thus, we assigned 'Age 17–30' images to A_1 and 'Age 31–45' images to A_2. Note that the age boundaries, such as exactly 30 years old, are handled differently in each dataset, but we used the above age group labels.

B. Stimulus images

We describe the stimulus images used to measure the gaze distribution in age group identification. Figure 1 shows examples of the stimulus images. To control for experimental conditions, the numbers of male and female subjects in the stimulus images were kept equal. The proportions of the subject body orientations (front, back, left, and right) in the images are the same. The same subject appeared only once in all stimulus images. The number of stimulus images was 32 for each dataset.

We assumed that the positions of the subject's body parts in the stimulus image are almost aligned. To confirm this alignment, we averaged the 32 stimulus images in each dataset. The results are shown in Figs. 2(a) and (b). In both figures, a black ellipse appears at the top of the average image. This ellipse corresponds to the head of the subjects. A large black area appears near the centre of the average image. This area corresponds to the torso of the subjects. In addition, a light black area appeared at the bottom of the average image. This area corresponds to the feet of the subjects.

C. Experimental conditions for gaze measurement

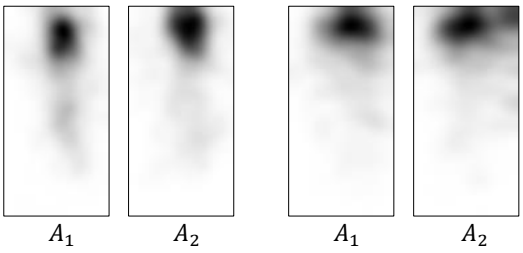
Sixteen observers (10 men and 6 women, 22.4±1.0 years old) participated. Figure 3 shows the experimental setting for gaze measurement. The size of the display was 24 inches (resolution: 1920×1080 pixels). An eye tracker (Gaze Point GP3) was used to measure the gaze distribution. The sampling rate of the eye tracker was 60 Hz. The stimulus image was enlarged to 960 pixels in height while keeping the aspect ratio fixed. We showed the enlarged stimulus image to the observer on the display.

In the age group identification task given to the observers, we asked the following question:

Q : Which age group label, A_1 or A_2, do you identify the subject in the image as belonging to?

The procedure for asking this question to the observers was as follows.

[1] We did not allow the same subject to appear in the images. Attribute labels in the RAP v1 dataset were used to identify images that could be of the same subject.



(a) CUHK (b) RAP

Fig. 4. Gaze distribution measured using stimulus images for each dataset in the task of identifying age group labels A_1 and A_2.

- P_1: We selected one observer randomly and asked question Q.
- P_2: The grey image was displayed for 2 seconds.
- P_3: We selected one stimulus image randomly and displayed that for 2 seconds.
- P_4: The observer verbally responded with the subject's age group label while a black image was displayed for 3 seconds.
- P_5: Procedures P_2 to P_4 were repeated for each stimulus image.
- P_6: Procedures P_1 to P_5 were repeated until all observers were finished.

D. Generation of gaze distribution

We describe the procedure for generating the gaze distribution. We used the observer's gaze positions measured in procedure P_3, and the method described in [3] was performed. Specifically, the gaze positions measured for a period of 2 seconds were accumulated in the time direction at the corresponding positions in the gaze distribution image. When accumulating these points, we used only the gaze positions that the eye tracker outputted as fixations. Note that an observer not only looks at the gaze position on the stimulus image at each time point, but also the region surrounding this position. Thus, we used a Gaussian kernel to filter the gaze distribution image during the accumulation. In addition, for all observers and all stimulus images, the gaze positions were accumulated at the corresponding positions in the gaze distribution image. This gaze distribution image was generated for each age group in each dataset.

E. Results of gaze distribution analysis

Figure 4(a) shows the gaze distribution measured during the task of identifying age group labels A_1 and A_2 using the stimulus images from the CUHK dataset. In the gaze distribution, the black areas indicates the points at which the observer gaze positions gathered. Moreover, white areas indicate areas where observer gaze positions were not gathered. To investigate which body part is the primary focus of the observers, we compared the gaze distribution in Fig. 4(a) with the body silhouette in the average image in Fig. 2(a). For both age group labels A_1 and A_2, we found that the observers focused on the subjects' heads in the stimulus images. By contrast, the observers did not focus near the subjects' feet. Next, Fig. 4(b) shows the gaze distribution measured during the task of identifying A_1 and A_2 using stimulus images from the RAP dataset. The gaze distribution in Fig. 4(b) is compared with the body silhouette in the average image in Fig. 2(b). We again found that the observers focused on the subjects' heads and not near the subjects' feet. The results in Fig. 4(b) for the RAP dataset show the same trend as the results in Fig. 4(a) for the CUHK dataset. The above results show that observers, when identifying age group labels, focus on the subject's head for both age group labels A_1 and A_2.

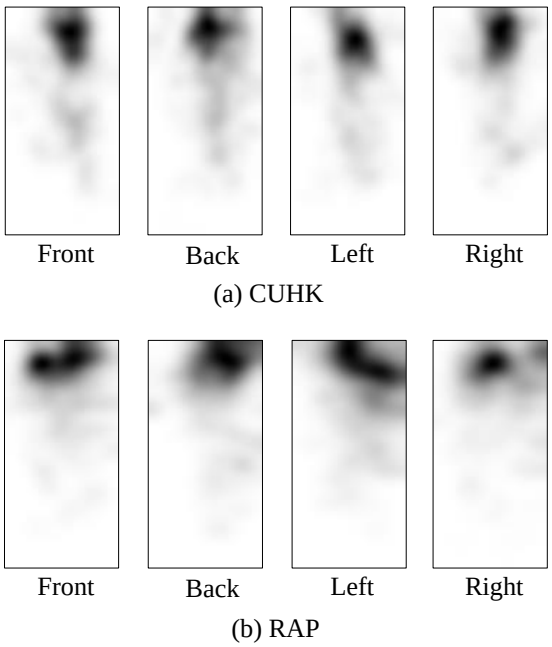


Fig. 5. Gaze distribution generated for each subject's body orientation.

We quantitatively evaluated how similar the gaze distribution for age group label A_1 was to the gaze distribution for A_2 for each dataset. Specifically, Pearson's correlation coefficients were calculated for all gaze distributions. The results show that the correlation coefficient between the A_1 and A_2 gaze distributions was 0.91 for the CUHK dataset and 0.92 for the RAP dataset. Thus, we confirmed that the gaze distribution is very similar regardless of the subject's age group.

Furthermore, we investigated whether the gaze distribution is affected by the subject's body orientation in the stimulus images. Figure 5 shows the gaze distribution generated for each subject's body orientation (front, back, left, or right). We show the gaze distributions on the CUHK dataset in Fig. 5(a) and those on the RAP dataset in Fig. 5(b). Pearson's correlation coefficient was calculated for the gaze distributions of all subject body orientations. The average correlation coefficient of ${}_4\mathrm{C}_2 = 6$ orientation combinations in the CUHK dataset was 0.92±0.03, and that in the RAP dataset was 0.84±0.03. We found that the orientation of the subject's body produces very similar gaze distributions when the observers viewed the stimulus images in the CUHK dataset. We also found that it

produces reasonably similar gaze distributions when observers viewed the stimulus images in the RAP dataset.

III. Age group identification using the gaze distribution

A. Gaze-guided feature extraction

We used a simple weighting technique to extract features from the training and test images for age group identification. Specifically, the method [3] was performed for gaze-guided feature extraction. In advance, we generated a single gaze distribution for each dataset by accumulating the gaze distributions generated for age group labels A_1 and A_2. We assigned large weights to the regions of accumulated gaze distributions where the observers frequently fixated and small weights to those regions they rarely viewed. The pixel values of the training and test images were multiplied by the gaze distribution weights pixel by pixel. Then, we examined the training and test images weighted by the gaze distribution using deep learning and machine learning techniques.

B. Accuracy of age group identification

To evaluate the accuracy of age group identification, we set the following comparison conditions:

M_1: Without gaze-guided feature extraction,
M_2: With gaze-guided feature extraction.

We used the CUHK and RAP datasets described in Section II and performed 10-fold cross-validation. For the CUHK dataset, each validation had 2286 training images and 254 test images. For the RAP dataset, each validation had 3317 training images and 1321 test images. Note that the training images were completely separate from the test images.

We used representative deep learning and machine learning techniques, a convolutional neural network (CNN), a gradient boosting decision tree (GB), a kernel support vector machine (SVM), and logistic regression (LR), for age group identification. In the CNN, we used a shallow network model [7], which consists of two convolution layers and two pooling layers. In the LR, we set the inverse of regularization strength $C = 1$. In the GB, we set the number of decision trees to 1000 and the maximum depth of nodes to 3. In the SVM, we used the radial basis function kernel and the regularization parameter $C = 1$.

Table I shows the accuracy of age group identification with and without gaze-guided feature extraction. We see that the accuracy of M_2 is better than that of M_1 for the deep learning and machine learning techniques on the CUHK dataset. We also see the same trend in the RAP dataset. Thus, we believe that the gaze-guided feature extraction improves the accuracy of age group identification.

TABLE I
Accuracy of age group identification with and without gaze-guided feature extraction.

Dataset	Technique	M_1: w/o gaze	M_2: w/ gaze
CUHK	CNN	0.57	**0.59**
	GB	0.58	**0.60**
	SVM	0.60	**0.63**
	LR	0.57	**0.59**
RAP	CNN	0.57	**0.61**
	GB	0.63	**0.64**
	SVM	0.63	**0.65**
	LR	0.62	**0.63**

IV. Conclusions

We analysed observer gaze distributions, measured when they perform the task of identifying the age group of individuals in images containing whole bodies. In addition, we evaluated whether the use of the gaze distribution improves the accuracy of age group identification. To measure the gaze distribution, we designed a task in which observers viewed full-body images of people as stimuli and were asked to identify the age group labels. Our gaze analysis revealed that observers identifying the subject's age group primarily viewed the subject's head. We suggested that there is the possibility of improving the accuracy of age group identification by using the gaze distribution for feature extraction. In future work, we will expand our research to investigate more detailed age group labels and analyse the gaze distribution using high resolution datasets.

Acknowledgment

This work was partially supported by JSPS KAKENHI Grant No. JP23K11145.

References

[1] X. Wang, S. Zheng, R. Yang, A. Zheng, Z. Chen, J. Tang, and B. Luo, "Pedestrian attribute recognition: A survey," *Pattern Recognition*, vol. 121, pp. 108220, 2022.

[2] N. Murrugarra-Llerena and A. Kovashka, "Learning attributes from human gaze," *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pp. 510–519, 2017.

[3] M. Nishiyama, R. Matsumoto, H. Yoshimura, and Y. Iwai, "Extracting discriminative features using task-oriented gaze maps measured from observers for personal attribute classification," *Pattern Recognition Letters*, vol. 112, pp. 241–248, 2018.

[4] D. Yadav, N. Kohli, E. Kalsi, M. Vatsa, R. Singh, and A. Noore, "Unraveling human perception of facial aging using eye gaze," in *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2018, pp. 2221–22217.

[5] Y. Deng, P. Luo, C. C. Loy, and X. Tang, "Pedestrian attribute recognition at far distance," *ACM*, pp. 789–792, 2014.

[6] D. Li, Z. Zhang, X. Chen, H. Ling, and K. Huang, "A richly annotated dataset for pedestrian attribute recognition," *CoRR, abs/1603.07054*, 2016.

[7] A. Grigory, B. Sid-Ahmed, R. Natacha, and D. Jean-Luc, "Learned vs. hand-crafted features for pedestrian gender recognition," in *Proceedings of 23rd ACM International Conference on Multimedia*, 2015, pp. 1263–1266.