Gender Classification using the Gaze Distributions of Observers on Privacy-Protected Training Images

Michiko Inoue, Masashi Nishiyama and Yoshio Iwai
Graduate School of Engineering, Tottori University, Tottori, Japan
nishiyama@tottori-u.ac.jp

Keywords: Gender classification, training images, gaze distributions, privacy-protection

Abstract: We propose a method for classifying the gender of pedestrians using a classifier trained by images containing privacy-protection of the head region. Recently, manipulated training images containing pedestrians have been required to protect the privacy of pedestrians. In particular, the head regions of the training images are manipulated. However, the accuracy of gender classification decreases when privacy-protected training images are directly used. To overcome this issue, we aim to use the human visual ability to correctly discriminate males from females even though the head regions have been manipulated. We measure the gaze distributions of observers who view pedestrian images and use them to pre-process gender classifiers. The experimental results show that our method using gaze distribution improved the accuracy of gender classification when the head regions of the training images have been manipulated with masking, pixelization, and blur for privacy-protection.

1 INTRODUCTION

Video surveillance cameras are now installed in various public spaces such as airports, stations, and shopping malls. Gender classification using pedestrian images acquired from video surveillance cameras is becoming widespread. The gender classification of pedestrian images enables the distribution of pedestrian gender to be collected for the development of various applications such as crime prevention and product marketing. The existing methods (Sudowe et al., 2015; Schumann and Stiefelhagen, 2017) achieve high accuracy using deep learning techniques for pedestrian attribute classification. However, in the existing methods, the collection of a large number of training images is required to achieve good deep learning performance.

When collecting training images, we must carefully handle the privacy of the human subjects in the images. Training images include personal information that makes human subjects feel nervous about allowing their images to be acquired. In particular, we must pay attention to how their faces are handled. Furthermore, these training images will be used repeatedly over a long period of time to improve the accuracy of gender classification. Therefore, techniques for protecting the privacy of the subjects in the images are required when collecting training images for gender classification.

To protect the privacy of subjects in pedestrian images, the head regions are generally manipulated. For example, head regions in magazine advertisements are masked, head regions in television interviews are pixelized, and head regions in Google Street View are blurred. Furthermore, methods to perform more complex manipulations for privacy protection have been proposed (Joon et al., 2016; Zhang et al., 2014; Yamada et al., 2013; Ribaric et al., 2016; Oh et al., 2017). However, existing methods do not fully discuss the privacy protection of training images in gender classification. When they are directly applied to training images, we believe that the accuracy of gender classification will be substantially decreased. In fact, (Ruchaud et al., 2015) showed that this decrease happens when the head regions of training images are manipulated with masks, pixelization, and blur to protect the privacy of the subjects.

In this paper, we focus on the human visual ability to distinguish gender to improve the accuracy of gender classification using privacy-protected training images. The existing method (Nishiyama et al., 2018) revealed that observers look at important features when distinguishing the gender of pedestrians in images. In that approach, gaze distributions were measured when observers viewed the pedestrian images. The existing method assigned weights to the

Figure 1: Observers can correctly determine gender when the head regions of pedestrian images are completely masked to protect the privacy of the subjects.

training images as preprocessing before feature extraction so that a classifier could be trained to emphasize the head region, which is where the gaze of the observers gathered. However, when the existing method is directly applied to privacy-protected training images, the accuracy of gender classification unfortunately decreases. The reason for this issue is that the head regions have been manipulated to protect the privacy of the pedestrians in the training images, even though head regions increase the accuracy of gender classification. We need to consider how to overcome this problem caused by privacy-protected training images. Here, we similarly focus on human visual abilities. When the head regions of pedestrian images are completely masked, observers can still correctly distinguish gender in many cases. We believe that the observers look at other important features obtained from regions other than the head, such as the shape and appearance of the torso, in the pedestrian image. As illustrated in Figure 1, observers can correctly discriminate gender; for example, it is a woman when the torso of the subject is rounded, and it is a man when the torso of the subject is muscular.

In this paper, we propose a method to improve the accuracy of gender classification when the head regions in the training images are manipulated for privacy protection. We make the following two contributions:

- We reveal the important regions where the gaze of observers tends to gather in privacy-protected images. We measure the gaze distributions of the observers when the head regions of the subjects are masked.
- We confirm whether or not the accuracy of gender classification is improved using these gaze distributions. We use the gaze distributions in feature extraction for gender classifiers.

To compare the important regions of the masked images, we also perform experiments in which the head regions are not masked.

2 Related work

2.1 Privacy protection

As described in (Joon et al., 2016; Flammini et al., 2013; Campisi, 2013), the privacy protection of subjects contained in images is an important issue, and has attracted attention in recent years. Various privacy-protection methods have been proposed for all stages in the process of recognition, such as during pedestrian image acquisition or when classification is performed using the pedestrian image. To prevent recognition in conventional face detection algorithms, a method that requires the user to wear protective glasses was proposed in (Yamada et al., 2013). To automatically shield the heads of subjects in images, a method that incorporates a special mechanism embedded in the camera system was proposed in (Zhang et al., 2014). The existing methods have the advantage that privacy-protected images are only recorded in video surveillance systems. However, it is necessary to prepare special equipment for image acquisition. Existing methods are not suitable for collecting training images. To protect the privacy of faces, a method that replaces the face of a subject with the face of a fictitious person has been proposed in (Ribaric et al., 2016). Although this method opens up new approaches to privacy-protection, it is not well accepted by the public. To hide personal features so that they are not visually discernible in an image for a classification task, a method that embeds the original feature in the image of a fictional person was proposed in (Oh et al., 2017). This is effective if the algorithm of the classification task is not changed in the future. However, the embedded features are not restored when a newly developed algorithm is applied. Instead, we consider a pre-processing method for feature extraction that can enable various classification algorithms to use privacy-protected training images.

2.2 Human visual abilities

In the fields of computer vision and pattern recognition, the use of human visual abilities has widely progressed. The estimation accuracy of a saliency map was improved using the distribution of gaze locations in (Xu et al., 2015). Action recognition and gaze-attention estimation were performed simultaneously using a wearable camera in (Fathi et al., 2012). Preference estimation was performed using gaze locations and image features in (Sugano et al., 2014), an eye-movement pattern for product recommendation was estimated in (Zhao et al., 2016), and an object recognition task was performed using only gaze

(a) S_1

(b) S_2

Figure 2: Examples of stimulus images for measuring gaze distributions.

distributions in (Karessli et al., 2017). The attributes of face images were classified using gaze-attention regions in (Murrugarra-Llerena and Kovashka, 2017), and the attribute classification of fashion clothes images was performed using a deep learning technique with gaze distributions in (Sattar et al., 2017). Although the existing methods handle various applications using gaze, they do not address the application of privacy protection. Thus, privacy protection in gender classification is a new application for gaze distributions.

3 Gaze distribution of observers when the head regions are masked

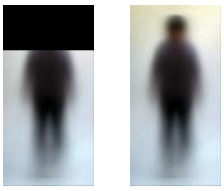
3.1 Stimulus images

We investigated which regions gather the gaze of observers when they determine the gender of subjects in images. Sixteen participants (10 males and 6 females, average age 22.4 ± 1.0 years) participated in the study. The head regions and their surrounding regions were completely masked so that the participants could not observe the head region of the subject in the images. Note that we refer to the pedestrian images observed by the participants as the stimulus images.

We used the following two conditions for the pedestrian images in the experiments:

S_1 : the head region was masked;
S_2 : the head region was *not* masked.

We measured S_2 in addition to S_1 for comparison. Figures 2(a) and (b) show the examples of the stimulus images of S_1 and S_2. We used the CUHK dataset,



(a) S_1 (b) S_2

Figure 3: Average images of pedestrian images with and without masked head regions.

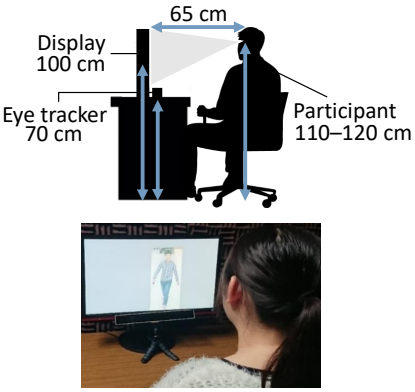


Figure 4: Experimental setting for measuring gaze distributions.

which is included in the PETA dataset (Yubin et al., 2014) as the stimulus images. The size of all stimulus images was 80×160 pixels for both S_1 and S_2.

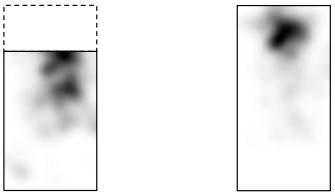
The pedestrian regions in the images of the CUHK dataset were manually aligned. We checked the alignment using the average image computed from all pedestrian images in the CUHK dataset. Figures 3(a) and (b) show the average images of S_1 and S_2. In Figure 3(b) , the black circle that appears at the top corresponds to the head region. The black ellipse that appears near the center of the image corresponds to the torso region. The light gray part that appears at the bottom of the image corresponds to the foot region. In Figure 3(a), we see that the head region is completely masked because the black circle at the top in Figure 3(b) is not observed.

To control the experimental conditions, the number of male and female subjects included in the stimulus images was equal. The proportions of all body orientations of the subjects in the stimulus images (front, back, left and right) were equal. In addition, the same person did not appear more than once in the stimulus images. Finally, the number of the stimulus images was 32 in S_1 and 32 in S_2.

Figure 4 shows the experimental setting for measuring gaze distributions. The participant was seated at a position 65 cm from the display. Each partici-

(a) S_1 (b) S_2

Figure 5: Examples of S_1 and S_2 stimulus images presented on the display.



(a) g_{S_1} (b) g_{S_2}

Figure 6: Gaze maps of S_1 and S_2.

pant adjusted the chair height while keeping the eye height between 110 and 120 cm. The display size was 24 inches ($1,920 \times 1,080$ pixels). We used the GP3 gaze measurement device (gazepoint), which has a sampling rate of 60 fps. The specifications of the device state that its angular resolution is between 0.5 and 1.0 degrees. We enlarged the stimulus image to 480×960 pixels. To avoid center bias, we presented the stimulus images at random positions on the display. Figures 5(a) and (b) show examples of the S_1 and S_2 stimulus images on the display. The pixel values of the masked regions in the stimulus images were set to be the same as those of the display background. We presented the stimulus images of S_1 for 8 participants (4 males and 4 females) randomly selected from all the participants and the stimulus images of S_2 for the remaining 8 participants (6 males and 2 females).

3.2 Protocol

The procedure for measuring the gaze distribution of the participants was as follows.

- P_1: We randomly selected a participant.
- P_2: We set the condition for measuring the gaze distribution to either S_1 or S_2.
- P_3: We explained how to perform the task of determining the gender of the subject using an example image.
- P_4: We presented a gray image for 2 s.
- P_5: We presented a randomly selected stimulus image for 2 s.
- P_6: We presented a black image for 3 s and asked the participant to state the gender of the subject in the stimulus image.
- P_7: We repeated steps P_4 to P_6 until all the stimulus images were presented.
- P_8: We repeated steps P_1 to P_7 until all the participants completed the experiment.

Here, we explain our method for generating a gaze map. We integrated the gaze distributions measured in P_5 using the existing method (Nishiyama et al., 2018) for a single gaze map. We used only the gaze locations of the participants, which the gaze measurement device output as fixations. We summed the locations of the fixations for S_1 or S_2 from all the participants and all stimulus images. The size of the gaze map was resized to the size of the stimulus image.

3.3 Analysis of the measured gaze distributions

Figure 6 shows the gaze maps of S_1 and S_2. We show the gaze map g_i for each $i \in \{S_1, S_2\}$. In the map, darker regions of the gaze map indicate that the gaze of the participants were more frequently measured at these locations than in the lighter regions. Comparing the gaze map g_{S_1} of Figure 6(a) with the average image of Figure 3(a), we see that the gaze locations of the participants gathered in the torso regions of the subjects when the head regions were masked. The body shape, clothes, and bag were observed in the subjects of the stimulus images. Next, we compared g_{S_2} of Figure 6(b) with the average image of Figure 3(b). We see that the participants mainly viewed the head regions of the subjects in the stimulus images when they were not masked. This tendency is the same as that of the existing study (Nishiyama et al., 2018). Note that the gaze locations of the participants did not gather near the feet regions of the subjects in both S_1 and S_2 stimulus images.

We describe the accuracy of gender classification performed by the participants. We counted correct answers when the responses of the participants matched the gender labels of the stimulus image. The accuracy of the participants was $87.0 \pm 5.5\%$ for S_1 and $95.9 \pm 3.1\%$ for S_2.

We investigated the differences in the gaze maps with respect to the participants' gender. We also investigated the differences in the stimulus images with respect to the subjects' gender. Figure 7(a) shows the gaze maps generated by each gender of the participants, and Figure 7(b) shows the gaze maps generated for each gender of the subjects. In the maps in Figure 7(a), there is no significant difference between the genders of the participants in terms of the person re-

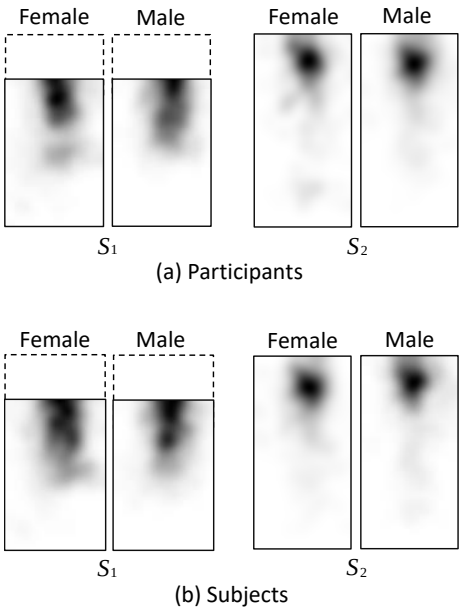


Figure 7: Gaze maps measured of (a) male and female participants and (b) male and female subjects.

Table 1: Correlation coefficients of the gaze maps for males and females with respect to participants or subjects.

Participants		Subjects	
S_1	S_2	S_1	S_2
0.88	0.97	0.97	0.92

gion gaze locations. In Figure 7(b), there was also no significant difference between the gender of the subjects in the maps. For both S_1 and S_2, we computed the correlation coefficients between the gaze maps with respect to participant gender and subject gender. We used Pearson's product–moment correlation coefficient. Table 1 shows the correlation coefficients of the gender differences in S_1 and S_2. We confirmed that the correlation coefficients between genders were very high under all combinations of conditions.

4 Gender classification using the gaze distributions for privacy-protected training images

4.1 Experimental conditions

We investigated whether or not the accuracy of gender classification was improved using the gaze distributions for the training images in which the head regions of the subjects were masked. We used images from the CUHK dataset as training and test images.

Figure 8: Examples of training images with and without masked head regions.

The stimulus images for gaze measurement were not included in the training and test images. To evaluate the accuracy of gender classification, we used 10-fold cross-validation for a total of 2,540 images of the CUHK dataset. In each cross-validation, the number of training images was 2,286 (1,143 male images and 1,143 female images), and the number of test images was 254 (127 male images and 127 female images). We define the training image conditions as follows:

- T_{S_1}: the head regions of the training images were masked;
- T_{S_2}: the head regions of the training images were *not* masked.

Figure 8(a) shows examples of the training images of T_{S_1} and Figure 8(b) shows those of T_{S_2}.

To confirm the effectiveness of the gaze map for privacy-protected training images, we compared the accuracy of gender classification under the following conditions:

G1 : our method using the masked gaze map g_{S_1};
G2 : the existing method using the *un-masked* gaze map g_{S_2};
G3 : a baseline method without the use of the gaze map.

The number of combinations of conditions of G1, G2, and G3 for T_{S_1} and T_{S_2} was six. We avoided gaze measurement for each training image and each test image and used only the gaze maps measured from the stimulus images in Section 3. We used the classifier with gaze-map based pre-processing described in (Nishiyama et al., 2018). The pre-processing is briefly explained below. Large weights are given to the pixels of the training images at which the gaze of the participants gathered. In contrast, small weights

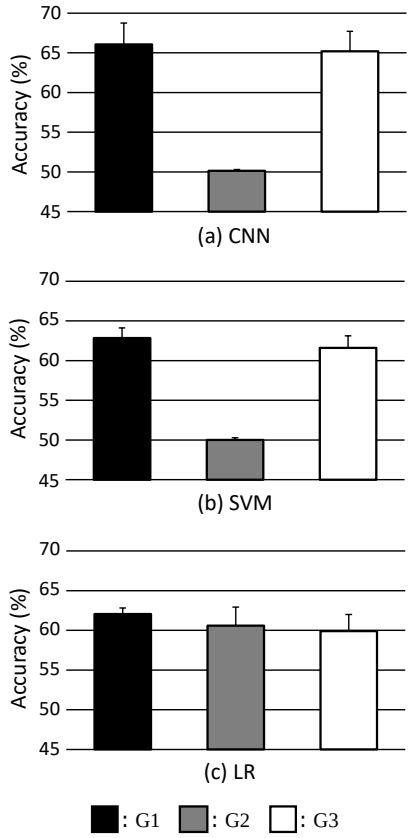


Figure 9: Accuracy of gender classification using the masked training images of T_{S_1}.

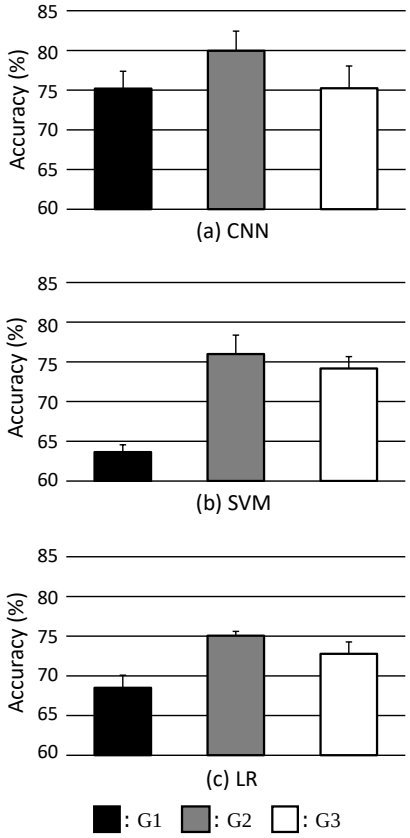


Figure 10: Accuracy of gender classification using the *unmasked* training images of T_{S_2}.

are given to the pixels at which gaze was not gathered. Specifically, let the pixel value of the gaze map be $g_i(x,y)$, where (x,y) is the location of the image. The range of $g_i(x,y)$ is $[0,1]$. In addition, the pixel value of the training image is represented as $t(x,y)$. The pixel value $t'(x,y)$ after pre-processing is computed as follows:

$$t'(x,y) = c(g_i(x,y))t(x,y), \quad (1)$$

where $c(\ )$ is a correction function. We set $c(z) = z^a + b$. When $a > 1$, the weight of the gaze map is emphasized. When $a < 1$, the weight of the gaze map is low. Variable b is an offset. We converted the RGB color space to the HSV color space and weighted only the V values using the correction function.

After assigning the weights to the training images, we employed the following classifiers:

- **CNN**: convolutional neural network. A mini-CNN (Grigory et al., 2015) with two convolution layers and two pooling layers was used.
- **SVM**: linear support vector machine (Corinna and Vladimir, 1995). The penalty parameter was set as 1.
- **LR**: logistic regression classifier (Cox, 1958). The normalization parameter was set as 1.

We did not mask the head regions of the test images. We weighted the test images using the gaze map using the same procedure as that used for the training images.

4.2 Performance of gender classification using the masked training images

Figure 9 shows the accuracy of gender classification for the masked training images of T_{S_1}. In Figure 9(a), we see that the accuracy of G1 was slightly better than that of G3, and the accuracy of G2 was significantly worse than those of G1 and G3. In Figure 9(b), we see the same tendencies shown in Figure 9(a). In Figure 9(c), we see that the accuracy of G1 is superior to those of G2 and G3, and the accuracy of G2 is slightly better than that of G3. These results confirm that gaze map g_{S_1} is more effective than g_{S_2} when the head regions of the training images are masked.

Figure 10 shows the accuracy of gender classification for the *unmasked* training images of T_{S_2}. In Fig-

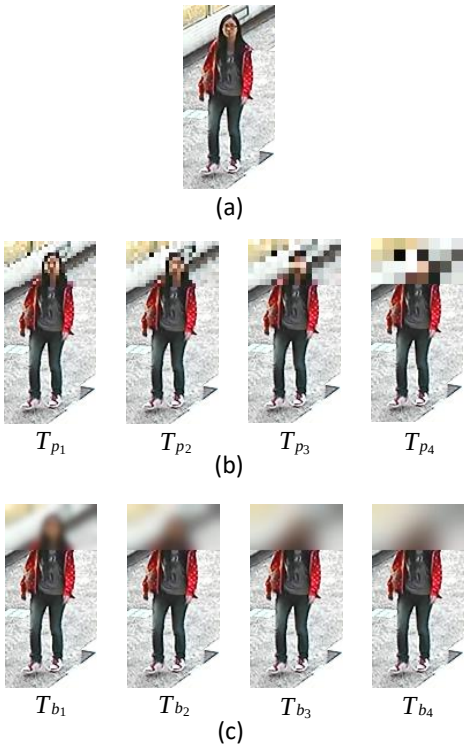


Figure 11: Examples of the training images after manipulating the head regions using pixelization and blur while changing their parameters.

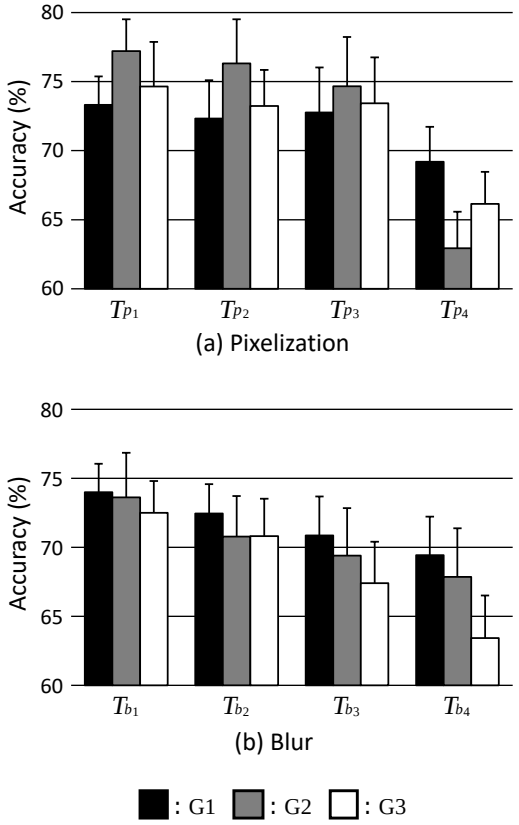


Figure 12: Accuracy of gender classification using the training images manipulated with pixelization and blur.

ure 10(a), there is no significant difference in accuracy between G1 and G3. We see that the accuracy of G2 is better than those of G1 and G3. In Figure 10(b), we see that the accuracy of G1 is worse than those of G2 and G3 and the accuracy of G2 is better than that of G3. In Figure 10(c), we see the same tendencies shown in Figure 10(b). These results confirm that gaze map g_{S_2} is more effective than g_{S_1} when the head regions of the training images are not masked. We believe that it is necessary to switch between the gaze maps suitable for each condition because the effectiveness of the gaze distributions depends on whether or not the training images are masked.

4.3 Gender classification using the training images with pixelization and blur

We investigated the accuracy of gender classification when the classifiers are trained by images modified with pixelization and blur for privacy protection. Figure 11(a) shows an example of a training image without privacy protection, Figure 11(b) shows images with pixelization, and Figure 11(c) shows images with blur. We used four manipulation levels $k \in \{1,2,3,4\}$ for pixelization T_{p_k} and blur T_{b_k}. We varied the level of pixelization (using 16×32, 12×24, 8×16, and 4×8 blocks) and the level of Gaussian blur ($\sigma = 3,7,11,15$). We used a CNN classifier and the conditions G1, G2, and G3 for the gaze maps described in Section 4.1.

Figure 12(a) shows the accuracy of gender classification using the training images with pixelization. In T_{p_1}, T_{p_2}, and T_{p_3}, G2, G3, and G1 have the highest, middle, and lowest accuracy, respectively. In contrast, in T_{p_4}, this order is G1, G3, and G2. These results hence confirm that our method based on gaze map g_{S_1} improves the accuracy for training images with a high level of pixelization.

Figure 12(b) shows the accuracy of gender classification using training images with blur. In T_{p_1}, T_{p_2}, T_{p_3}, and T_{p_4}, G1, G2, and G3 have the highest, middle, and lowest accuracy, respectively. We hence confirm that our method based on gaze map g_{S_1} improves the accuracy for training images with various levels of blur.

5 Conclusions

We proposed a method for improving the accuracy of gender classification using the gaze distribution of human observers on training images in which the privacy of the subjects was protected. We used stimulus images with masked head regions and measured the gaze distributions of observers. We confirmed that the participants mainly observed the torso regions of the subjects in the stimulus images. Next, we conducted gender classification experiments using privacy-protected training images with masking, pixelization, and blur. The experimental results confirm that our method, which uses the gaze map with masked head regions, improved the accuracy of gender classification. In future work, we intend to continue developing the method to increase the accuracy by combining gaze maps with and without masking. We will expand this investigation into gaze maps with privacy protection for various classification tasks related to attributes other than gender. This work was partially supported by JSPS KAKENHI under grant number JP17K00238 and MIC SCOPE under grant number 172308003.

REFERENCES

Campisi, P. (2013). Security and privacy in biometrics: Towards a holistic approach. *Security and Privacy in Biometrics, Springer*, pages 1–23.

Corinna, C. and Vladimir, V. (1995). Support-vector networks. *Machine Learning*, 20(3):273–297.

Cox, D. R. (1958). The regression analysis of binary sequences. *Journal of the Royal Statistical Society. Series B (Methodological)*, 20(2):215–242.

Fathi, A., Li, Y., and Rehg, J. (2012). Learning to recognize daily actions using gaze. In *Proceedings of 12th European Conference on Computer Vision*, pages 314–327.

Flammini, F., Setola, R., and Franceschetti, G. (2013). Effective surveillance for homeland security: Balancing technology and social issues. *CRC Press*.

Grigory, A., Sid-Ahmed, B., Natacha, R., and Jean-Luc, D. (2015). Learned vs. hand-crafted features for pedestrian gender recognition. In *Proceedings of 23rd ACM International Conference on Multimedia*, pages 1263–1266.

Joon, O. S., Rodrigo, B., Mario, F., and Bernt, S. (2016). Faceless person recognition: Privacy implications in social media. In *Proceedings of European Conference on Computer Vision*, pages 19–35.

Karessli, N., Akata, Z., Schiele, B., and Bulling, A. (2017). Gaze embeddings for zero-shot image classification. In *Proceedings of IEEE conference on computer vision and pattern recognition*, pages 4525–4534.

Murrugarra-Llerena, N. and Kovashka, A. (2017). Learning attributes from human gaze. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pages 510–519.

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

Oh, S. J., Fritz, M., and Schiele, B. (2017). Adversarial image perturbation for privacy protection a game theory perspective. In *Proceedings of IEEE International Conference on Computer Vision*, pages 1491–1500.

Ribaric, S., Ariyaeeinia, A., and Pavesic, N. (2016). De-identification for privacy protection in multimedia content. *Image Communication*, 47(C):131–151.

Ruchaud, N., Antipov, G., Korshunov, P., Dugelay, J. L., Ebrahimi, T., and Berrani, S. A. (2015). The impact of privacy protection filters on gender recognition. *Applications of Digital Image Processing XXXVIII*, page 959906.

Sattar, H., Bulling, A., and Fritz, M. (2017). Predicting the category and attributes of visual search targets using deep gaze pooling. In *Proceedings of IEEE International Conference on Computer Vision Workshops*, pages 2740–2748.

Schumann, A. and Stiefelhagen, R. (2017). Person re-identification by deep learning attribute-complementary information. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1435–1443.

Sudowe, P., Spitzer, H., and LeibeSudowe, B. (2015). Person attribute recognition with a jointly-trained holistic cnn model. In *Proceedings of IEEE International Conference on Computer Vision Workshop*, pages 329–337.

Sugano, Y., Ozaki, Y., Kasai, H., Ogaki, K., and Sato, Y. (2014). Image preference estimation with a data-driven approach: A comparative study between gaze and image features. *Eye Movement Research*, 7(3):862–875.

Xu, M., Ren, Y., and Wang, Z. (2015). Learning to predict saliency on face images. In *Proceedings of IEEE International Conference on Computer Vision*, pages 3907–3915.

Yamada, T., Gohshi, S., and Echizen, I. (2013). Privacy visor: Method for preventing face image detection by using differences in human and device sensitivity. In *Proceedings of International Conference on Communications and Multimedia Security*, pages 152–161.

Yubin, D., Ping, L., Change, L. C., and Xiaoou, T. (2014). Pedestrian attribute recognition at far distance. In *Proceedings of 22nd ACM International Conference on Multimedia*, pages 789–792.

Zhang, Y., Lu, Y., Nagahara, H., and Taniguchi, R. (2014). Anonymous camera for privacy protection. In *Proceedings of 22nd International Conference on Pattern Recognition*, pages 4170–4175.

Zhao, Q., Chang, S., Harper, F. M., and J. A. Konstan, J. (2016). Gaze prediction for recommender systems. In *Proceedings of 10th ACM Conference on Recommender Systems*, pages 131–138.