information

MDPI

Article

Prevention of Unintended Appearance in Photos Based on Human Behavior Analysis

Yuhi Kaihoko [1], Phan Xuan Tan [2,*] and Eiji Kamioka [1,*]

[1] Graduate School of Engineering and Science, Shibaura Institute of Technology, Tokyo 135-8548, Japan; ma18028@shibaura-it.ac.jp
[2] Department of Information and Communications Engineering, Shibaura Institute of Technology, Tokyo 135-8548, Japan
* Correspondence: tanpx@shibaura-it.ac.jp (P.X.T.); kamioka@shibaura-it.ac.jp (E.K.)

Received: 23 July 2020; Accepted: 29 September 2020; Published: 2 October 2020

Abstract: Nowadays, with smartphones, people can easily take photos, post photos to any social networks, and use the photos for various purposes. This leads to a social problem that unintended appearance in photos may threaten the facial privacy of photographed people. Some solutions to protect facial privacy in photos have already been proposed. However, most of them rely on different techniques to de-identify photos which can be done only by photographers, giving no choice to photographed person. To deal with that, we propose an approach that allows a photographed person to proactively detect whether someone is intentionally/unintentionally trying to take pictures of him. Thereby, he can have appropriate reaction to protect the facial privacy. In this approach, we assume that the photographed person uses a wearable camera to record the surrounding environment in real-time. The skeleton information of likely photographers who are captured in the monitoring video is then extracted and put into the calculation of dynamic programming score which is eventually compared with a threshold for recognition of photo-taking behavior. Experimental results demonstrate that by using the proposed approach, the photo-taking behavior is precisely recognized with high accuracy of 92.5%.

Keywords: photo-taking behavior; photo capturing and sharing; bystanders; human behavior analysis; identity protection; facial privacy

1. Introduction

For years, smartphones have been increasingly become one of the most indispensable personal devices, allowing people to easily take photos recording every desirable moment with just a simple click. According to the Global Digital 2019 report, the number of people around the world who use a mobile phone accounts for 67%—more than two-thirds of the total global population [1]. In Japan, the statistics obtained from the Ministry of Internal Affairs and Communications show that the ownership rate of the smartphone is about 60.9%, and especially the rate of owners who are under 40 was over 90% in 2018 [2,3]. This facilitates the explorations of various Social Networking Services (SNS) (e.g., Facebook, Twitter, etc.). In fact, about 3.5 billion people accounting for 45% of the global population are using SNS [1]. As a result, a social problem potentially occurs when people are unintentionally captured in others' photos and which are then published on social networks. More seriously, the photos along with the photographed person's identity can be used by photographers for their own purposes. As a consequence, the facial privacy of the photographed person is severely violated. Recent advances in computer vision and machine learning techniques make this problem become more serious. Indeed, these techniques can automatically recognize people with extremely high accuracy, facilitating the possibility of searching for a specific person in vast image collections [4]. To combat

this privacy problem, numerous approaches have been introduced. One straightforward approach is to manually specify the regions containing subjects and apply appearance obscuration. However, this approach is time-consuming and not suitable for real-time privacy protection. For automatic privacy-protection purposes, the existing methods might be done at either photographer's site or the site of the photographed person. The methods in the former category leverage the power of computer vision and machine learning techniques to hide the identity of photographed persons to avoid their identification [5]. For example, Google developed cutting-edge face-blurring technology which can blur identifiable faces of bystanders on the images [6]. Other solutions aim to automatically recognize photographed persons in images and obscure their identities [7–9]. Unfortunately, these approaches give no choice to the photographed person to control over his privacy protection since this process is totally done at the photographer site. The methods in the latter category attempt to proactively prevent the photos of the photographed person from being taken. For example, some techniques force for the privacy of photographed person to be respected based on his privacy preferences represented by visual markers [10] or hand gestures [11] or offline tags [12] which are visible to everyone. However, the privacy preferences might vary widely among individuals and change from time to time, following patterns which cannot be conveyed by static visual markers [13]. More sophisticated techniques rely on cooperation between photographers and photographed persons, which enables photographed people's privacy preferences to be detected by nearby photo-taking devices (via peer-to-peer short-range wireless communications) [14–16]. However, this approach requires the photographed persons to broadcast their preferences, leading to other aspects of privacy. More importantly, this approach might not be effective in a situation where the photographers proactively switch off the communication function on their devices or ignore the advertised privacy choices of the nearby photographed persons because they are intentionally and secretly taking photos of pre-targeted persons. Indeed, this is also a common problem of most of existing studies, which mainly focus on the privacy-protection of the "bystander" which is defined as either "a person who is present and observing an event without taking part in" [9] or "a person who is unexpectedly framed in" [7] or "a person who is not a subject of the photo and is thus not important for the meaning of the photo, e.g., the person was captured in a photo only because they were in the field of view and was not intentionally captured by the photographer" [9]. In other words, the situation where the photographer intentionally takes photos of targeted persons has not been taken into account.

In this study, we propose an approach that allows the photographed person to proactively detect a situation where someone is intentionally/unintentionally trying to take photos of him using a mobile phone, without broadcasting his privacy preferences as well as identifying information. Afterward, the photographed person will have an appropriate reaction such as leaving the shared space or asking the photographer to stop taking the photos in order to protect his privacy. Importantly, in order to sufficiently cover as many of cases of photo-taking as possible, we use the notion of "photographed person" instead of "bystander". We assume that the photographed person has strong motivation to protect his privacy and is willing to use a wearable camera to monitor the surrounding environment. The behavior of the likely photographers is recognized via the analysis of their skeleton information obtained from the monitored video. Note that, in this study, we only use a normal camera to evaluate the potential of the proposed approach. In practice, a thermographic camera should be used to hide the facial identities of people who are captured in the monitored video. However, there is no technical difference between the normal camera and the thermographic camera in detecting photo-taking behaviors of the photographer since the proposed method uses only the photographer's skeleton information which can be precisely obtained by both types of cameras. On another front, we argue that misdetection possibly occurs when there is behavior—i.e., net-surfing—which is similar to photo-taking behavior. Basically, the human arm parts are believed to significantly contribute to the precise recognition of photo-taking behavior. Thus, only skeleton information of the arm parts including length and angle transition is focused on in the analysis process. In our study, such information is extracted by OpenPose [17] in real-time. Afterwards, dynamic programing (DP) matching between

monitored data and reference data is performed to generate monitored DP scores which are then compared with a pre-determined DP threshold. The comparison results decide whether the input data represents photo-taking behavior. The experimental results demonstrate that the proposed approach achieve an accuracy of 92.5% in recognizing photo-taking behavior.

The remainder of the paper is organized as follows: related work is provided in Section 2. Meanwhile, Section 3 describes the proposed method. Performance evaluation of the proposed method is discussed in Section 4, and Section 5 concludes this study.

2. Related Works

Prior works on handling photographed people's facial privacy can be classified into two categories: photographer-site methods, which leverage obfuscation techniques to hide the identity of photographed persons; and photographed person-site methods, which deny third party devices the opportunity to collect data.

2.1. Photographer-Site Methods

As the image sources are explosively growing and easily accessible, de-identification has become extremely important. It refers to the reversible process of removing or obscuring any personally identifiable information from an individual [18]. Thus, to deal with this privacy problem, the common approaches are blurring and pixelization. For example, Frome et al. [5] proposed a method for automatic privacy protection for all people captured in Google Street View, where a fast sliding-window approach was applied for face detection and post-processing was performed to blur the faces. Koyama et al. [7] introduced a new system to automatically generate privacy-protected videos in real-time to protect the privacy of non-intentionally captured persons (ICPs). In real scenarios, social videos posted via social networks include not only ICPs but also non-ICPs. The authors also claimed that existing privacy protection systems simply blur out all the people in the video without distinguishing between ICPs and non-ICPs, resulting in making an unnatural video. Meanwhile, their proposed privacy-protection system automatically discriminates ICPs from non-ICPs in real-time based on the spatial and temporal characteristics of the video, and then, only the non-ICPs can be localized and hidden. To protect privacy of persons captured in videos, Kitahara et al. proposed a system called Stealth Vision [19], which applies pixelization to persons. To locate persons in a mobile camera's frame, their system uses fixed cameras installed in the target environment. Meanwhile, by leveraging the power of machine learning, some interesting de-identification techniques have been introduced. For example, Yifan et al. [20] proposed a framework called Privacy-Protective-GAN that adapts generative adversarial network (GAN) for the face de-identification problem to ensure generating de-identified output with retained structure similarity according to a single input. In order to mitigate the privacy concern of the photographed persons in egocentric video, Dimiccoli et al. developed a convolutional neural networks (CNN)-based computational method for recognizing everyday human activities while mitigating privacy concerns by intentionally degrading the quality of egocentric photos [21]. Even though these de-identification techniques provide effective solutions for privacy-protection, the photographed person has no control over privacy-protection. This might lead to another aspect of privacy issue if the photographers intentionally use the photos for their own purpose without hiding the photographed person's identity.

2.2. Photographed Person-Site Methods

Photographed person-site methods can be classified into two groups: (1) cooperation between photographer and photographed person; and (2) photographed person-based.

In former group, some solutions require the photographed person to advertise his privacy preferences based on which the photographer's smart device will have appropriate actions (e.g., take no photos, blur subject's identity). The implementation of these methods mainly depends on: the way the photographed person express his intention/requirements in privacy-protection; and cooperation

methods between the photographer and the photographed person. Some methods require the photographed person to wear visible specialized tags. For examples, Pallas et al. in [12] introduced a set of four elementary privacy preferences represented by corresponding symbols—"Offlinetags" which are invisible and easily to be detected by detection algorithms. These privacy preferences are: "No photos", "blur me", "upload me", and "tag me". COIN [15] enables a photographed person to broadcast his privacy policies and empowers the photo service provider (or photographer) to exert the privacy protection policy. This approach is similar to the one in [16]. Some other methods require stronger cooperation between the photographer and the photographed person. For example, Li et al. presented PrivacyCamera [14], an application working on both the photographer's and the photographed person's mobile phone. Upon detecting a face, the app automatically sends notifications to nearby photographed people who are registered users of the application using short-range wireless communication. If the photographed person does not want to appear in the photo, they will indicate so to the photographer. His face will be blurred once the photographer confirms the appearance of the photographed person in the photo. However, this solution cannot completely solve the privacy problem if the photographer intentionally ignores the requests from the photographed people.

In the latter group, the photographed person proactively takes actions to protect his privacy. For examples, Yamada et al. [22] proposed a method to avoid unintended appearance in photos physically using a privacy visor that uses near-infrared light. That privacy visor's shape is like a pair of glasses that are equipped with near-infrared LEDs. The purpose of the use of near-infrared light is to saturate the charged-coupled device (CCD) sensor of digital cameras to distort the Haar-like features. Farinella et al. developed FacePET [23] to prevent the unintentional capture of facial images by distorting the region containing the face. This work is similar to the work in [22] since it makes use of glasses to emit light patterns designed to distort the Haar-like features which are used in some face detection algorithms. The noticeable difference is that the work in [22] used near-infrared light, while the visible light was used in [23]. However, these systems might not be effective for other types of face detection algorithm such as deep learning-based approaches. Additionally, these prototype glasses seem to be burdensome for users. In previous study [24], we proposed a method to identify photo-taking behavior using optical flow technique. To recognize such the behavior, the movements of arms and/or hands of the photographer were studied. However, the detection accuracy of this proposal was not so high, and it focused on only photo-taking behavior without considering other behavior with similar characteristics.

Our proposed solution in this study belongs to the photographed person-based category, allowing the photographed person to proactively make decisions in controlling over his facial privacy. More concretely, it helps him to detect the situation where someone is intentionally/unintentionally trying to take photos of him and has appropriate reaction to protect his facial privacy.

3. Proposed Approach

3.1. Photo-Taking Recognition Algorithm

In this section, the proposed algorithm for recognizing photo-taking behavior is presented. In a general scenario, we assume that a photographed person uses a wearable camera to monitors the surrounding environment. Then, based on the monitored video, the proposed algorithm will examine whether there is someone is trying to take the photos of the photographed person. Typically, our propped algorithm focuses on detecting photo-taking behavior and classifying it from net-surfing behavior. Note that, the net-surfing behavior is taken into account in the proposed algorithm due to its popularity. In fact, with smartphone, people can perform similar activities to net-surfing, for examples, texting, retrieving data, etc. In our definition, net-surfing includes web-surfing, social media (e.g., Facebook, Instagram) surfing, etc. Indeed, according to [25], Americans spend an average of 3 h a day on their smartphone for net-surfing compared to 41 min per day for texting. Typically, both photo-taking and net-surfing behaviors share common motions of moving arms which are defined

by the changes in arm's length and the angle from the view of the photographed person. Therefore, the transition of the arm's length and angle of the bending arm are crucial inputs for the detection mechanism. The proposed algorithm (Algorithm 1) is clearly described as follows:

Algorithm 1: Proposed Algorithm

Input: ***Monitored Video, DP Threshold***

Output: ***0*: Photo-Taking Behavior, *1*: Net-Surfing Behavior**

Initiate OpenPose
    Analyze the monitored video
    return *arm parts' skeleton information*
Calculate the arm's length and angle of bending arm
    (I) length of upper arm, (II) length of lower arm, (III) angle of bending arm
    return *(I)~(III) value*
Calculate DP scores
DP matching (reference data: photo-taking behavior)
return *DP score*
If DP score ≤ threshold **Then**
Judged as photo-taking behavior
return *0: photo-taking behavior*
Else
Judged as net-surfing behavior
    return *1: net-surfing behavior*
End if

3.2. How Does the Proposed Approach Work in Reality?

Figure 1 depicts an assumed scenario for the implementation of our proposal. Accordingly, a photographed person uses a wearable camera to monitor the surrounding environment all the time or in specific event that he wants to protect his facial privacy. The monitored video as input data is continuously fed into the detection algorithm which runs on his mobile phone or cloud-based device for further analysis. If the photo-taking behavior is detected, a vibration signal as the output is activated to notify the photographed person. This allows him to perform some types of physical actions. For example, he simply leaves the shared space or asks the photographers to stop taking photos. In practice, this approach can be used by a person who wishes to proactively protect his privacy from the violation of an individual. It means that only a suspected individual is captured in the monitoring video. However, using a normal wearable camera possibly leads to the facial privacy issue of other people who are captured in the video unintentionally. In reality, we believe that using thermographic camera, particularly long-wave infrared camera, is a potential solution to deal with this problem. Typically, long-wave infrared imagery is independent of illumination since thermal infrared sensors operating at particular wavelength bands measure heat energy emitted and not the light reflected from the objects [26]. As mentioned in Section 2, the photographed person is assumed to be the one who has strong motivation in securing his facial privacy. Thus, by wearing such a thermographic camera, the photographed person can proactively control over privacy-protection without violating the others' facial privacy. However, using a thermographic camera probably poses challenges in recognizing photo-taking behavior in thermal video. This will be considered in our future study.

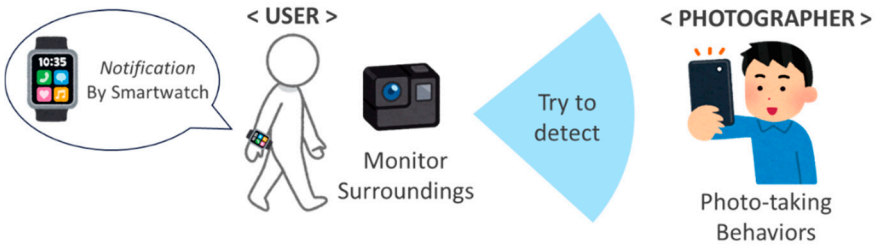


Figure 1. A scenario of photo-taking behavior detection and its notification.

3.3. Extract Human Skeleton Information

As stated in Section 3.1, human skeleton information is the key input of our proposed algorithm. Typically, both photo-taking and net-surfing behaviors share similar motions of moving arms. Thus, the skeleton information of the arm's length and the transition of angle of bending arms from the view of the photographed person is focused on. Such information can be obtained by monitoring: (I) upper arm, (II) lower arm, and (III) the angle of the bending arms.

In order to obtain the skeleton information, OpenPose [17] (an open-source tool) is used. By leveraging this tool, the human skeleton information can be extracted in real-time from two-dimensional video frames. Figure 2 illustrates an example of skeleton information extracted from OpenPose. Accordingly, there are 25 points connected by joint parts, establishing "BODY_25" human skeleton estimation model. In practice, OpenPose allows the joint coordinates in each frame to be obtained and stored in json files. Thus, the skeleton data is formed as [x, y, *confidence score*].

In the following subsections, the details of each step in the proposed algorithm will be explained.

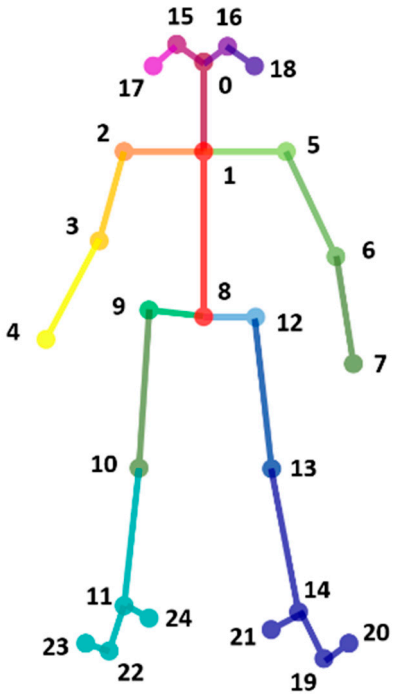


Figure 2. "BODY_25" human skeleton estimation model.

Here, the x and y are coordinates indicating body part locations in the input image. The *confidence score* indicates the accuracy of the coordinates calculated by OpenPose tool. As we assumed earlier, there are some potential differences in the arm's length and the angle among photo-taking and net-surfing behaviors. Therefore, we only focus on these parts which are numerically calculated from joints' information. Accordingly, the joints: "2, 3, 4, 5, 6, 7" in "BODY_25" model (depicted in Figure 2)

are used for further behavior analysis. The points and joints of the utilized arm parts are visualized in Figure 3. Table 1 provides brief information of joint positions of the arm parts and the according expressions used in this paper.

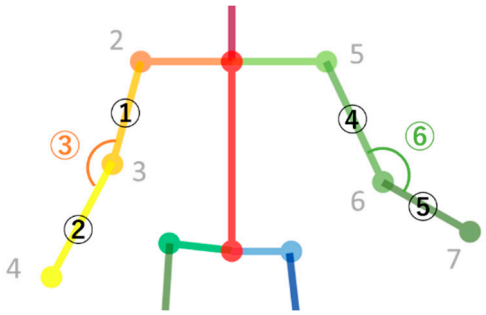


Figure 3. Focusing parts in proposed approach.

Table 1. Correspondence between joint position and body part.

	Index Factors	Joint Position (Keypoint)	Arm Parts	Index Number in Figure 3	Expression in This Paper
Right	I	2–3	Right Upper arm	①	Length-23
	II	3–4	Right Lower arm	②	Length-34
	III	2–3–4	Angle of the bending right arm	③	Angle-234
Left	I	5–6	Left Upper arm	④	Length-56
	II	6–7	Left Lower arm	⑤	Length-67
	III	5–6–7	Angle of the bending left arm	⑥	Angle-567

In proposed approach, the numerical values of the arm length and angle are determined by using the distance between two points and inner product of coordinates, which are obtained from OpenPose. The detailed calculations are presented in Figure 4.

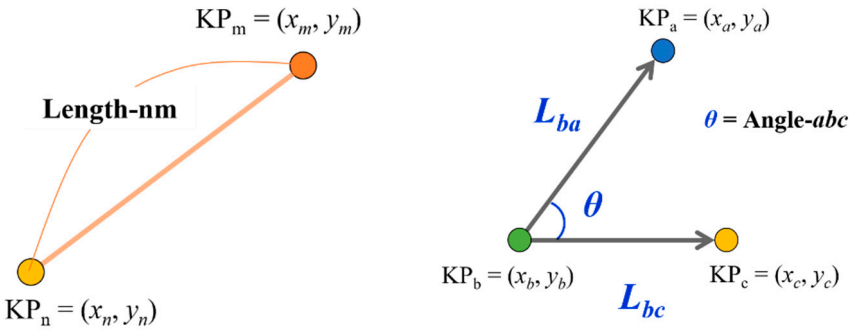


Figure 4. Calculation of the arm's length and angle of the bending arm. (**a**) Calculation of the arm's length from two coordinates by using the distance between two points KP_n and KP_m. This method is applied to calculate the length of ①, ②, ④, ⑤ in Table 1; (**b**) Calculation of the angle of the bending arm from three coordinates which are indexed by ③, ⑥ in Table 1 by using the inner product.

- Calculation of the arm's length (I) and (II)

According to Figure 4a, a certain joint position (keypoint) denoted as KP_p is presented as

$$\mathrm{KP}_p = \left(x_p, y_p\right) \tag{1}$$

where, the p indicates a joint position (keypoint) number.

The arm length can be calculated by the equation

$$\text{Length-nm} = \sqrt{(x_m - x_n)^2 + (y_m - y_n)^2} \tag{2}$$

In this case, "Length-nm" stands for arm's length between joint position number of n and m. By using Equation (2), the length of the right and left upper/lower arm (indexed by ①, ②, ④, ⑤) can be properly determined.

- Calculation of the angle of bending arm (III)

The angle of bending arm θ is formed by $\mathrm{KP_a} = (x_a, y_a)$, $\mathrm{KP_b} = (x_b, y_b)$, $\mathrm{KP_c} = \left(x_c, \mathrm{y_c}\right)$ as shown in Figure 4b. Accordingly, the procedure to calculate the angle from these points is as follows: First, it is required to perform vectorization where the vectorization $\boldsymbol{L}_{pq}$ is expressed as

$$\boldsymbol{L_{ba}} = \begin{pmatrix} x_a - x_b \\ y_a - y_b \end{pmatrix} = \begin{pmatrix} ba_1 \\ ba_2 \end{pmatrix}, \boldsymbol{L_{bc}} = \begin{pmatrix} x_c - x_b \\ y_c - y_b \end{pmatrix} = \begin{pmatrix} bc_1 \\ bc_2 \end{pmatrix} \tag{3}$$

Then, by using these vectors, the angle can be calculated as

$$\theta = \cos^{-1}\left(\frac{\boldsymbol{L_{ba}} \cdot \boldsymbol{L_{bc}}}{|\boldsymbol{L_{ba}}| \cdot |\boldsymbol{L_{bc}}|}\right) \tag{4}$$

where, $0 \leq \theta \leq \pi$.

Therefore, based on this procedure, the angles of the bending right and left arms (indexed by ③, ⑥) can be calculated.

3.4. Threshold for Recognizing Photo-Taking Behavior

In this subsection, we present the determination of DP threshold which plays an important role in deciding whether a series of human hand movements form photo-taking behavior or not. Typically, the threshold value can be obtained from the point of equal error rate (ERR), where the false acceptance rate (FAR) and false rejection rate (FRR) curves meet. The following parts will provide brief explanations of DP matching, FAR, FRR, and ERR using in our proposed approach.

3.4.1. DP Matching

DP matching is a pattern matching technique which evaluates the similarity between two sequenced data. For examples, given two patterns of sequenced data (X and Y) which are expressed as

$$\mathrm{X} = x_1, x_2, \cdots, x_i, \cdots, x_I \tag{5}$$

$$\mathrm{Y} = y_1, y_2, \cdots, y_j, \cdots, y_J \tag{6}$$

where X and Y represent a sequenced input data and the reference sequenced data, respectively.

Meanwhile, I and J indicate the number of data points of X and Y, respectively. Let $d\left(x_i, y_j\right)$ express the distance between the elements: X and Y. It will be transformed from x-y coordinate space to i-j coordinate space as

$$l(i, j) = d(x_i, y_i) = \left|x_i - y_j\right| \tag{7}$$

In addition, the accumulated distance is expressed by $g(i, j)$ in i-j coordinates space. $g(i, j)$ basically can be obtained by calculating the minimum DP path in an optimal distance problem. Figure 5 shows the weight for calculating optimal distance as the definition of DP path. According to Figure 5, the dissimilarity $g(i, j)$ can be defined by

$$g(i,\ j) = \min\left\{ \begin{array}{ll} g(i-1,\ j) + d(i,\ j) & : (a) \\ g(i-1,\ j-1) + 2d(i,\ j) & : (b) \\ g(i,\ j-1) + d(i,\ j) & : (c) \end{array} \right\} \tag{8}$$

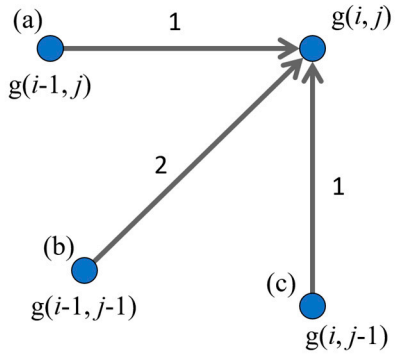


Figure 5. Definition of DP path. To calculate the accumulated distance, (**a**) to (**c**) indicates a pattern of distance in i-j coordinates space. Each number shown in (**a**) to (**c**) expresses the weighted score for calculating the distance by the following Equation (8).

Finally, DP matching score between X and Y is obtained by normalizing g (i, j) with the number of each data points as shown in Equation (9).

$$\text{DP score} = \frac{g(I,J)}{I+J} \tag{9}$$

The smaller the DP score is, the higher the similarity between the two data is. In this study, the use of DP score is two-fold. First, in the training phase, DP scores are calculated from matching process upon human skeleton information in training dataset. The DP scores are then used to generate the values of FAR and FRR. Theoretically, the curves which represent FAR and FRR are expected to intersect at a point of EER value. As a result, the DP score which reflects EER value is eventually determined as the DP threshold. Second, in the testing phase, DP scores which are calculated from the matching process are compared with the determined DP threshold to conclude whether the input monitored hand movements characterize photo-taking behavior.

3.4.2. FAR and FRR and DP Threshold Determination

In this part, we provide the explanations on how we define FAR and FRR for the determination of EER value which is referred to threshold value. The terms of FAR, FRR, and ERR are common in the topics of biometric security systems [27]. False acceptance rate (FAR) is defined by the percentage of identification instances in which unauthorized persons are incorrectly accepted (this is also known as false match rate.) False rejection rate (FRR) refers to the percentage of identification instances in which authorized persons are incorrectly rejected (this is also known as false non-match rate). In other words, FAR implies how high your system's security level is, whereas FRR reflects the level of comfortableness of the users. In order to evaluate the operating performance of a security system, the equal error rate (EER), which is also known as the crossover error rate (CER), must be taken into account. It means that the system has parameters that can be turned to adjust FAR and FRR to the point where both of them are equal. Importantly, the smaller the ERR is, the better the performance is. In this study, the FAR is defined as the error rate in which the net-surfing behavior is recognized as photo-taking

behavior, whereas FRR refers to the non-detection rate of photo-taking behavior. Accordingly, the obtained EER is illustrated in Figure 6 as the intersection point of the curves of FAR and FRR. The DP score corresponding to this ERR value will be desirable threshold.

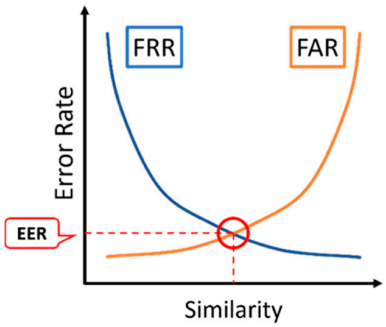


Figure 6. Ideal FRR–FAR curves image and EER crossing point.

4. Evaluations

In this section, we present three major tasks regarding the evaluation of the proposed approach: (1) collecting necessary datasets; (2) determining DP threshold score; (3) evaluating performance of the proposed approach.

4.1. Dataset Collection

4.1.1. Experimental Setup and Data Acquisition

In this part, we present the experiments which were conducted to obtain the necessary datasets for the study. The experimental setup is described in Figure 7 where a user (assumed as a photographer) is taking action of either taking the photo or net-surfing using a smartphone, while the other plays the role of the photographed person. In our experiment, the photographer's behavior was continuously recorded by another smartphone worn by the photographed person. Note that, the recorded videos were taken from the right side of all participants as shown in Figure 7. Therefore, we hypothesize that the information of the movements of the participants' right arm will significantly contribute to detection purpose. All the videos were taken by Apple iPhone5s with a frame rate of 30 *fps*. There were 15 subjects participating in this experiment. Openpose was then used to automatically extract skeleton information of the participants in the videos, forming our datasets. As mentioned earlier, we only focus on three major parts: ((I) upper arm, (II) lower arm, and (III) the angle of the bending arms.

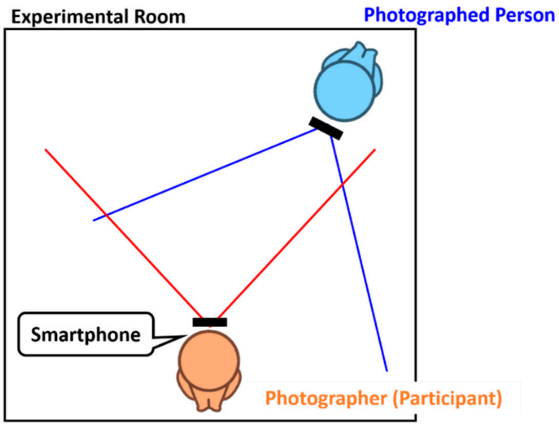
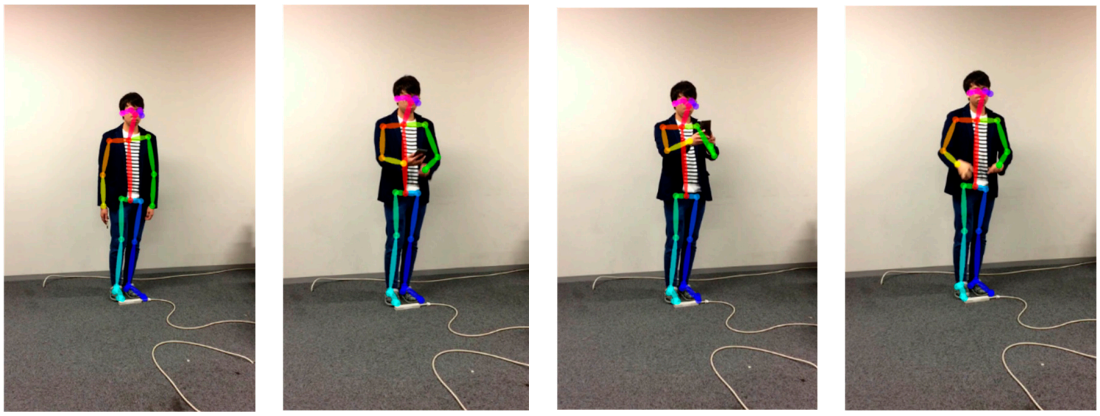


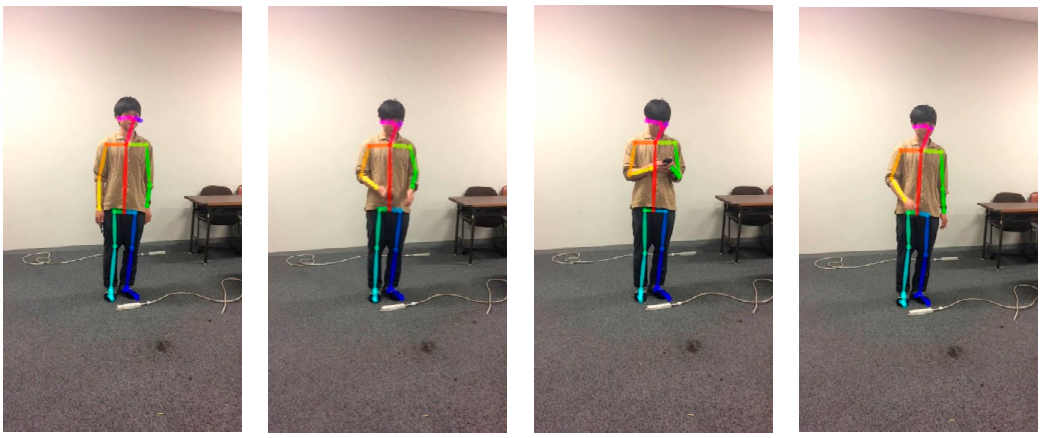
Figure 7. Experimental environment where photographed person records the video of the photographer, while the photographer is performing either photo-taking behavior or net-surfing behavior with a smartphone.

Figures 8 and 9 partly illustrate the visual outputs obtained from OpenPose for photo-taking and net-surfing behaviors, respectively. Specifically, (a–d) in these two figures shows the example frame expressing: initial position of subject's arms, moment when the behavior starts, moment during the behavior, and moment when behavior ends, respectively. The OpenPose data obtained from the participant who performed a photo-taking behavior was denoted as "*Px*". Meanwhile, the data obtained from the participant who performed net-surfing behavior was denoted as "*Nx*". We divided the obtained dataset into two sub-datasets, namely, dataset1 and dataset2 with the ratio of 50:50. The details are tabulated in Table 2. Accordingly, the dataset1 which consists of *P*1 to *P*7 and *N*1 to *N*3 was used for determining DP threshold. Note that, since this sub-dataset was small, cross-validation was applied beforehand. On the other hand, the dataset2 which consists of P8 to P15 and N4 to N6 was used for evaluating the performance of the proposed approach.



(**a**) (**b**) (**c**) (**d**)

Figure 8. Visual skeleton information of photo-taking behavior extracted from OpenPose. (**a**) Initial position of subject's arm; (**b**) before subject start taking photo; (**c**) when subject is taking photo (during photo-taking behavior); (**d**) when subject finishes photo-taking behavior.



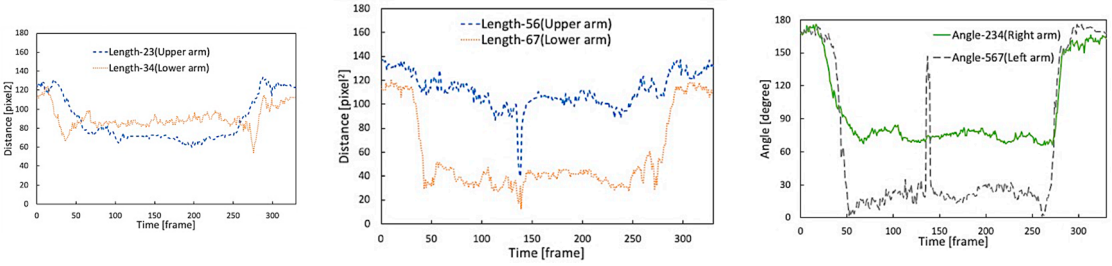
(**a**) (**b**) (**c**) (**d**)

Figure 9. Visual skeleton information of net-surfing behavior. (**a**) Initial position of subject's arm; (**b**) before subject starts the net-surfing behavior; (**c**) when subject is surfing the Internet on smartphone (during net-surfing behavior); (**d**) when subject finishes net-surfing behavior.

Table 2. Obtained datasets from experiments.

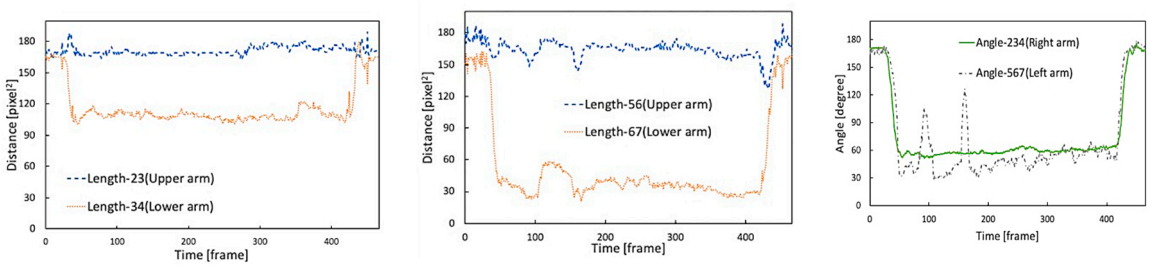
	Photo-Taking Behavior	**Net-Surfing Behavior**
Dataset1	*P*1, *P*2, . . . , *P*7	*N*1, *N*2, *N*3
Dataset2	*P*8, *P*9, . . . , *P*15	*N*4, *N*5, *N*6

We visualize some sample data of skeleton information obtained from OpenPose to demonstrate the transitions of three considered human parts (I, II, and III). Figure 10 illustrates the transitions drawn from $P1$ which is the data of the photo-taking behavior performed by subject 1. Meanwhile, the transitions depicted in Figure 11 plotted from $N1$ (net-surfing behavior performed by subject 1). Note that, for each subject, six joint components (joint-23, joint-34, joint-56, joint-67, angle-234, and angle-567) in total were considered for further analysis. Qualitatively, according to Figure 10b,c and Figure 11b,c, there are no significant differences between the photo-taking and net-surfing behaviors. Meanwhile, the differences among these behaviors are obvious when observing Figures 10a and 11a.



(**a**) (**b**) (**c**)

Figure 10. Arm's length and angle of bending arms of subject 1 when taking photo. (**a**) Right upper and right lower arm's lengths; (**b**) left upper and left lower arm's lengths; (**c**) angles of right bending and left bending arms. In (**a**,**b**), vertical axis represents distance between joints (length) in pixel^2. The horizontal axis indicates frame number it means time (in frames). In (**c**), the vertical axis represent angle in degree. The horizontal axis indicates frame number it means time (in frames).



(**a**) (**b**) (**c**)

Figure 11. Arm's length and angle of bending arms of subject 1 when performing net-surfing. (**a**) Right upper and right lower arm's lengths; (**b**) left upper and left lower arm's lengths; (**c**) angles of right bending and left bending arms.

4.1.2. Data Pre-Processing

In order to eliminate non-detection and misdetection caused by OpenPose, data pre-processing is needed. Figure 12 illustrates an example of non-detection where the numerical values of the joint components cannot be calculated. In such a situation, the coordinates of the undetected joint in a specific frame are predicted by performing interpolation using the information of preceding frame and succeeding frame. The misdetection, on the other hand, introduces a sudden change in the numerical values of the investigating joint components as shown in Figure 10b,c. Figure 13 shows an example (Example 1) of misdetection in both captured video frame and visual graph. Note that Figure 13a,b are the same graphs as Figure 10b,c, respectively. In these graphs, the 139th frame in which a joint was mis-detected, was emphasized by a yellow rectangle. Figure 13c illustrates 135th, 139th, and 142nd frame extracted from video.

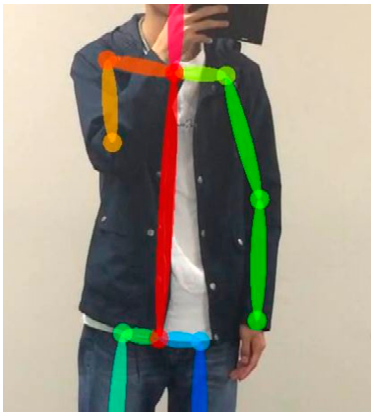


Figure 12. Example of non-detection frame.

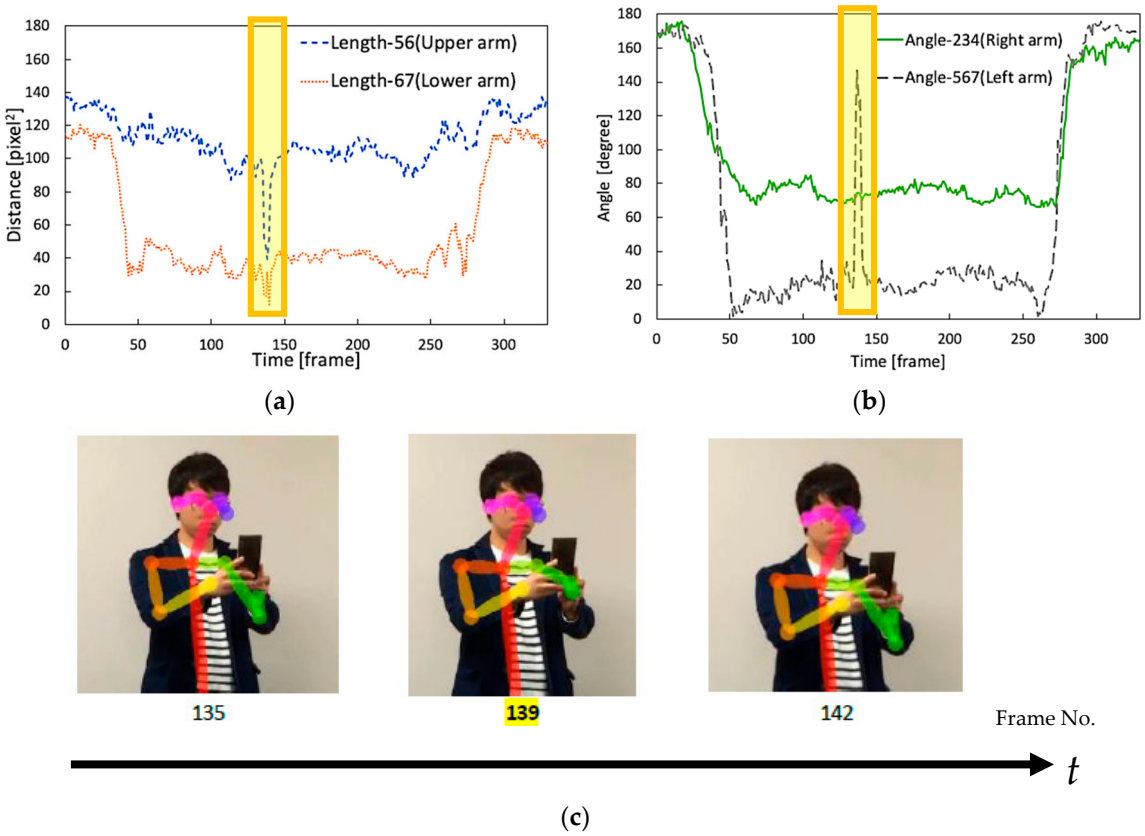


Figure 13. Example 1 of misdetection generated by OpenPose (taken from P1). (**a**) Left upper and left lower arm's lengths; (**b**) angles of right bending and left bending arms; (**c**) Example frame (135th, 139th, and 142nd).

Figure 14 shows another example of misdetection (Example 2). In Figure 14a, the ideal detection result is drawn in white color. According to this figure, the result obtained from OpenPose was different from the ideally estimated result. Figure 14b shows some numerical values calculated from the coordinates of joints of the subject. The values fluctuate several times due to such a misdetection. In fact, both non-detection and misdetection create noise in the obtained data. Thus, in order to remove the noise, low pass filter (LPF) was used. Thereby, we first extracted the frequency components from the obtained data by using fast Fourier transform (FFT), then the cut-off frequency was assigned. In this case, a general Butterworth filter was considered as the filter, whereas the desirable cut-off frequency of the LPF was determined as 40 Hz from the preliminary experiment. Figure 15 visualizes the processed

photo-taking behavior data of subject 1, after performing LPF. Vertical axes indicate the am length in pixel^2 and the angle in degrees. The horizontal axis indicates the frame corresponding to time. The line in red in each graph indicates the result of the filtering. Qualitatively, it is obvious that the transitions of joint components become smoother, allowing us to apply data to further analysis steps.

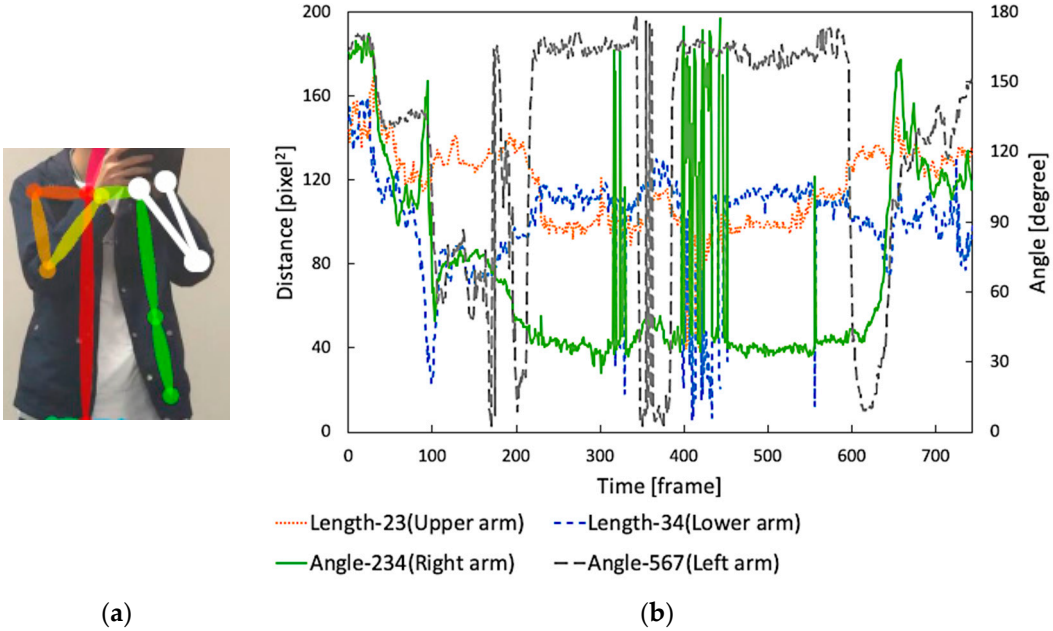


Figure 14. Example 2 of misdetection generated by OpenPose (taken from P5). (**a**) Misdetection frame (white line presents an expected detection result); (**b**) result of the right upper and lower arms' lengths and the angle of right/left bending arms. In (**b**), first vertical axis indicates distance between joints (length) (pixel^2). Second vertical axis indicates angle (degree). Horizontal axis indicates frame number corresponding to time (frames).

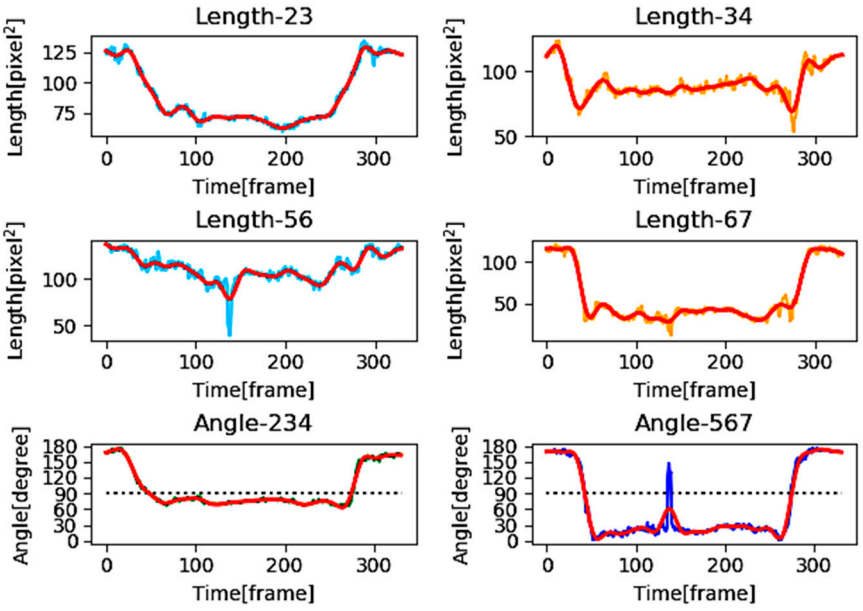


Figure 15. Sample P1 data after applying LPF.

4.2. Determination of DP Threshold

In order to determine the DP threshold (denoted as Th_{DP}), we first performed DP matching on dataset 1. Since the data volume was small, we used cross-validation. It means that the data of a participant was used as the reference data for the rest of data in DP matching.

Table 3 tabulates an example of DP matching scores of each joint component when data *P*1 was considered as the reference data. Since the monitored videos were taken from the right side of the photographer rather than from their front, thus, it was expected that not all the joint components are equally important. Therefore, to select the appropriate components for further investigation, the average DP scores with error bars of all joint components across participants are obtained in Table 3 and plotted in Figure 16. Accordingly, the right upper arm (length-23) shows the biggest difference between photo-taking and net-surfing behaviors. Meanwhile, there is not so much difference between these two behaviors can be found in other components. Note that the cases where monitored videos are captured in other sides of the photographer will be considered in future work. Therefore, in this study, we only focus on length-23 in determining DP threshold and in performance evaluation of our proposed approach.

Table 3. DP scores obtained from the case where P1 was used as the reference data (Reference data: *P*1, Input data: *P*2, *P*3, . . . , *P*7 and *N*1, . . . , *N*3).

		Length-23	Length-34	Length-56	Length-67	Angle-234	Angle-567
Photo-taking behavior	*P*2	1.69	2.41	5.84	6.23	9.11	4.33
	*P*3	6.54	14.57	3.41	4.62	8.98	2.73
	*P*4	3.39	3.36	4.76	4.38	2.28	1.65
	*P*5	3.92	6.40	10.28	17.9	14.53	21.34
	*P*6	30.64	10.93	16.87	5.9	4.44	3.91
	*P*7	10.89	7.21	13.2	4.4	3.91	4.80
	Average	9.51	7.48	9.06	7.24	7.21	6.46
	S.T.	9.89	4.20	4.84	4.82	4.15	6.74
Net-surfing behavior	*N*1	58.88	8.50	28.6	2.04	3.48	6.31
	*N*2	21.48	3.18	11.17	3.26	2.73	8.20
	*N*3	29.05	3.27	9.5	1.25	1.39	6.94
	Average	36.47	4.98	16.42	2.18	2.54	7.15
	S.T.	16.15	2.49	8.64	0.83	0.86	0.78

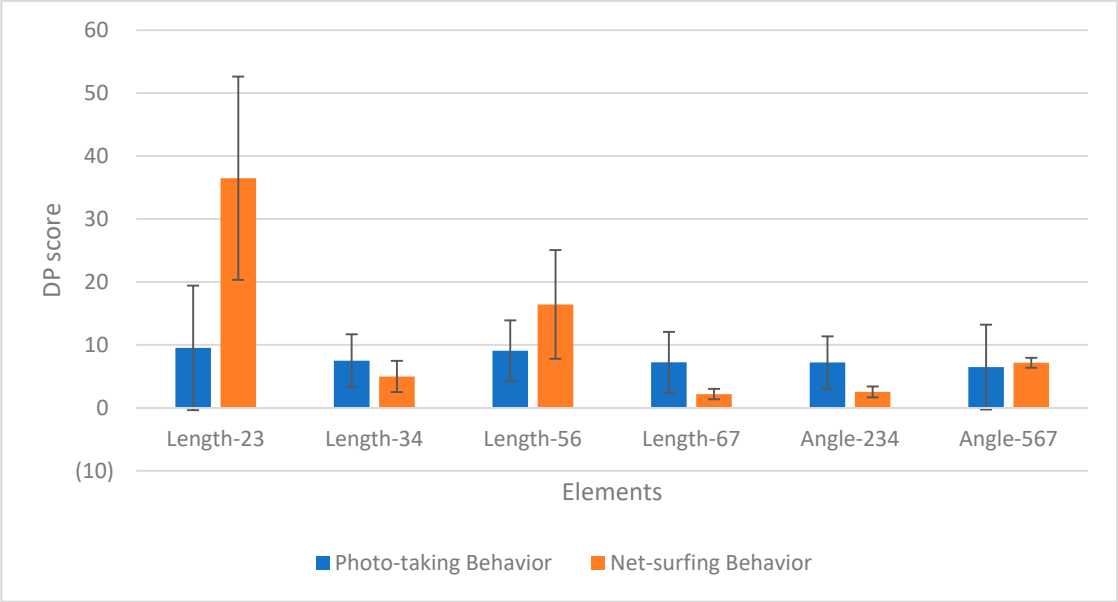


Figure 16. Average DP score for each behavior obtained from the results in Table 3 (reference data: P1).

The next step is to calculate the values of FAR, FRR, and EER. In our study, FAR and FRR can be calculated by using the equations

$$\text{FAR} = \frac{\text{Number of mis-recognition as Photo-taking behavior}}{\text{Number of all Net-surfing behavior}} \quad (10)$$

$$\text{FRR} = 1 - \frac{\text{Number of correct recognition as Photo-taking behavior}}{\text{Number of all Photo-taking behavior}} \quad (11)$$

The calculated FAR and FRR were plotted along with the so-called 'assigned DP threshold' which was ranged from 0 to 35 with an increasing step of 2.5. If the DP score of each joint component in the reference data of particular participant (as an example of P1 in Table 3) is less than 'assigned DP threshold', it is determined that this participant performed a photo-taking behavior. Oppositely, a net-surfing behavior was determined when the DP score is more than 'assigned DP threshold'. Similarly, with the data of seven participants who performed photo-taking behavior, we could obtain seven graphs depicting the visual values of FAR and FRR of these participants. Figure 17 depicts two of seven graphs of the reference data P1 and P6. Thereby, seven values of EER were easily extracted. As mentioned earlier, EER is the intersection point of FAR and FRR where both of those values are equal. To determine DP threshold, a general value of EER across all the participants must be obtained. Figure 18 depicts all of seven EER values. Accordingly, most of EER values which are less than 0.2 represent photo-taking behavior. On the other hand, the data with an obviously high error rate was recognized as outliers and must be removed. Thereby, we excluded two data points: *P6* and *P7* with EER values higher than 0.4. The average value of eligible EER was then calculated as about 0.17. Therefore, in accordance with the average value of 0.17 of EER, the DP threshold for our proposed approach was determined as $\boldsymbol{Th_{DP}}$ = 15.9. In the next subsection, the performance of our approach is evaluated by using this DP threshold and dataset 2.

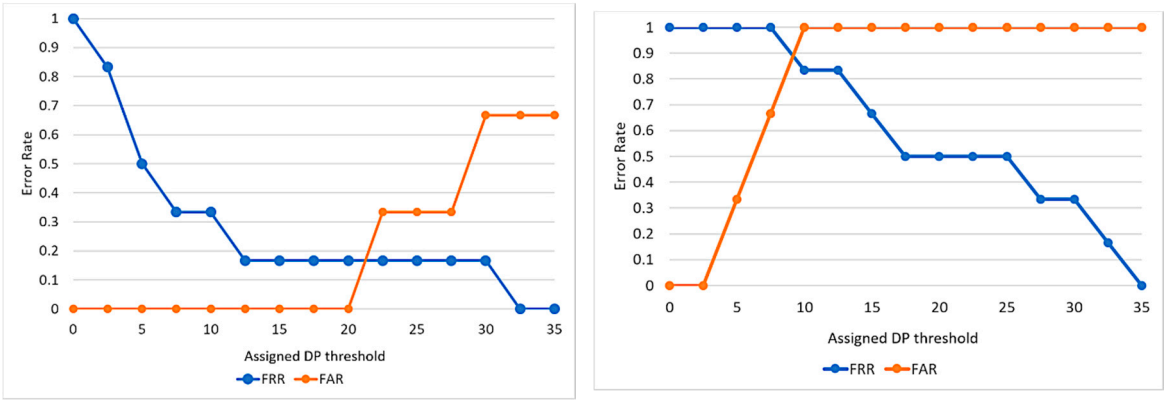


Figure 17. Examples of FRR–FAR curves in the cases where: (**a**) reference data is P1; (**b**) reference data is P6. In each graph, the horizontal axis indicates assigned DP thresholds. The vertical axis indicates error rate.

4.3. Performance Evaluation of the Proposed Approach

In order to evaluate the proposed approach, DP matching was performed on the dataset2 using the dataset1 as the reference data. We removed *P6* and *P7* from the reference data because they were considered as outliers with very high EER values (shown in Figure 18). The obtained DP scores were then compared with the DP threshold to identify photo-taking behavior. The DP matching results are presented in Table 4. Note that, as explained in Section 4.2, we only focused on the joint of length-23, thus, Table 4 provides the results of DP matching with respect to this joint. In addition, the detection decision is expressed in cell colors. Accordingly, the yellow cells indicate that the behaviors were decided as photo-taking behavior whose DP scores are less than the threshold ($\boldsymbol{Th_{DP}}$= 15.9). In other words, the yellow cells show the correct detection using DP score. In addition, the light gray cells

indicate the correct decision in net-surfing behavior since those DP scores are higher than the threshold. On the other hand, the gray cells with white numbers indicate the incorrect decision.

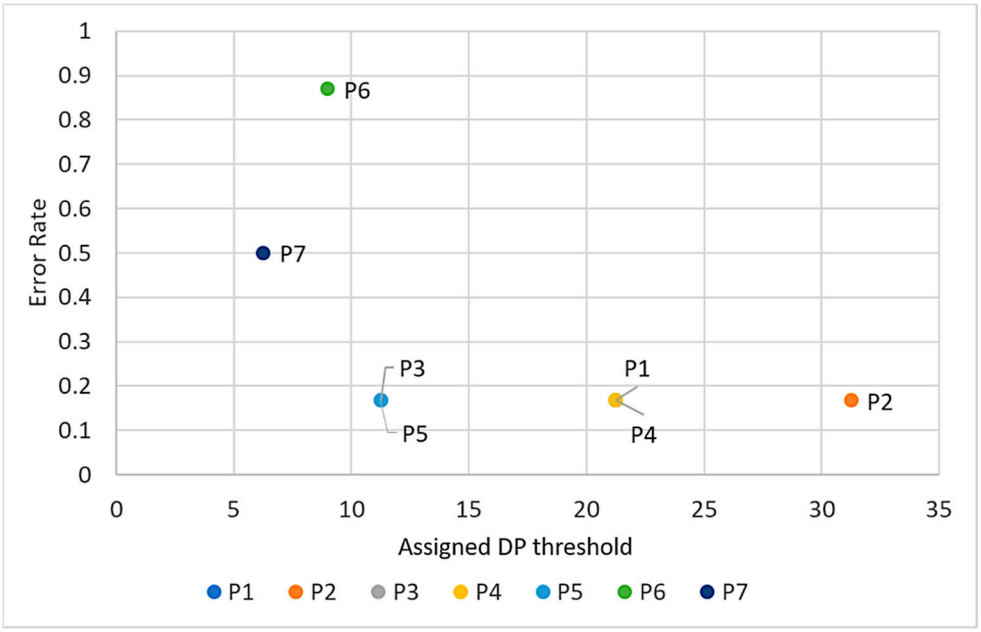


Figure 18. EER distribution obtained from all FRR–FAR curves by cross-validation for right upper arms (f_c = 40 Hz). The horizontal axis indicates the assigned DP thresholds. The vertical axis indicates the error rate. The legend shows photo-taking behavior data used as reference data.

Table 4. DP matching result of dataset2 when using dataset1 as the reference data (Reference data: dataset1 except outlier [*P*1, . . . , *P*5], Input data: dataset2 [*P*8, . . . , *P*15 and *N*3, . . . , *N*6]).

			Reference Data: Photo-Taking Behavior (Dataset1)				
			length-23(Right Upper arm)				
			***P*1**	***P*2**	***P*3**	***P*4**	***P*5**
Input Data (Dataset2)	Photo-taking behavior	***P*8**	8.10	7.89	5.69	8.27	3.40
		***P*9**	4.60	6.68	3.84	4.04	5.91
		***P*10**	0.93	2.84	5.17	2.96	3.28
		***P*11**	8.44	9.49	2.07	3.33	3.22
		***P*12**	1.14	2.30	5.74	2.81	2.91
		***P*13**	5.61	6.00	1.42	1.93	3.65
		***P*14**	21.10	28.43	7.01	17.08	10.18
		***P*15**	13.59	13.74	10.18	13.51	3.14
	Net-surfing behavior	***N*4**	35.44	43.55	15.57	32.08	22.06
		***N*5**	24.26	33.34	10.02	22.68	12.71
		***N*6**	14.62	19.47	11.26	17.28	6.67

In order to obtain detection accuracy, Equation (12) was utilized. In overall, by using proposed approach, we achieve 92.5% of accuracy in recognizing photo-taking behavior. Looking at the result in detail, the detection accuracy when particular reference data is applied, is not the same. It might be varied, but not introducing so big difference. For example, if the reference data *P*1, *P*2, or *P*4 are used, the detection accuracy of the photo-taking behavior is 87.5%, whereas, when either the reference data *P*3 or *P*5 is used, the accuracy is 100%. Such variations might come from individual differences in term of photo-taking posture. On the other hand, based on Equation (12), the detection accuracy of net-surfing behavior is calculated as 60%, which is not so as high as we expected. We speculate that some subjects, especially subject 3, might turn his body while performing net-surfing behavior;

thus, the arm lengths and the angle of bending arms subsequently changed. Additionally, too small dataset of net-surfing behavior could be the reason. In the future works, the proposed method will be evaluated with larger datasets. It is worth noting that the high accuracy in the detection of photo-taking behavior and sufficient accuracy in the detection of net-surfing behavior confirm the reliability of determined DP threshold.

$$\text{Accuracy} = \frac{number\ of\ correct\ detection}{total\ number\ of\ DP\ scores} \times\ 100\% \tag{12}$$

4.4. Overal Discussion and Future Works

Overall, this study provides a potential approach to privacy-protection of photographed person. Most of the state-of-the-art methods have been proposed for working at photographer site. However, this does not provide any opportunity to the photographed person to control his facial privacy. This becomes a serious problem when the photographers ignore the photographed person's facial privacy preferences, and intentionally take photos and share the photos on social networks or use the photos without hiding the photographed person's identity for some purposes. Our proposed method, on the other hand, allows the photographed person to make proactive decisions based on his facial privacy preferences. This method potentially works even in the case of protecting the facial privacy of a 'bystander' who was not intentionally captured by the photographers. Once the photo-taking behavior is detected, the photographed person will receive notifications from his smart phone. This enables him to flexibly perform physical actions such as leaving the shared space or asking the photographers to stop taking photos. However, there are some concerns which would be considered in the future works. First, to apply this method, the photographed person must record videos of the photographers all the time, this leads to the facial privacy issue for such persons and other unwanted photographing people. In this study, we tried to mitigate this problem by making an assumption that only the arms of a photographer were recorded. However, this assumption might not be sufficient to guarantee that the proposed method can completely solve the problem since the face part is probably captured in the video. Therefore, we believe that using a thermographic camera, which can produce long-wave infrared images, is more realistic approach. Indeed, the facial information of bystanders will not be easily recognized from such long-wave infrared images [26]. Second, there are several real scenarios that have not been considered in this study. In practice, the photographed person probably has to protect his facial privacy in a crowded place. In fact, OpenPose can generate the skeleton information of many subjects in a video, providing potential of recognizing photo-taking behavior of more than one subject in real-time. For the feasibility of this study, we do not take into account the "crowded case", instead, we assume that the photographed person can use our method in situations/events if he is aware that his facial privacy can be violated by an individual. In addition, the cases where recorded videos are taken from different sites of the photographer should be considered. Third, computational cost of real-time processing is also a considerable challenge. However, by leveraging the power of fog/edge computing in addition to offloading techniques [28], the limitations such as the lack of computational power, restrictions in storage capacity, and processing delay will be potentially solved.

5. Conclusions

We have presented the proposed approach to prevent the unintended appearance in photos by recognizing photo-taking behavior performed by photographer. In this study, we argue that it is difficult to differentiate photo-taking behavior and net-surfing behavior because they are formed by very similar motions of moving arms. Thus, to correctly recognize photo-taking behavior, human skeleton information was proposed to be analyzed. The analysis was based on DP matching technique. In a real scenario, our proposed approach allows a photographed person proactively to protect his facial privacy, especially in the case where a particular photographer is intentionally capturing them in the photos for some purpose. In the future, we will further investigate various aspects related to real-time processing problems as well as scenarios where the photographed person is in a crowded place.

Author Contributions: Conceptualization, Y.K.; writing—original draft preparation, Y.K.; writing—review and editing, P.X.T. and E.K.; supervision, P.X.T. and E.K. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest.

References

1. We Are Social, Digital in 2019. Available online: https://wearesocial.com/global-digital-report-2019 (accessed on 31 January 2020).
2. Ministry of Internal Affairs and Communications, White Paper Information and Communications in Japan," Part2, Figure 5-2-1-1, Transitions in Household Ownership Rates for ICT Devices. p. 65. Available online: https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/whitepaper/2018/index.html (accessed on 2 October 2020).
3. Ministry of Internal Affairs and Communications, White Paper Information and Communications in Japan, Part1, Figure 4-2-1-2, Terminals Connected to the Internet. p. 42. Available online: https://www.soumu.go.jp/main_sosiki/joho_tsusin/eng/whitepaper/2018/index.html (accessed on 2 October 2020).
4. McCarthy, B.C.; Feis, A. Rogue NYPD Cops Are Using Facial Recognition App Clearview. Available online: https://nypost.com/2020/01/23/rogue-nypd-cops-are-using-sketchy-facial-recognition-app-clearview/ (accessed on 15 July 2020).
5. Frome, A.; Cheung, G.; Abdulkader, A.; Zennaro, M.; Wu, B.; Bissacco, A.; Adam, H.; Neven, H.; Vincent, L. Large-scale privacy protection in Google Street View. In Proceedings of the IEEE Computer Vision (ICCV2009), Kyoto, Japan, 29 September–2 October 2009; pp. 2373–2380.
6. Google Street View, Google-Contributed Street View Imaginary Policy. Available online: https://www.google.com/streetview/policy/#blurring-policy (accessed on 17 July 2020).
7. Koyama, T.; Nakashima, Y.; Babaguchi, N. Real-time privacy protection system for social videos using intentionally-captured persons detection. In Proceedings of the 2013 IEEE International Conference on Multimedia and Expo (ICME), San Jose, CA, USA, 15–19 July 2013; Volume 6, pp. 1–6.
8. Guardian Project, ObscuraCam: Secure Smart Camera. Available online: https://guardianproject.info/apps/obscuracam/ (accessed on 17 July 2020).
9. Hasan, R.; Crandall, D.; Fritz, M.; Kapadia, A. Automatically Detecting Bystanders in Photos to Reduce Privacy Risks. In Proceedings of the 2020 IEEE Symposium on Security and Privacy (SP), San Francisco, CA, USA, 18–20 May 2020; pp. 318–335. [CrossRef]
10. Bo, C.; Guobin, S.; Jie, L.; Xiang-Yang, L.; YongGuang, Z.; Feng, Z. Privacy. tag: Privacy concern expressed and respected. In Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems, Memphis, TN, USA, 23–26 November 2014; pp. 163–176.
11. Li, S.; Deng, W.; Du, J. Reliable Crowdsourcing and Deep Locality-Preserving Learning for Expression Recognition in the Wild. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 2584–2593.
12. Pallas, F.; Ulbricht, M.-R.; Jaume-Palasí, L.; Höppner, U. Offlinetags: A novel privacy approach to online photo sharing. In Proceedings of the Extended Abstracts of the 32nd Annual ACM Conference on Human Factors in Computing Systems—CHI EA '14. Association for Computing Machinery, New York, NY, USA, 26 April–1 May 2014; pp. 2179–2184. [CrossRef]
13. Shu, J.; Rui, Z.; Pan, H. Cardea: Context-aware visual privacy protection for photo taking and sharing. In Proceedings of the 9th ACM Multimedia Systems, Amsterdam, The Netherlands, 12–15 June 2018; pp. 304–315.
14. Li, A.; Li, Q.; Gao, W. PrivacyCamera: Cooperative Privacy-Aware Photographing with Mobile Phones. In Proceedings of the 2016 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), London, UK, 27–30 June 2016; pp. 1–9.
15. Aditya, P.; Sen, R.; Druschel, P.; Oh, S.J.; Benenson, R.; Fritz, M.; Schiele, B.; Bhattacharjee, B.; Wu, T.T. Epub I-Pic: A platform for privacy-compliant image capture. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services-MobiSys '16, Singapore, 26–30 June 2016; pp. 235–248. [CrossRef]

16. Zhang, L.; Liu, K.; Li, X.-Y.; Liu, C.; Ding, X.; Liu, Y. Privacy-friendly photo capturing and sharing system. In Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, Heidelberg, Germany, 12–16 September 2016; pp. 524–534. [CrossRef]
17. Cao, Z.; Šimon, T.; Wei, S.-E.; Sheikh, Y. Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 1302–1310.
18. Ribaric, S.; Ariyaeeinia, A.; Pavešić, N. De-identification for privacy protection in multimedia content: A survey. *Signal Process. Image Commun.* **2016**, *47*, 131–151. [CrossRef]
19. Kitahara, I.; Kogure, K.; Hagita, N. Stealth vision for protecting privacy. In Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 26 August 2004; Volume 4, pp. 404–407.
20. Wu, Y.; Yang, F.; Ling, H. Privacy-Protective-GAN for Face De-identification. *arXiv* **2018**, arXiv:1806.08906.
21. Dimiccoli, M.; Marín, J.; Thomaz, E. Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *1*, 1–18. [CrossRef]
22. Yamada, T.; Gohshi, S.; Echizen, I.; Yamada, T. Privacy Visor: Method Based on Light Absorbing and Reflecting Properties for Preventing Face Image Detection. In Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Washington, DC, USA, 13–16 October 2013; pp. 1572–1577.
23. Perez, A.; Zeadally, S.; Matos Garcia, L.J.; Mouloud, J.A.; Griffith, S. FacePET: Enhancing Bystanders Facial Privacy with Smart Wear- ables/Internet of Things. *Electronics* **2018**, *7*, 379. [CrossRef]
24. Kaihoko, Y. Identification of Photo-taking behaviors using Optical Flow Vector. *Int. J. Adv. Trends Comput. Sci. Eng.* **2019**, *8*, 306–312. [CrossRef]
25. ZDNet, Americans Spend Far More Time on Their Smartphones than They Think. Available online: https://www.zdnet.com/article/americans-spend-far-more-time-on-their-smartphones-than-they-think/ (accessed on 28 August 2020).
26. Bhowmik, M.K.; Saha, K.; Majumder, S.; Majumder, G.; Saha, A.; Sarma, A.N.; Nasipuri, M. Thermal Infrared Face Recognition—A Biometric Identification Technique for Robust Security system. *Rev. Refinements New Ideas Face Recognit.* **2011**. [CrossRef]
27. False Acceptance Rate (FAR) and False Recognition Rate (FRR) in Biometrics. Available online: https://www.bayometric.com/false-acceptance-rate-far-false-recognition-rate-frr/ (accessed on 15 July 2020).
28. Tsakanikas, V.; Dagiuklas, T. Video surveillance systems-current status and future trends. *Comput. Electr. Eng.* **2018**, *70*, 736–753. [CrossRef]