Reduction in Communication via Image Selection for Homomorphic Encryption-based Privacy-Protected Person Re-identification

Shogo Fukuda[1], Masashi Nishiyama[2] and Yoshio Iwai[2]
[1]*Graduate School of Sustainability Science, Tottori University, Tottori, Japan*
[2]*Graduate School of Engineering, Tottori University, Tottori, Japan*
nishiyama@tottori-u.ac.jp

Keywords: Image selection, homomorphic encryption, features, person re-identification, privacy protection

Abstract: We propose a method for reducing the amount of communication by selecting pedestrian images for privacy-protected person re-identification. Recently, it has become necessary to pay attention to how features corresponding to personal information can be protected. Owing to homomorphic encryption, which enables a distance between features to be computed without decryption, our method can use a cloud server on a public network while protecting personal information. However, we must consider the problem of the large amount of communication that occurs between camera clients and the server when homomorphic encryption is used. Our method aims to reduce the amount of this communication by selecting appropriate pedestrian images using reference leg postures. In our experiment, we confirmed that the amount of communication dynamically reduces without significant degradation in the accuracy of privacy-protected person re-identification with homomorphic encryption.

1 INTRODUCTION

Multiple-camera monitoring systems are needed for realizing a safe and secure society. Such systems are useful when searching for a criminal or lost person. In this paper, we consider a multiple-camera monitoring system set in indoor corridors where pedestrians must pass when moving between rooms.

The postures of the pedestrians change over time while they are walking. To correctly match pedestrian images acquired from the same individual, it is necessary to extract features that represent identities from the images and compute the distances between these features, as described in (Gong et al., 2014; Wang et al., 2018; Almasawa et al., 2019; Ye et al., 2020; Leng et al., 2020). For example, some existing methods (Nakajima et al., 2003; Bird et al., 2005; Farenzena et al., 2010) extract low-level image features such as color and gradient histograms, and other existing methods (Bourdev et al., 2011; Layne et al., 2012; Zhang et al., 2014; Lin et al., 2019; Zhao et al., 2019) extract human attribute features such as gender and age.

When designing a multi-camera monitoring system, we must pay attention to the fact that the features used in person re-identification are personal information. This is because the features are personally identifiable[1]. Thus, multiple-camera monitoring systems need to handle features with encryption so that third parties cannot access them. To protect the personal information of the features, we focus on homomorphic encryption (Brakerski et al., 2012). There are some prevalent methods of pattern recognition using homomorphic encryption, such as principal component analysis (Lu et al., 2016) and deep learning (Dowlin et al., 2016). Homomorphic encryption also enables the distance to be computed using addition and multiplication operations without decrypting the target features, as described in (Morita et al., 2018). In this paper, we apply homomorphic encryption to features for person re-identification and perform the distance computation while the features are still encrypted.

When homomorphic encryption is used, this causes a severe problem in the multi-camera monitoring system in that the amount of communication between each camera client and the server becomes very large. Since the data size of features after homomorphic encryption is much larger than that before encryption, the amount of communication per feature increases. Furthermore, the data size is proportional to the number of pedestrian images, which is the stage

[1]General Data Protection Regulation (May 25, 2018)

before the conversion to features, and thus the amount of communication further increases when the number of images is increased.

To avoid the problem of the data size, it is necessary to devise ways to reduce the amount of communication. There are two possible ways: by reducing the number of dimensions of the features or by selecting the features to be sent to the server. In this paper, we focus on the feature selection approach. In particular, we consider reducing the number of pedestrian images (before feature extraction), which then reduces the amount of communication. Many pedestrian images are acquired as a person walks in the field of view of a camera client. Specifically, the number of pedestrian images increases as a multiplication of the number of people in the field of view, the time length, and the camera's frame rate. Moreover, the number of cameras causes a problem for the server side. If all the acquired pedestrian images are directly received at the cloud server, the amount of communication increase drastically.

In this paper, to reduce the number of pedestrian images and the amount of communication, we propose a method for selecting appropriate pedestrian images for feature extraction for person re-identification from among the large number of pedestrian images acquired from each camera. We incorporate an image selection approach into a privacy-protected person re-identification system using homomorphic encryption. To do this, we focus on the cyclic motion of leg postures during walking as a cue for selecting pedestrian images. Using the leg posture as a cue, our method extracts the features from the pedestrian images containing the same appearances of the persons from the cameras. In our experiments, we confirmed that our method effectively reduces the number of pedestrian images while maintaining the accuracy of privacy-protected person re-identification.

2 RELATED WORK

We consider an idea to reduce the amount of communication between the camera client and the server for person re-identification. One idea is to apply a tracking process (Shu et al., 2012; Kawanishi et al., 2017; Li et al., 2013) to pedestrians in the video sequence and extract features from the time-series images linked by the tracking. An existing method (Khan and Brémond, 2017) extracts features of body parts from multiple images of a time series generated from the tracking process. This method uses Gaussian mixture models (GMMs) to represent the features of each body part of a person. However, GMMs have the disadvantage of a large feature size per person caused by maintaining the mean vector and the covariance matrix. Thus, it is not suitable for reducing the amount of communication, which is the purpose of this paper. An existing method (Han and Bhanu, 2006) focuses on gait features to integrate multiple images of a time series generated from the tracking process into a single averaged image. The gait features containing cyclic limb movements are represented in the single image. Because this method assumes that one complete gait cycle has been acquired, it is not possible to extract the features when the tracking process is interrupted. In this paper, instead of using tracking methods, we design a method based on image selection.

We consider one other idea to reduce the amount of communication, which is to reduce the dimensionality of the features. For example, some existing methods (Weinberger et al., 2009; Shi et al., 2009; Chen et al., 2015; Ullah et al., 2020; Ma et al., 2019; Fang et al., 2020) encode features using hash functions, and others (Gong et al., 2013; Ambai et al., 2015; Irie et al., 2017; Wu et al., 2020; Zang et al., 2019; Qayyum et al., 2019) use vector quantization and sparse coding to reduce the dimensionality. The large number of pedestrian images increases is still a problem, even if hashing and sparse coding have been applied. Thus, we develop a method to directly solve the problem of the high number of pedestrian images.

Existing face recognition methods (Chang et al., 2008; Z. Yang et al., 2004; Wong et al., 2011) select frontal face images from the sets of images detected in a video sequence using head pose estimation. Although these existing methods aim to obtain similar effect as that of our method, it is not easy to apply them directly to person re-identification, which uses the whole body of a person. Indeed, the existing methods for facial image selection are not suitable when the person's back is facing the camera. Thus, our method selects appropriate pedestrian images from the sets of images acquired from each camera client.

3 PEDESTRIAN IMAGE SELECTION FOR PERSON RE-IDENTIFICATION

3.1 Overview

We use the leg posture to select a small number of appropriate pedestrian images from the many pedestrian

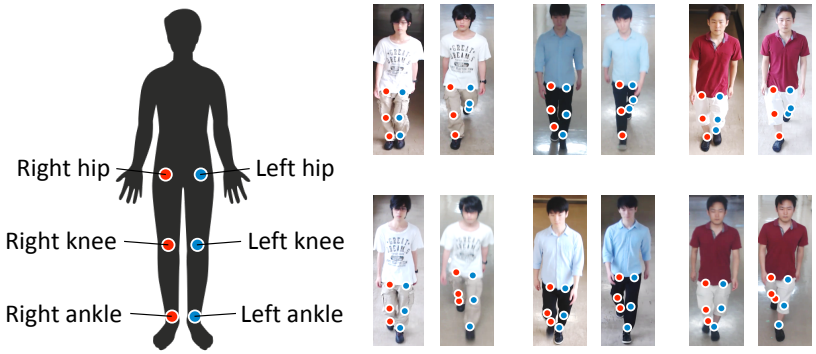

Figure 1: Body joints for representing a spatial cue of the change of leg posture. The temporal cue refers to the cyclic movements of the legs of the body joints that change over time.

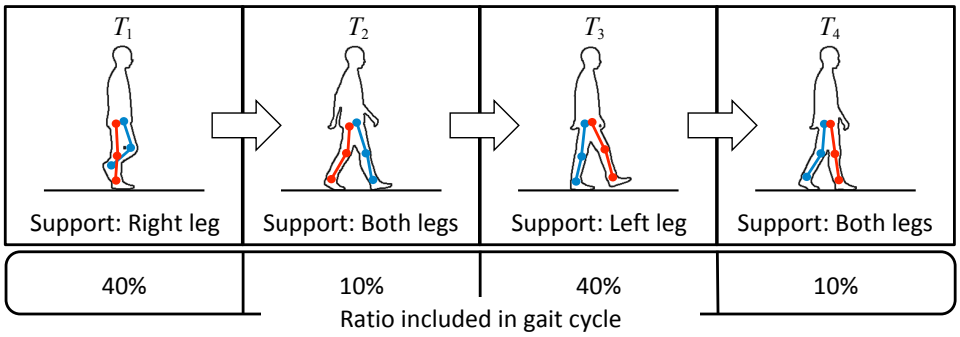

Figure 2: Representation of the temporal cues of leg posture using the categories based on the gait cycle. We represent each category in time direction from T_1 to T_4. In general, the duration of one cycle of the gait cycle is about one second.

images acquired from each camera. A person walking in a corridor repeats a cyclic motion of her/his legs. Thus, we assume that there will be many opportunities to acquire pedestrian images containing reference leg postures. If we acquire a small number of pedestrian images with the same reference leg posture, we can expect to re-identify a person without decreasing the accuracy.

We discuss the reason for the use of leg posture below. The same arm posture cannot be repeatedly acquired in the same way as leg posture because the arm movements during walking are more flexible, for instance, if the pedestrian is carrying something or folding his/her arms. The head and torso postures are less effective in reducing the number of images than the leg postures because the same posture of the head and torso frequently appears during walking. Thus, we adopt the leg posture to reduce the number of pedestrian images while maintaining the person re-identification accuracy.

The leg posture of a walking person is represented using spatial and temporal cues. A spatial cue refers to the change of the locations of the body joints. Specifically, we use the cue of the spatial locations of the body joints shown in Figure 1. The temporal cue refers to the cyclic movements of the legs of the body joints that change over time. In the next section, we describe the motion cycle of the leg posture in detail to consider the reference leg postures for image selection.

3.2 Gait cycle of the leg posture

The motion change in the leg posture during walking is called the gait cycle in (Perry and Burnfield, 2010). It is stated that the gait cycle is repeated as a cyclic leg movement while walking. The gait cycle is divided into four categories $T_c (c = 1, \ldots, 4)$ according to the differences in the leg postures, as shown in Figure 2. Although the start time of the gait cycle is not clearly defined in (Perry and Burnfield, 2010), we treat the moment when the left leg leaves the ground as the start time of the gait cycle. We represent each category in time direction from T_1 to T_4. In general, the duration of one cycle of the gait cycle is about one second. Categories T_1 and T_3 are both 40% of the duration, and T_2 and T_4 are both 10% of the duration.

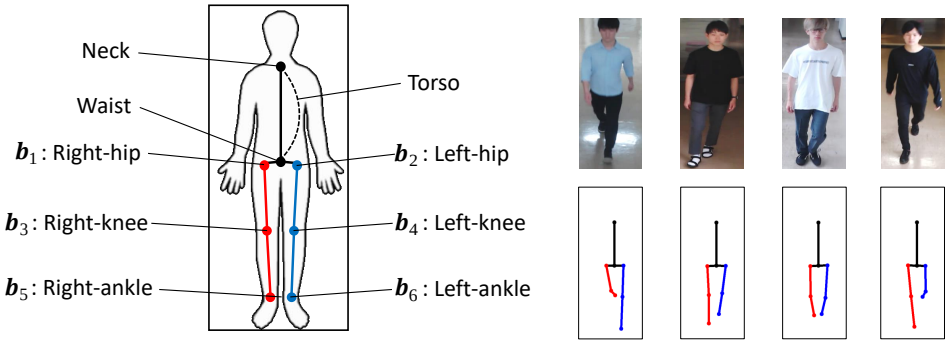

Figure 3: Representation of body joints $\boldsymbol{b}_j$ after the length of the torso has been normalized. We assign two-dimensional locations to the body joints $\boldsymbol{b}_j(j = 1,\ldots,6)$ of each person at each time in each category T_c.

3.3 Reference leg postures for image selection

We describe how to generate reference leg postures $P_c(c = 1,\ldots,4)$, which are the cues for selecting appropriate pedestrian images. We represent the temporal cue of the leg posture using each T_c of the gait cycle. We also represent the spatial cue as the average locations of the body joints in each T_c. To generate the reference leg postures, we use time-series signals of the locations of the body joints in the gait cycle observed from various persons. To do this, we need to consider the temporal variation of the locations for each person and the spatial variation of the locations among them. We describe the detail generation of the reference postures below.

Suppose that the three-dimensional locations of the body joints are observed by a motion capture system during the gait cycle of a large number of individuals. We project the body joints onto the image plane using a predefined perspective projection matrix. Here, our discussion assumes that the height and angle of each camera set in the corridor are the same. We normalize the scale and rotation to suppress the differences in the body height of the persons. We then make the length of the torso, which connects the neck and the waist, constant in the image plane.

Figure 3 shows our representation of the body joints after normalizing the length of the torso. According to the definition of the gait cycle described in Section 3.2, we assign two-dimensional locations to the body joints $\boldsymbol{b}_j(j = 1,\ldots,6)$ of each person at each time in each category T_c. We generate the reference posture P_c for T_c by calculating the average values of these two-dimensional locations. We experimentally determine which of the four reference postures P_c should be used in Section 4.

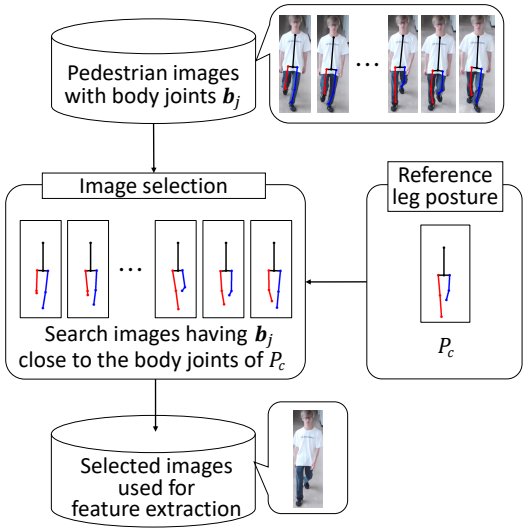

Figure 4: Overview of appropriate pedestrian image selection for person re-identification. Our method searches for pedestrian images with leg postures close to the reference posture P_c.

3.4 Pedestrian image selection

In this section, we describe how we select appropriate pedestrian images from a set of images detected in each camera client. Our method searches for pedestrian images with leg postures close to the reference posture. An overview of this pedestrian image selection is shown in Figure 4. We use the reference posture P_c described in Section 3.3. After image selection, our method extracts features for person re-identification from the selected images.

In the following, we describe the detail of the image selection. Given an image of a pedestrian, the locations of the body joints $\boldsymbol{b}_j$ are determined using a two-dimensional posture prediction technique, e.g., OpenPose (Cao et al., 2017). To compensate for the differences in body length between persons, we nor-

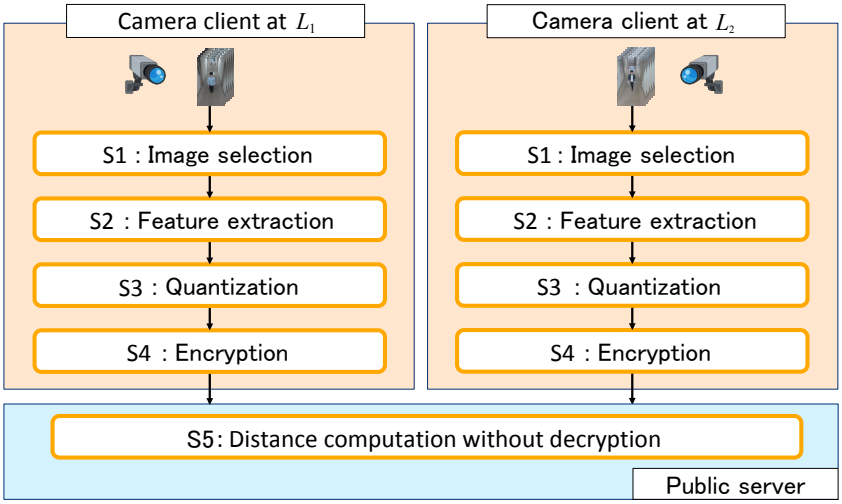

Figure 5: Overview of our method for privacy-protected person re-identification. In S1, our method selects the appropriate pedestrian images at each camera client. In S2, our method extracts features from the selected pedestrian images using co-occurrence attributes. In S3, our method applies linear quantization to the features as pre-processing for homomorphic encryption. In S4, our method encrypts the quantized features. Then, our method sends them from each camera client to the server. In S5, our method calculates the distance between the received features without decrypting them in the server.

malize the locations of $\boldsymbol{b}_j$ using the geometrical transformation described in Section 3.3 so that the length and rotation of the torso of an input pedestrian image is equal to that of the reference posture.

We compute the cost of the difference between the location of the body joints $\boldsymbol{b}_j$ in the reference posture and that of the body joints in the normalized posture using the L2 norm. If this cost is less than or equal to a threshold R, our method determines that the pedestrian image containing a similar $\boldsymbol{b}_j$ is suitable for feature extraction for person re-identification. Note that we experimentally determined R so that one or more images of each person are selected for each camera client while she/he exists in the field of view of the camera.

3.5 Privacy-protected person re-identification

Figure 5 shows an overview of our method for privacy-protected person re-identification. In S1, our method selects the appropriate pedestrian images at each camera client (L_1, L_2) using body joint $\boldsymbol{b}_j$ with the procedure described in Section 3.4. In S2, our method extracts features $\boldsymbol{f}_q$ and $\boldsymbol{f}_t$ from the selected pedestrian images using co-occurrence attributes proposed by (Nishiyama et al., 2016). Co-occurrence attributes are represented by combinations of physical and adhered human characteristics (e.g., a man wearing a suit, a 20-something woman, or a woman with long hair who is wearing a skirt). In S3, our method applies linear quantization to the features as pre-processing for homomorphic encryption. In S4, our method encrypts the quantized features using the fully homomorphic encryption (FHE) library[2]. Then, our method sends them from each camera client to the server. In S5, our method calculates the distance between the received features $\boldsymbol{f}_q$ and $\boldsymbol{f}_t$ without decrypting them in the server. We use the large margin nearest neighbor (LMNN) proposed by (Weinberger and Saul, 2009), which is a metric learning technique. The details of the distance computation for privacy-protected person re-identification are described in (Morita et al., 2018). In our system, it is no longer necessary to decrypt the features on the server. Thus, we believe that it is possible to operate a secure system for privacy-protected person re-identification.

4 EXPERIMENTS

4.1 Experimental conditions

We confirmed whether our method could reduce the amount of communication used for privacy-protected person re-identification. In particular, we evaluated the effectiveness of pedestrian image selection using the reference leg posture. We used the first match rate to evaluate the accuracy of person re-identification. We also used the reduction rate to evaluate the amount

[2]HElib https://github.com/shaih/HElib

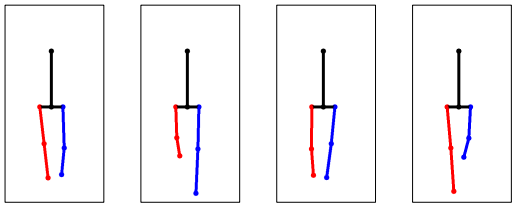

Figure 6: Reference leg postures $P_c(c = 1,\ldots,4)$ generated using the projection matrix of the TUP dataset.

of communication. We used pedestrian images from one camera client as the target and pedestrian images from another camera client as the query. We selected pedestrian images using each reference posture $(P_1,\ldots,P_4)$. Next, we computed the average of the first match rate for the target and query images described above. The reduction ratio is the ratio of the number of images not selected by the method to the number of pedestrian images when no image selection is used. We averaged the reduction rates over various combinations of camera clients.

In this experiment, we used the TUP (Tottori University Pedestrian) dataset, which we originally collected, and a publicly available MARS (Motion Analysis and Re-identification Set) (Zheng et al., 2016) dataset. The details of each dataset are described in Sections 4.3 or 4.4, respectively. First, we describe the parameters of person re-identification. We set the dimension of features $\boldsymbol{f}_q$ or $\boldsymbol{f}_t$ to 95. We used 200 pedestrian images from the SARC3D dataset (Baltieri et al., 2010) to generate a metric matrix for LMNN in the experiments on the TUP dataset. For the experiments on the MARS dataset, we used 1000 randomly selected images for the LMNN metric matrix from the training images included in the MARS dataset.

4.2 Generation of the reference leg postures

We generated the reference postures using the AIST Gait Database 2015 provided by (Kobayashi et al., 2015). The dataset consists of 214 healthy Japanese participants. The three-dimensional locations of the body joints were measured using a motion capture system from walking participants. We used 100 participants randomly selected from the dataset and averaged the locations of the body joints of each participant during one gait cycle using the procedure described in Section 3.3. By averaging the locations of the participants, we generated a reference posture P_c for each category T_c. Note that the motion capture sensor attached to the surface of the body cannot directly measure the body joints $\boldsymbol{b}_j(j = 3,\ldots,6)$ in Figure 3. We used average locations from several peripheral joints. Because the camera settings are different for the TUP and MARS datasets, we used a perspective projection matrix manually adjusted to each dataset. Figure 6 shows the reference leg postures $P_c(c = 1,\ldots,4)$ generated using the projection matrix of the TUP dataset.

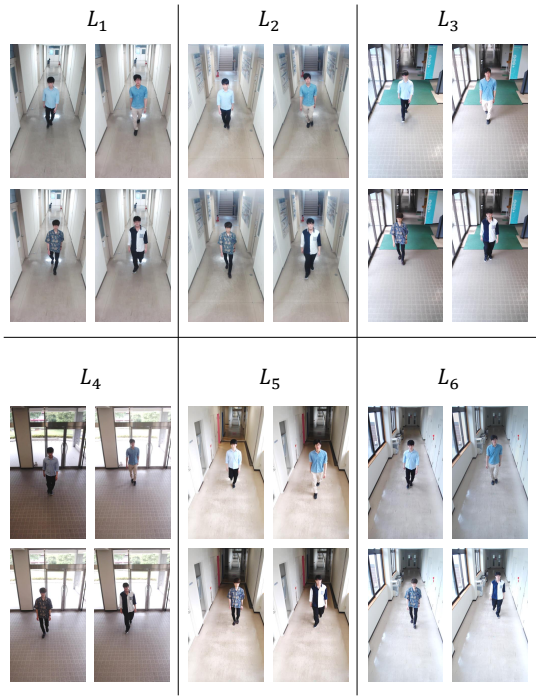

Figure 7: Examples of images acquired from the camera clients $(L_1,\ldots,L_6)$ in the TUP dataset.

Next, we describe the parameters of the image selection for pedestrians. OpenPose (Cao et al., 2017) was used to predict the two-dimensional locations of the body joints shown in Figure 3 from the pedestrian images. We adjusted the threshold R of image selection for each dataset. As described in Section 3.4, we searched for the best R such that one or more images were selected for each participant. We obtained $R = 100$ for the TUP dataset and $R = 375$ for the MARS dataset.

4.3 Evaluations on the TUP dataset

4.3.1 Conditions of image acquisition

We collected pedestrian images when walking along indoor corridors using six camera clients. Figure 7 shows examples of images acquired from the camera clients $(L_1,\ldots,L_6)$ in the TUP dataset. The cameras were set at the same height of 2.4 m from the floor. The angles were set the same so that the center of the camera image was located 3 m in front of it. We

Figure 8: Examples of pedestrian images of five different individuals contained in the TUP dataset. We acquired the pedestrian images from the camera clients $(L_1, \ldots, L_6)$.

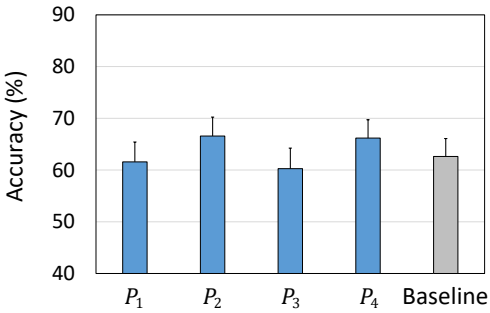

Figure 9: Average accuracy of person re-identification for the camera clients $(L_1, \ldots, L_6)$ of the TUP dataset.

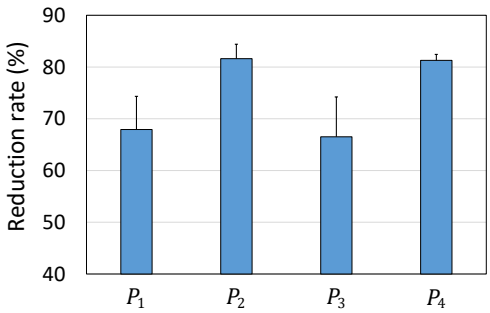

Figure 10: Average reduction rate of pedestrian images for the camera clients $(L_1, \ldots, L_6)$ of the TUP dataset.

used a USB camera (logicool C920, 1920×1080 pixels), and the frame rate was set to 30 fps. The number of participants was 31, but the number of people walking at the same time was limited to one. We instructed the participants to pass through the center of the corridor while walking. Figure 8 shows examples of pedestrian images in the TUP dataset. There were 5,117 pedestrian images collected from camera client L_1, 5,060 collected from L_2, 2,895 collected from L_3, 2,418 collected from L_4, 4,841 collected from L_5, and 5,082 collected from L_6.

4.3.2 Evaluation results

Figure 9 shows the average person re-identification accuracy for camera clients $(L_1, \ldots, L_6)$ of the TUP dataset. Our image selection method using reference leg postures P_2 and P_4 obtained better accuracy than the baseline method using all pedestrian images without image selection. Figure 10 shows the average reduction rate when using reference posture $P_c (c = 1, \ldots, 4)$. If image selection is not applied (the baseline method), the reduction rate was 0%, and all pedestrian images were used. The reduction rates of P_2 and P_4 were higher than those of P_1 and P_3. We believe that reference postures P_2 and P_4, which had high reduction rates and person re-identification accuracy, are suitable for the TUP dataset.

Figure 11 shows examples of selected pedestrian images. We used reference posture P_2 at camera client L_4. In the figure, we listed pairs of pedestrian images and their leg postures picked up from one gait cycle.

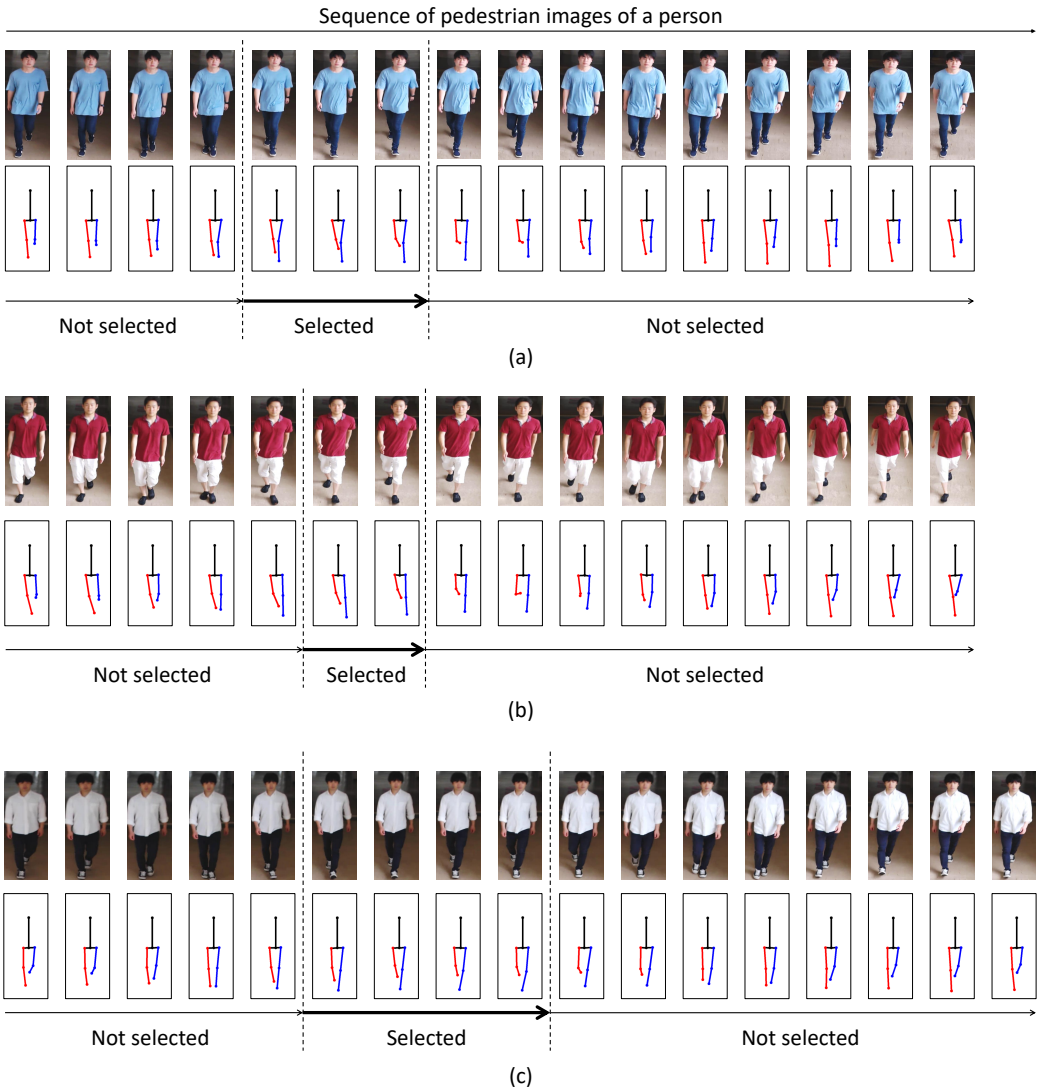

Figure 11: Examples of pedestrian images selected using the reference posture P_2 from camera client L_4 of the TUP dataset. In the figure, we listed pairs of pedestrian images and their leg postures picked up from one gait cycle.

Our method selected 3 out of 16 images of person 1 in (a), 2 out of 16 images of person 2 in (b), 4 out of 17 images of person 3 in (c). We see that the leg postures of the selected pedestrian images are similar to the reference posture P_2 in Figure 6.

4.3.3 Amount of communication from the camera client to the server

We evaluated the amount of communication for the case in which L_1 and L_2 camera clients send features to the cloud server. The amount of communication per feature increased from 760 B to 524.4 KB before and after encryption. The number of pedestrian images detected by camera client L_1 was 5,117 and 5,060 were detected by L_2. Moreover, the amount of communication of the baseline method with the encrypted features extracted from all the pedestrian images was 2.68 GB for L_1 and 2.65 GB for L_2. In contrast, when the pedestrian images were selected using reference pose P_2, the number of pedestrian images was reduced to 903 for L_1 and 807 for L_2. The amount of communication of our method was 0.47 GB for L_1 and 0.42 GB for L_2. We see that the amount of communication and number of images were both reduced by 82.4% in L_1 and 84.1% in L_2. These results confirm that our method can substantially reduce the amount of communication for privacy-protected person re-identification using homomorphic encryption.

Figure 12: Examples of pedestrian images contained in the MARS dataset.

4.4 Evaluations on the MARS dataset

4.4.1 Conditions of image acquisition

We also evaluated our method on the publicly available MARS dataset. Figure 12 shows examples of pedestrian images contained in the MARS dataset. This dataset includes people walking freely in an outdoor plaza, but it can be assumed that the pedestrian traffic line is constant. We used the image sets of camera clients C1 and C2. These sets contain only the pedestrian image sequences after detection and tracking. Because the tracking results of some image sequences are often interrupted, we used only those containing more than 20 pedestrian images in one sequence. We evaluated the accuracy of 4,540 pedestrian images for 41 subjects of C1 and 3,948 pedestrian images for 40 subjects of C2. We searched for persons walking in the same direction in the field of view of cameras C1 or C2 at different times.

4.4.2 Evaluation results

Figure 13 shows reference leg postures generated using the projection matrix of the MARS dataset. Because of the different heights and angles of cameras C1 and C2, different respective perspective projection matrices were applied. The reference posture shown in (a) was used in the evaluation of C1 and the reference leg posture shown in (b) was used in the evaluation of C2.

Figure 14 shows the accuracy of person re-identification on the MARS dataset. We see that the accuracies of our method using image selection with reference leg postures P_2 and P_4 were better than that of the baseline method without image selection. Figure 15 shows the reduction rate of pedestrian images in the MARS dataset. If image selection is not applied (the baseline method), the reduction rate was 0%, and all pedestrian images were used. The reduction rates of P_2 and P_4 were higher than those of P_1 and P_3. We see that reference postures P_2 and P_4, which had both good reduction rates and person re-identification accuracy, were suitable for the camera settings of the MARS dataset. These results confirm that our method can reduce the number of pedestrian images in the MARS dataset while maintaining the accuracy of privacy-protected person re-identification.

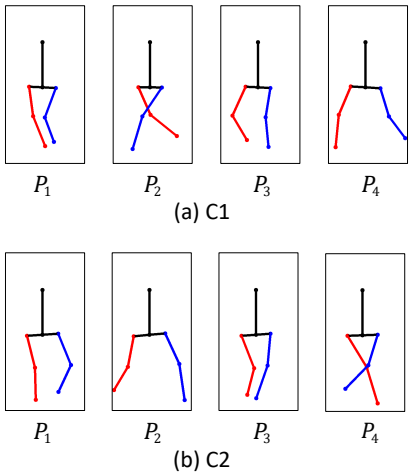

Figure 13: Reference leg postures generated using the projection matrix of the MARS dataset.

5 CONCLUSIONS

To reduce the amount of communication needed for privacy-protected person re-identification, we proposed a method for selecting appropriate pedestrian images. In the proposed method, we focused on the gait cycle of the leg posture as a cue for image selection. We generated the reference leg postures from the body joints distributed in each category of the gait cycle. We incorporated the image selection into a privacy-protected person re-identification system using homomorphic encryption and evaluated the effectiveness of our method on the TUP and MARS datasets. We confirmed that the amount of communication could be substantially reduced while achieving the same or better accuracy of person re-identification using the reference leg postures for selecting appropriate pedestrian images. In future work, we intend

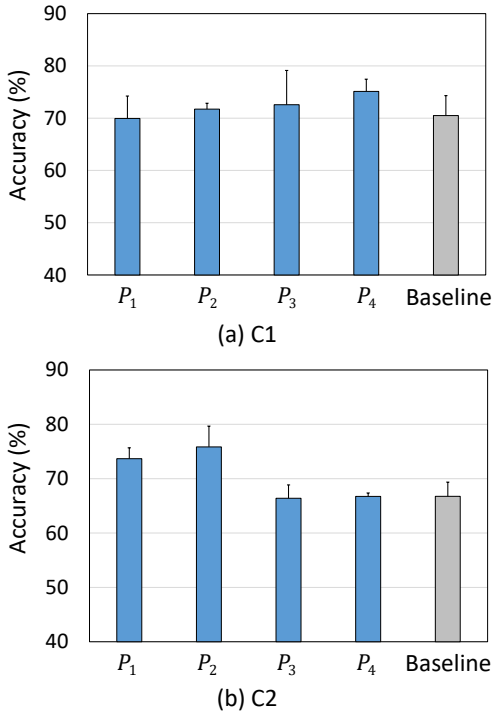

Figure 14: Accuracy of person re-identification on the MARS dataset.

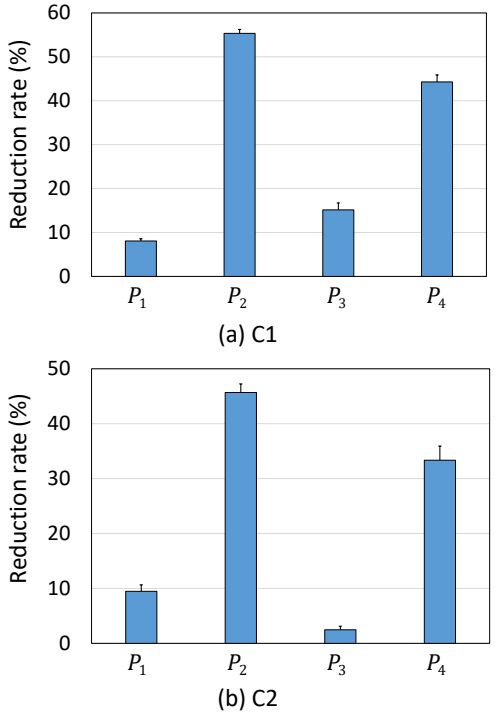

Figure 15: Reduction rate of pedestrian images in the MARS dataset.

to develop a reference posture generation for the case where the pedestrian traffic line is not constant as well as when the height and angle of the camera are not constant. We will expand our evaluations to investigate the effectiveness of our method when the number of cameras and number of people increase. We also intend to combine our method with dimensional reduction, e.g., hashing and sparse coding.

ACKNOWLEDGEMENT

This work was partially supported by JSPS KAKENHI under grant number JP17K00238, JP18H04114, and JP20K11865.

REFERENCES

Almasawa, M. O., Elrefaei, L. A., and Moria, K. (2019). A survey on deep learning-based person re-identification systems. *Journal of IEEE Access*, 7:175228–175247.

Ambai, M., Kimura, T., and Sakai, C. (2015). Fast and accurate object detection based on binary co-occurrence features. *Journal of IPSJ Transactions on Computer Vision and Applications*, 7:55–58.

Baltieri, D., Vezzani, R., and Cucchiara, R. (2010). 3d body model construction and matching for real time people re-identification. In *Proceedings of Eurographics Italian Chapter Conference*.

Bird, N. D., Masoud, O., Papanikolopoulos, N. P., and Isaacs, A. (2005). Detection of loitering individuals in public transportation areas. *Journal of IEEE Transactions on Intelligent Transportation Systems*, 6(2):167–177.

Bourdev, L., Maji, S., and Malik, J. (2011). Describing people: A poselet-based approach to attribute classification. In *Proceedings of International Conference on Computer Vision*, pages 1543–1550.

Brakerski, Z., Gentry, C., and Vaikuntanathan, V. (2012). Fully homomorphic encryption without bootstrapping. In *Proceedings of Innovations in Theoretical Computer Science Conference*, pages 309–325.

Cao, Z., Simon, T., Wei, S., and Sheikh, Y. (2017). Realtime multi-person 2d pose estimation using part affinity fields. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1302–1310.

Chang, L., Rodés, I., Méndez, H., and del Toro, E. (2008). Best-shot selection for video face recognition using fpga. In *Proceedings of Iberoamerican Congress on Pattern Recognition*, pages 543–550.

Chen, W., Wilson, J., Tyree, S., Weinberger, K., and Chen, Y. (2015). Compressing neural networks with the hashing trick. In *Proceedings of International Conference on Machine Learning*, pages 2285–2294.

Dowlin, N., Gilad-Bachrach, R., Laine, K., Lauter, K., Naehrig, M., and Wernsing, J. (2016). Cryptonets: applying neural networks to encrypted data with high

throughput and accuracy. In *Proceedings of International Conference on Machine Learning*, volume 48, pages 201–210.

Fang, S., Wang, J., Yang, C., and Tong, P. (2020). Fast retrieval method of image data based on learning to hash. *Journal of Physics: Conference Series*, 1631:012029.

Farenzena, M., Bazzani, L., Perina, A., Murino, V., and Cristani, M. (2010). Person re-identification by symmetry-driven accumulation of local features. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2360–2367.

Gong, S., Cristani, M., Yan, S., and Loy, C. C. (2014). *Person Re-Identification*. Springer-Verlag London.

Gong, Y., Lazebnik, S., Gordo, A., and Perronnin, F. (2013). Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(12):2916–2929.

Han, J. and Bhanu, B. (2006). Individual recognition using gait energy image. *Journal of IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(2):316–322.

Irie, G., Arai, H., and Tanigushi, Y. (2017). Multimodal learning of geometry-preserving binary codes for semantic image retrieval. *Journal of IEICE Transactions on Information and Systems*, E100.D(4):600–609.

Kawanishi, Y., Deguchi, D., Ide, I., and Murase, H. (2017). Trajectory ensemble: Multiple persons consensus tracking across non-overlapping multiple cameras over randomly dropped camera networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 1471–1477.

Khan, F. M. and Brémond, F. (2017). Multi-shot person re-identification using part appearance mixture. In *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, pages 605–614.

Kobayashi, Y., Hobara, H., and Mochimaru, H. (2015). Aist gait database 2015. https://www.airc.aist.go.jp/dhrt/gait2015/index.html.

Layne, R., Hospedales, T. M., and Gong (2012). Person re-identification by attributes. In *Proceedings of the British Machine Vision Conference*, number 24, pages 1–11.

Leng, Q., Ye, M., and Tian, Q. (2020). A survey of open-world person re-identification. *Journal of IEEE Transactions on Circuits and Systems for Video Technology*, 30(4):1092–1108.

Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A., and Hengel, A. V. D. (2013). A survey of appearance models in visual object tracking. *Journal of ACM Transactions on Intelligent Systems and Technology*, 4(4):58:1–58:48.

Lin, Y., Zheng, L., Zheng, Z., Wu, Y., Hu, Z., Yan, C., and Yang, Y. (2019). Improving person re-identification by attribute and identity learning. *Journal of Pattern Recognition*, 95:151 – 161.

Lu, W., Kawasaki, S., and Sakuma, J. (2016). Using fully homomorphic encryption for statistical analysis of categorical, ordinal and numerical data. In *Proceedings of Network and Distributed System Security Symposium*, pages 201–210.

Ma, X.-Q., Yu, C.-C., Chen, X.-X., and Zhou, L. (2019). Large-scale person re-identification based on deep hash learning. *Journal of Entropy*, 21(5):449.

Morita, K., Yoshimura, H., Nishiyama, M., and Iwai, Y. (2018). Protecting personal information using homomorphic encryption for person re-identification. In *Proceedings of IEEE 7th Global Conference on Consumer Electronics*, pages 166–167.

Nakajima, C., Pontil, M., Heisele, B., and Poggio, T. (2003). Full-body person recognition system. *Journal of Elsevier Pattern Recognition*, 36(9):1997–2006.

Nishiyama, M., Nakano, S., Yotsumoto, T., Yoshimura, H., Iwai, Y., and Sugahara, K. (2016). Person re-identification using co-occurrence attributes of physical and adhered human characteristics. In *Proceedings of International Conference on Pattern Recognition*, pages 2085–2090.

Perry, J. and Burnfield, J. M. (2010). *Gait Analysis: Normal and Pathological Function*. Slack Inc.

Qayyum, A., Malik, A., M Saad, N., and Mazher, M. (2019). Designing deep cnn models based on sparse coding for aerial imagery: a deep-features reduction approach. *Journal of Remote Sensing*, 52(1):221–239.

Shi, Q., Petterson, J., Dror, G., Langford, J., Smola, A., Strehl, A., and Vishwanathan, S. V. N. (2009). Hash kernels. In *Proceedings of International Conference on Artificial Intelligence and Statistics*, volume 5, pages 496–503.

Shu, G., Dehghan, A., Oreifej, O., Hand, E., and Shah, M. (2012). Part-based multiple-person tracking with partial occlusion handling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1815–1821.

Ullah, A., Muhammad, K., Hussain, T., Baik, S. W., and De Albuquerque, V. H. C. (2020). Event-oriented 3d convolutional features selection and hash codes generation using pca for video retrieval. *Journal of IEEE Access*, 8:196529–196540.

Wang, K., Wang, H., Liu, M., Xing, X., and Han, T. (2018). Survey on person re-identification based on deep learning. *Journal of CAAI Transactions on Intelligence Technology*, 3(4):219–227.

Weinberger, K., Dasgupta, A., Attenberg, J., Langford, J., and Smola, A. (2009). Feature hashing for large scale multitask learning. In *Proceedings of Annual International Conference on Machine Learning*, pages 1113–1120.

Weinberger, K. Q. and Saul, L. K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10:207–244.

Wong, Y., Chen, S., Mau, S., Sanderson, C., and Lovell, B. C. (2011). Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition. In *Proceedings of IEEE Computer Vision and Pattern Recognition*, pages 74–81.

Wu, P., Liu, J., Li, M., Sun, Y., and Shen, F. (2020). Fast sparse coding networks for anomaly detection in videos. *Journal of Pattern Recognition*, 107:107515.

Ye, M., Shen, J., Lin, G., Xiang, T., Shao, L., and Hoi, S. C. C. (2020). Deep learning for person re-identification: A survey and outlook. *CoRR*, abs/2001.04193.

Z. Yang, H. Ai, B. Wu, S. Lao, and L. Cai (2004). Face pose estimation and its application in video shot selection. In *Proceedings of the 17th International Conference on Pattern Recognition*, volume 1, pages 322–325.

Zang, M., Wen, D., Liu, T., Zou, H., and Liu, C. (2019). A fast sparse coding method for image classification. *Journal of Applied Sciences*, 9(3):505.

Zhang, N., Paluri, M., Ranzato, M., Darrell, T., and Bourdev, L. (2014). Panda: Pose aligned networks for deep attribute modeling. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1637–1644.

Zhao, Y., Shen, X., Jin, Z., Lu, H., and , Hua, X. (2019). Attribute-driven feature disentangling and temporal aggregation for video person re-identification. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4908–4917.

Zheng, L., Bie, Z., Sun, Y., Wang, J., Su, C., Wang, S., and Tian, Q. (2016). Mars: A video benchmark for large-scale person re-identification. In *Proceedings of Computer Vision and Pattern Recognition European Conference on Computer Vision*, pages 868–884.