 doi:10.20944/preprints202003.0376.v2

Article

Prevention of Unintended Appearance in Photos Based on Human Behaviors Analysis

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

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

Abstract: Nowadays, with smartphones people can easily take photos, post photos to any social networks and use the photos for some purposes. This leads to a social problem that unintended appearance in photos may threaten the privacy of photographed person. 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 photographed person to proactively detect whether someone is intentionally/unintentionally trying to take pictures of him/her. Thereby, he/she can have appropriate reaction to protect the 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 to be 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

1. Introduction

For years, smartphone has been increasingly becoming one of the indispensable personal devices, allowing people to easily take photos recording every desirable moment just by 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 is over 90% in 2018 [2, 3]. This facilitates the explorations of various Social Networking Services (SNS) (e.g., Facebook, Twitter, etc.). In fact, there are bout 3.5 billion people accounting for 45% of the global population are using SNS [1]. As the results, a social problem is potentially occurred when people are unintendedly appeared in others' photos and be published on any social networks. More seriously, the photos along with photographed person's identity can be used by photographers for their own purposes. As the consequence, the privacy of 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 people in vast image collections [17]. 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 purpose, the existing methods might be done at either photographer site or photographed person site. The methods in former category leverage the power of computer vision and machine learning techniques to hide the identity of photographed persons to avoid their identification [4]. For example, Google developed cutting-edge face blurring technology which allows to blur identifiable faces of bystanders on the images [5]. Other solutions aim to automatically recognize photographed persons in images and obscure their identities [6][7][8]. Unfortunately, these approaches give no choice to photographed persons to control over their privacy protection since this process is totally done at photographer site. The methods in latter category attempt to proactively prevent the photos of photographed person from being taken. For example, some techniques make the privacy of photographed persons to be respected based on their privacy preferences represented by visual markers [9] or hand gestures [10] or offlinetags [11] 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 [12]. More sophisticated techniques rely on cooperation between photographers and photographed persons, which allows photographed persons' privacy preferences can be detected by nearby photo-taking devices (via peer-to-peer short-range wireless communications) [13][14][15]. However, this approach requires the photographed persons to broadcast their preferences, leading to the other aspects of privacy. More importantly, this approach might not be effective in the situation in which the photographers proactively switch-off communication function on their devices or ignore the advertised privacy choices of nearby photographed persons because they intentionally and secretly take photo of pre-targeted persons. Indeed, this is also common problem of most of existing studies when they mainly focus on privacy-protection of "bystander" which is defined as either "a person who is present and observing an event without taking part in"[8] or "a person who is unexpectedly framed in"[6] 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" [8]. 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 the situation when someone is intentionally/unintentionally trying to take photos of him/her using mobile phone, without broadcasting his/her privacy preferences as well as identifying information. Afterward, photographed person will have appropriate reaction such as leaving the shared space or asking the photographer to protect the privacy. Note that, in order to sufficiently cover as many as possible the cases of photo-taking, we use the notion of "photographed person" instead of "bystander". We assume that the photographed person has strong motivation in protecting his privacy and willing to use a wearable camera to monitor the surrounding environment. The behavior of likely photographer is recognized via the analysis of his/her skeleton information obtained from monitored video. We argue that misdetection possibly occurs when there exists 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 in analysis process. In our study, such the information is extracted by OpenPose [16] in real-time. Afterward, 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 input data represents photo-taking behavior. The experimental results demonstrate that the proposed approach achieved 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.

 doi:10.20944/preprints202003.0376.v2

2. Related Works

Prior works on handling photographed person's 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 individual record [18]. Thus, to deal with this privacy problem, the common approaches are blurring and pixelization. For example, Frome et al. [4] 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 detector and post-processor was performed to blur the faces. Koyama et al. [6] 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 scenario, 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 distinguish 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 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. [25]. 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 photographed person to advertise his/her privacy preferences based on which 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 photographed person express his/her intention/requirements in privacy-protection*; and *cooperation methods between photographer and photographed person*. Some methods require photographed person to wear visible specialized tags. For examples, Pallas et.al in [11] 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 [23] enables photographed person to broadcast his/her privacy policies and empowers the photo service provider (or photographer) to exert the privacy protection policy.

 doi:10.20944/preprints202003.0376.v2

This approach is similar to the one in [24]. Some other methods require stronger cooperation between photographer and photographed person. For example, Li et al. presented PrivacyCamera [13], an application working on both photographer's and photographed person's mobile phone. Upon detecting a face, the app automatically sends notifications to nearby photographed person who are registered users of the application using short-range wireless communication. If the photographed person does not want to appear in the photo, he/she will request to the photographer. His/her face will be blurred once the photographer confirms the appearance of photographed person in the photo. However, this solution cannot completely solve the privacy problem if photographer intentionally ignore the requests from photographed person

In latter group, the photographed person proactively takes actions to protect his/her privacy. For examples, Yamada et al. [21] proposed a method to avoid the 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 [22] to prevent the unintentional capture of facial images by distorting the region containing the face. This work is similar to the work in [21] 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 [21] used near-infrared light, while the visible light was used in [22]. 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 [26], 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 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 photographed person to proactively make decisions in controlling over his/her privacy. More concretely, it helps him/her to detect the situation when someone is intentionally/unintentionally trying to take photos of him/her and has appropriate reaction to protect the privacy.

3. Proposed Approach

3.1. Photo-taking recognition algorithm

In this section, the proposed algorithm for recognizing photo-taking behavior is presented. In 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 a photo. Typically, our propped algorithm focuses on detecting photo-taking behavior and classifying it from net-surfing behavior. Note that, net-surfing 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 [29], Americans spend an average of 3 hours a day on their smartphone for net-surfing compared to 41 minutes 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 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 is clearly described as follows:

Table 1. proposed algorithm

Proposed Algorithm

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

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

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

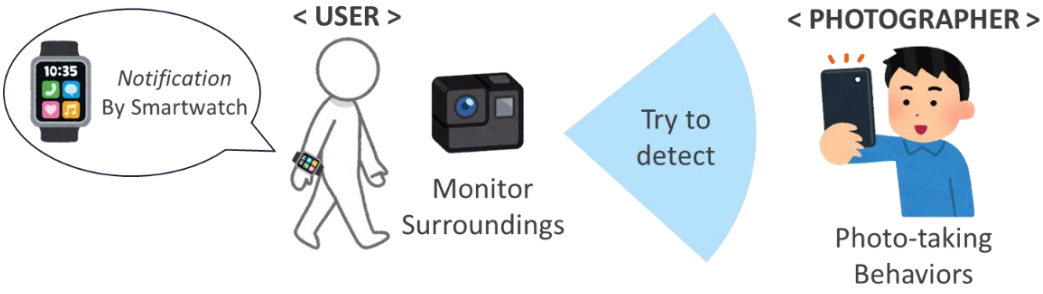


Figure 1. A scenario of photo-taking behavior detection and its notification

3.2. How does proposed approach work in reality?

Figure 1 depicts an assumed scenario for the implementation of our proposal. Accordingly, the photographed person uses a wearable camera to monitor the surrounding environment all the time or in specific event that he/she wants to protect his/her privacy. The monitored video as input data is continuously fed into the detection algorithm which runs on his/her mobile phone or cloud-based device for further analysis. If a photo-taking behavior is detected, a vibration signal as the output is activated to notify photographed person. This allows him/her to perform some types of physical actions. For examples, he/she simply leaves the shared space, or asks the photographers to stop taking photos. However, using a normal wearable camera possibly leads to the 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 [30]. As mentioned in section 2, the photographed person is assumed to be the one who has strong motivation in securing his privacy. Thus, by wearing such a thermographic camera, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

photographed person can proactively control over privacy-protection without violating the others' 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.

3.3. Extract human skeleton information

As stated in subsection 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 the information can be obtained by monitoring: (I) upper arm or (II) lower arm, and (III) the angle of the bending arms.

In order to obtain the skeleton information, OpenPose [16] - an open-source tool is used. By leveraging this tool, the human skeleton information can be extracted in real-time from two-dimension 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*]. Here, the

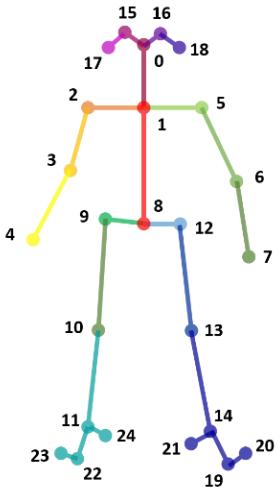
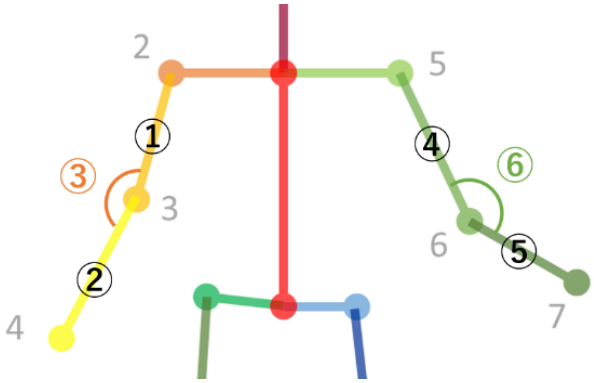


Figure 2. "BODY_25" human skeleton estimation model

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 2 provides brief information of joint positions of the arm parts and the according expressions used in this paper.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

Figure 3. Focusing parts in proposed approach

Table 2. 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.

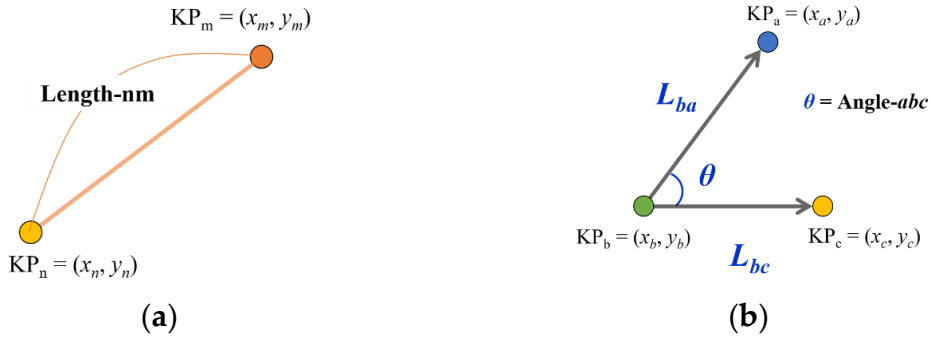


Figure 4. Calculation of the arm's length and angle of the bending arm. (**a**) Calculation of the arm's length from two coordinates by using the distance between two points KP_n and KP_m. This method is applied to calculate the length of ①, ②, ④, ⑤ in Table 2; (**b**) Calculation of the angle of the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

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

According to Figure 4(a), a certain joint position (keypoint) denoted as KP$_p$ is presented as follow:

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

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

The arm length can be calculated by the following equation:

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

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)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

The angle of bending arm θ is formed by $\mathrm{KP_a} = (x_a, y_a)$, $\mathrm{KP_b} = (x_b, y_b)$, $\mathrm{KP_c} = (x_c, y_c)$ as shown in Figure 4(b). 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 below equation:

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

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

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

where, $0 \le \theta \le \pi$.
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 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 follows:

$$\mathrm{X} = x_1, x_2, \cdots, x_i, \cdots, x_I \tag{5}$$
$$\mathrm{Y} = y_1, y_2, \cdots, y_j, \cdots, y_J \tag{6}$$

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's $d(x_i, y_j)$ expresses the distance between the elements: X and Y. It will be transformed from x-y coordinate space to i-j coordinate space as follow:

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

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:

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

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).

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

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

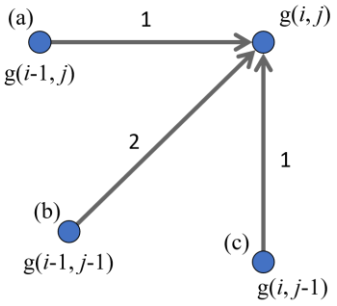
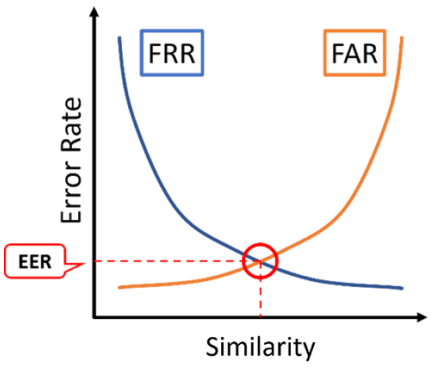


Figure 5. Definition of DP path. To calculate the accumulated distance, (a) to (c) indicates a pattern of distance in *i-j* coordinates space. Each number shown in (a) to (c) expresses the weighted score for calculating the distance by the following Equation (8).

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 [28]. **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.



Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

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 a photo or net-surfing using a smartphone, while the other plays the role of 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 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 8 and Figure 9 partly illustrate the visual outputs obtained from OpenPose for photo-taking and net-surfing behaviors, respectively. Specifically, (a), (b), (c) and (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 3. Accordingly, the dataset1 which consists of $P1$ to $P7$ and $N1$ to $N3$ 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.".

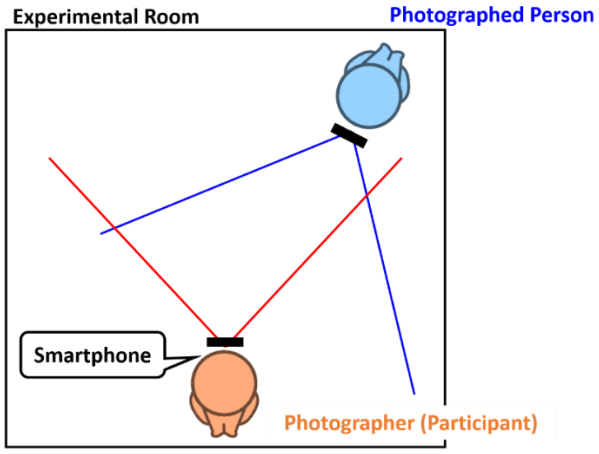
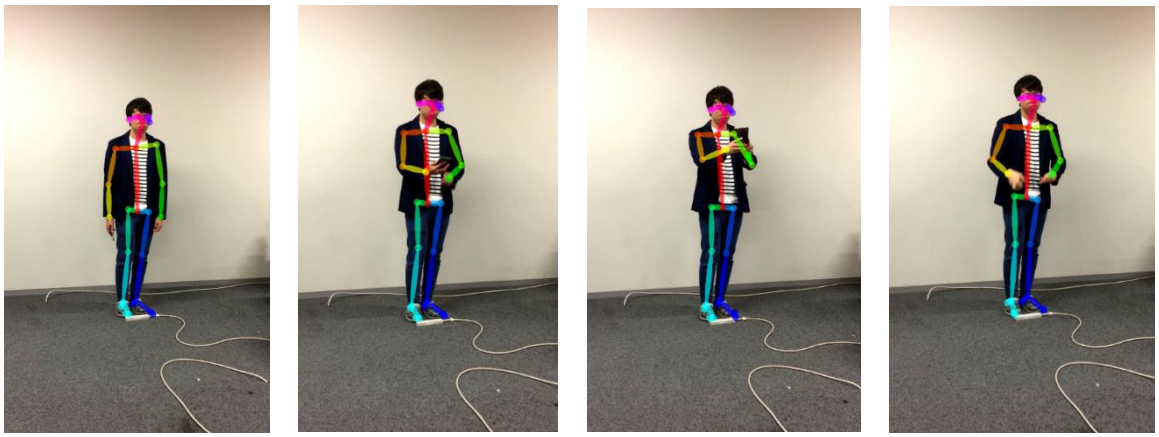


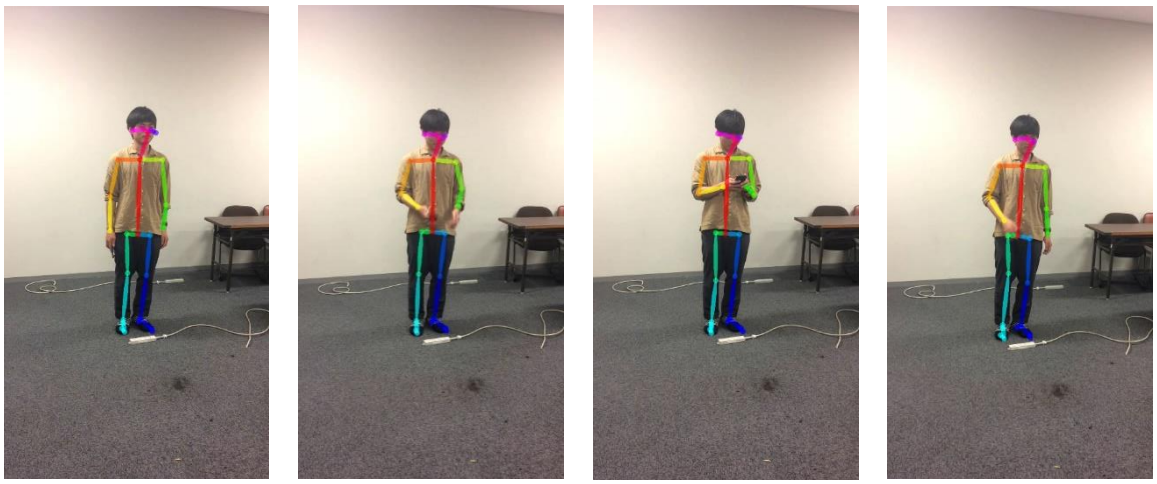
Figure 7. Experimental environment where photographed person records the video of the photographer, while the photographer is performing either photo-taking behavior or net-surfing behavior with a smartphone.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2



(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 3. The obtained datasets from experiments

	Photo-taking behavior	Net-surfing behavior
Dataset1	$P1, P2 \ldots, P7$	$N1, N2, N3$
Dataset2	$P8, P9 \ldots P15$	$N4, N5, N6$

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 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 10 (b)(c) and Figure 11 (b)(c), there are no significant differences between the photo-taking and net-surfing behaviors. Meanwhile, the differences among these behaviors are obvious when observing Figure 10 (a) and Figure 11 (a).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

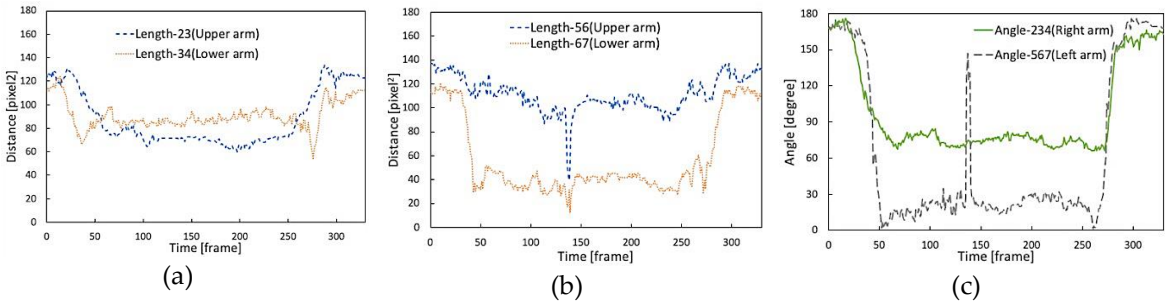


Figure 10. The arm's length and angle of bending arms of subject 1 when taking photo. (**a**) Right upper and right lower arm's lengths; (**b**)Left upper and left lower arm's lengths; (**c**) angles of right bending and left bending arms. In (a) and (b), vertical axis represents distance between joints (length) in pixel². The horizontal axis indicates frame number it means time [frames]; In (c), the vertical axis represent angle in degree. The horizontal axis indicates frame number it means time [frames].

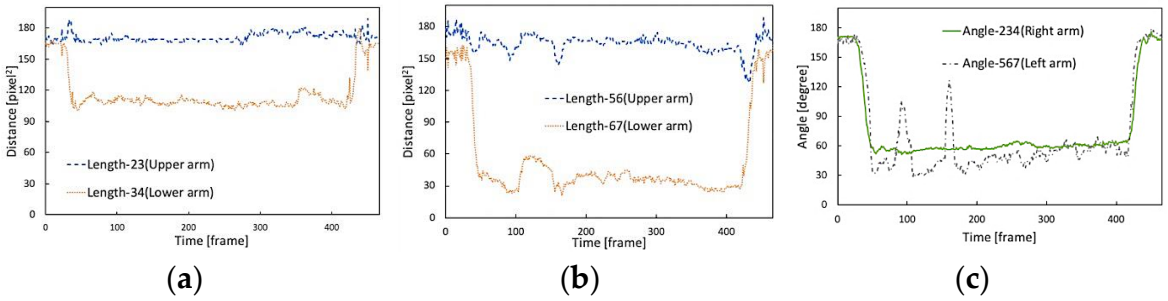


Figure 11. The arm's length and angle of bending arms of subject 1 when performing net-surfing. (**a**) Right upper and right lower arm's lengths; (**b**)Left upper and left lower arm's lengths; (**c**) angles of right bending and left bending arms.

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 10 (b) and (c). Figure 13 shows an example (Example 1) of misdetection in both captured video frame and visual graph. Note that Figure 13 (a) and (b) are the same graphs as Figure 10 (b) and (c), respectively. In these graphs, the 139^{th} frame in which a joint was mis-detected, was emphasized by a yellow rectangle. Figure 13 (c) illustrates 135^{th}, 139^{th} and 142^{nd} frame extracted from video.

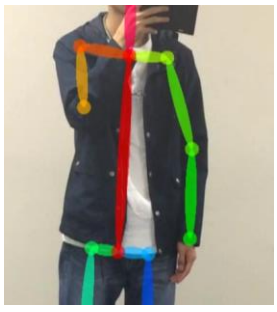


Figure 12. Example of non-detection frame

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

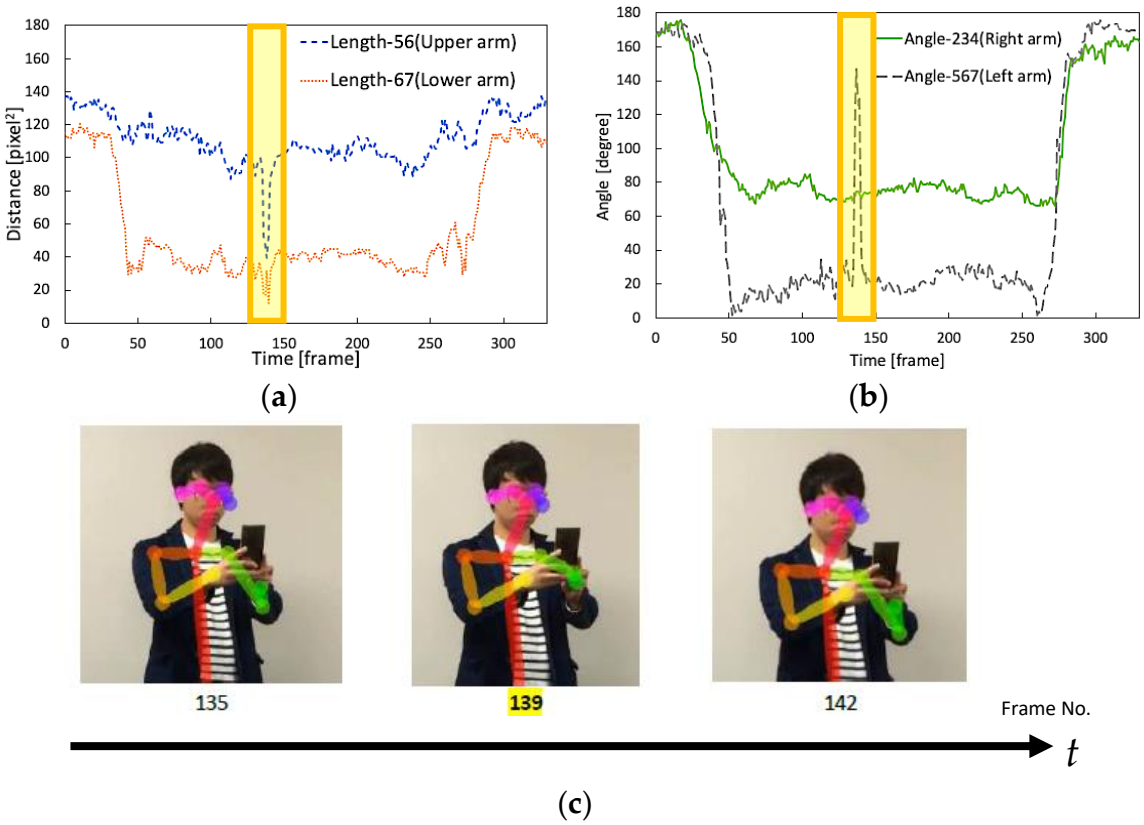


Figure 13. Example 1 of misdetection generated by OpenPose (taken from P1). (**a**) Left upper and left lower arm's lengths; (**b**) angles of right bending and left bending arms; (**c**) Example frame (135^{th}, 139^{th} and 142^{nd})

Figure 14 shows another example of misdetection (Example 2). In Figure 14 (a), 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 14 (b) 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 pixel² 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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

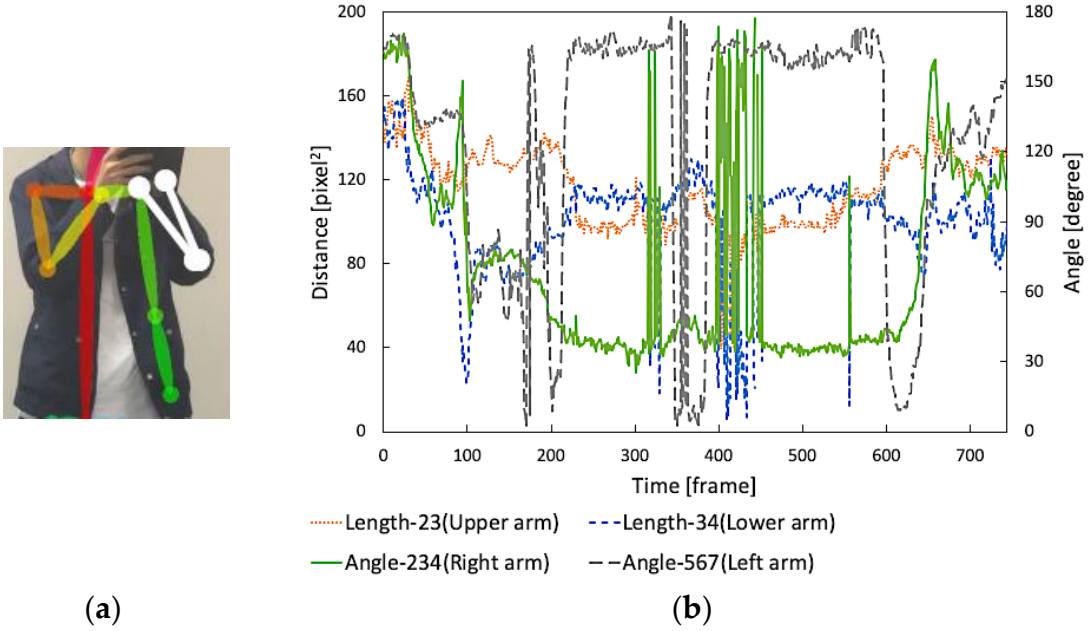


Figure 14. Example 2 of misdetection generated by OpenPose (taken from P5). (a) Misdetection frame (white line presents an expected detection result); (b) result of the right upper and lower arms' lengths and the angle of right/left bending arms. In (b), first vertical axis indicates distance between joints (length) [pixel²]. Second vertical axis indicates angle [degree]. Horizontal axis indicates frame number corresponding to time [frames].

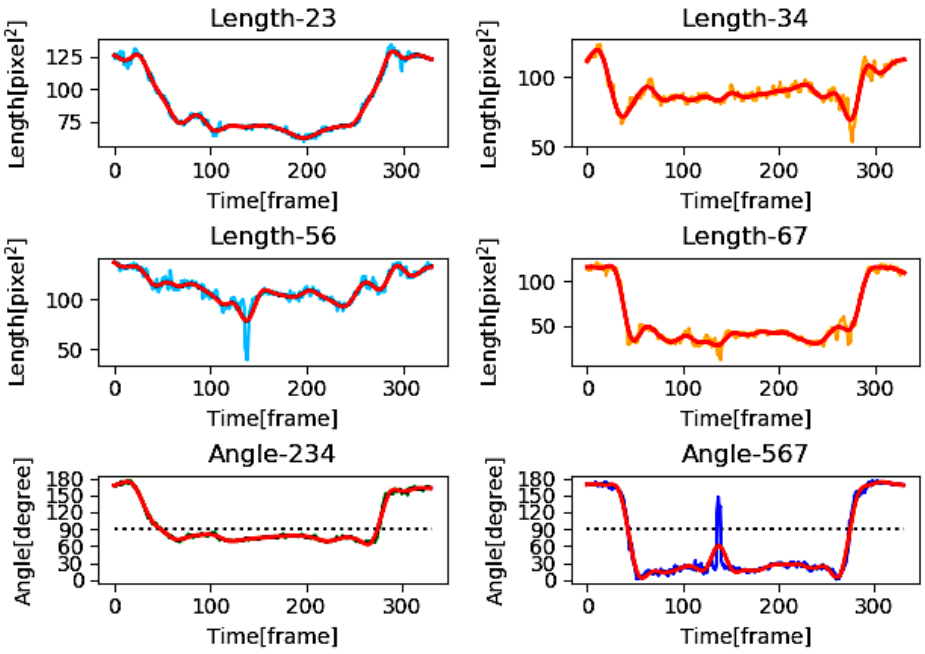


Figure 15. Sample P1 data after applying LPF

4.2. Determination of DP threshold

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

In order to determine the DP threshold (denoted as Th_{DP}), we first performed DP matching on dataset1. 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 4. DP scores obtained from the case where P1 was used as the reference data (Reference data: P1, Input data: P2, P3, …, P7 and N1, …, N3)

		Length-23	Length-34	Length-56	Length-67	Angle-234	Angle-567
Photo-taking behavior	P2	1.69	2.41	5.84	6.23	9.11	4.33
	P3	6.54	14.57	3.41	4.62	8.98	2.73
	P4	3.39	3.36	4.76	4.38	2.28	1.65
	P5	3.92	6.40	10.28	17.9	14.53	21.34
	P6	30.64	10.93	16.87	5.9	4.44	3.91
	P7	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	N1	58.88	8.50	28.6	2.04	3.48	6.31
	N2	21.48	3.18	11.17	3.26	2.73	8.20
	N3	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**

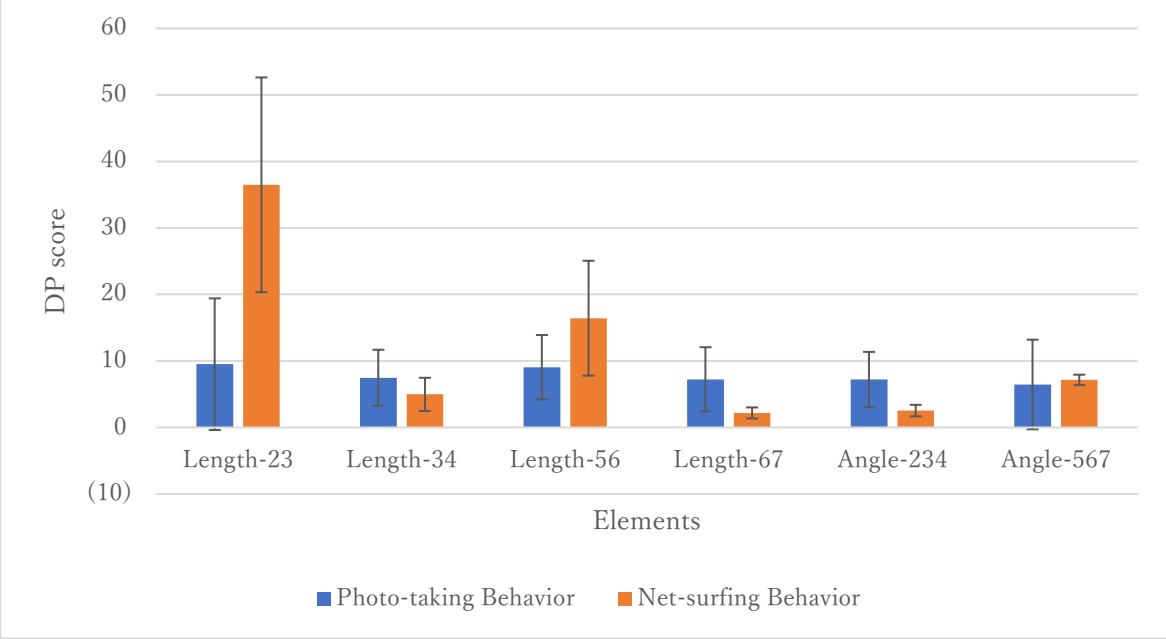


Figure 16. Average DP score for each behavior obtained from the results in Table 4 (Reference data: P1)

Table 4 tabulates an example of DP matching scores of each joint component when data $P1$ was considered as the reference data. Since the monitored videos were taken from the right side of the photographer rather than from his/her 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 4 and plotted in Figure 16. Accordingly, the right upper arm (length-23) shows the biggest difference

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

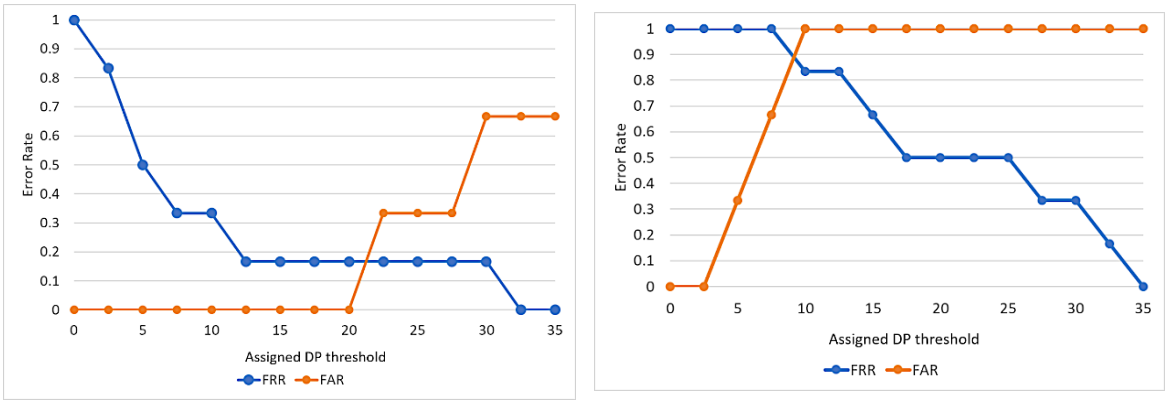
between photo-taking and net-surfing behaviors. Meanwhile, there is not so much difference between these two behaviors can be found in other components. Note that the cases where monitored videos are captured in other sides of the photographer will be considered in future work. Therefore, in this study, we only focus on length-23 in determining DP threshold and in performance evaluation of our proposed approach.

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 following 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 4) 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 $Th_{DP} = 15.9$. In the next subsection, the performance of our approach is evaluated by using this DP threshold and dataset2.



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

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.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 13 September 2020 doi:10.20944/preprints202003.0376.v2

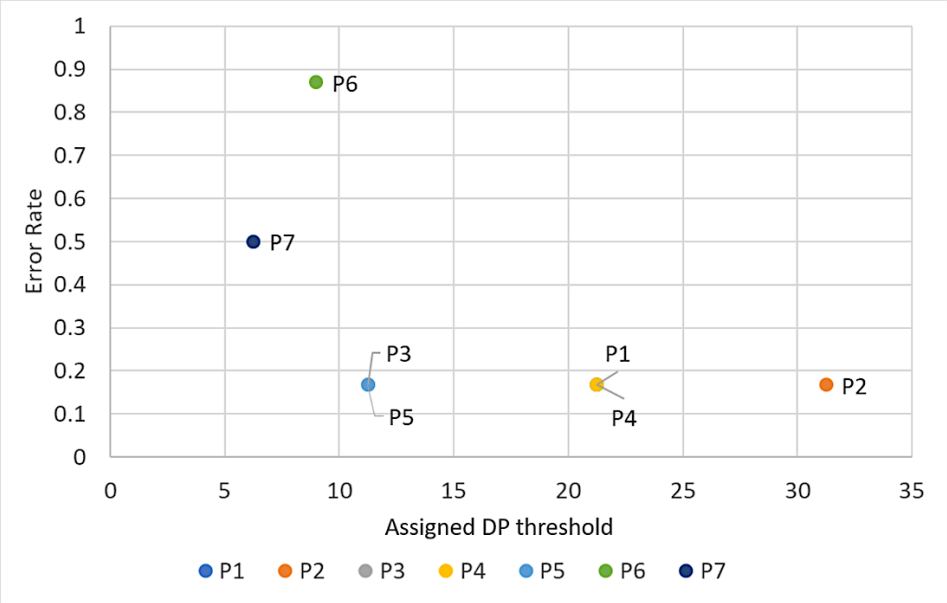


Figure 18. EER distribution obtained from all FRR-FAR curves by cross-validation for right upper arms (f_c=40Hz). The horizontal axis indicates the assigned DP thresholds. The vertical axis indicates the error rate. The legend shows photo-taking behavior data used as reference data.

4.3. Performance Evaluation of the proposed approach

Table 5. DP matching result of dataset2 when using dataset1 as the reference data
(Reference data: dataset1 except outlier [P1, …, P5], Input data: dataset2[P8, …, P15 and N3, …, N6])

			Reference Data: Photo-taking behavior (dataset1)				
			length-23(Right Upper arm)				
			P1	P2	P3	P4	P5
Input Data (dataset2)	Photo-taking behavior	**P8**	8.10	7.89	5.69	8.27	3.40
		P9	4.60	6.68	3.84	4.04	5.91
		P10	0.93	2.84	5.17	2.96	3.28
		P11	8.44	9.49	2.07	3.33	3.22
		P12	1.14	2.30	5.74	2.81	2.91
		P13	5.61	6.00	1.42	1.93	3.65
		P14	21.10	28.43	7.01	17.08	10.18
		P15	13.59	13.74	10.18	13.51	3.14
	Net-surfing behavior	**N4**	35.44	43.55	15.57	32.08	22.06
		N5	24.26	33.34	10.02	22.68	12.71
		N6	14.62	19.47	11.26	17.28	6.67

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 19). The obtained DP scores were then compared with the DP threshold to identify photo-taking behavior. The DP matching results are presented in Table 5. Note that, as explained in subsection 4.2, we only focused on the joint of length-23, thus, Table 5 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 behavior were decided as photo-taking behavior whose DP scores are less than the threshold (Th_{DP}=

 doi:10.20944/preprints202003.0376.v2

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.

$$\text{Accuracy} = \frac{number\ of\ correct\ detection}{total\ number\ of\ DP\ scores} x\ 100\% \quad (12)$$

In order to obtain detection accuracy, the 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, are not the same. It might be varied but not introducing so large difference. For examples, if the reference data P1, P2 or P4 are used, the detection accuracy of the photo-taking behavior is 87.5%, while when either the reference data P3 or P5 is used, the accuracy is 100%. Such variations might come from individual difference 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 number of data for 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.

4.4. Overal discussion and future works

In 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 chance to photographed person to control his/her privacy. This becomes serious problem when photographers ignore photographed person's privacy preferences, and intentionally take photos and share the photos to Social Networks or use the photos without hiding photographed person's identity for some purposes. Our proposed method, on the other hand, allows photographed person to make proactive decision based on his/her privacy preferences. This method potentially works even in the case of protecting the privacy of "bystander" who was not intentionally captured by photographers. Once the photo-taking behavior is detected, the photographed person will receive notifications from their smart phone. This facilitates him/her 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 photographers all the time, this leads to the 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 arm parts of photographer were recorded. However, we believe that this assumption is not sufficient to guarantee the proposed method to completely solve the problem since the face part is probably captured in the video. Therefore, using a thermographic camera instead of normal camera is more realistic approach. Second, there are several real scenarios has not been considered in this study. In practice, the photographed person probably has to protect his/her privacy in a crowded place. In fact, OpenPose can generate 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 photographed person can use our method in situations/events if he/she is aware that the privacy can be violated by a particular photographer. In addition, the case where recorded videos are taken from different sites of 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 [27], the limitations such as 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 real scenario, our proposed approach allows photographed person proactively to protect his/her privacy upon on the privacy preferences, especially in the case where particular photographer intentionally capture him/her in the photos for some purposes. In the future, we will investigate more on various aspects related to real-time processing problems as well as scenario where photographed person is in crowded place.

References

1. We are social, “Digital in 2019”, https://wearesocial.com/global-digital-report-2019 (accessed on January 31, 2020).
2. Ministry of Internal Affairs and Communications (2018), “2018 WHITE PAPER Information and Communications in Japan,” Part2, Figure 5-2-1-1, Transitions in household ownership rates for ICT devices, p.65.
3. Ministry of Internal Affairs and Communications (2018), “2018 WHITE PAPER Information and Communications in Japan,” Part1, Figure 4-2-1-2, Terminals connected to the Internet, p.42.
4. Aditya, P.; Sen, R.; Druschel, P.; Joon Oh, S.; Benenson, R.; Fritz, M.; Schiele, B.; Bhattacharjee, B.; Wu, T.T. I-pic: A platform for privacy-compliant image capture. In Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys), Singapore, 25–30 June 2016; pp. 249–261.
5. Google Street View, “Google-Contributed street view imaginary policy,” https://www.google.com/streetview/policy/#blurring-policy (accessed on July 17, 2020).
6. Tatsuya Koyama, Yuta Nakashima and Noboru Babaguchi, “Real-time privacy protection system for social videos using intentionally-captured persons detection,” Proceedings IEEE of Multimedia and Expo (ICME2013), 6 pages, 2013.
7. Guardian project, “ObscuraCam: Secure Smart Camera,” https://guardianproject.info/apps/obscuracam/ (accessed on July 17, 2020).
8. Hasan, Rakibul, David Crandall, Mario Fritz, and Apu Kapadia. "Automatically Detecting Bystanders in Photos to Reduce Privacy Risks." (2020).
9. Bo, Cheng, Guobin Shen, Jie Liu, Xiang-Yang Li, YongGuang Zhang, and Feng Zhao. "Privacy. tag: Privacy concern expressed and respected." In Proceedings of the 12th ACM conference on embedded network sensor systems, pp. 163-176. 2014
10. Li, Shan, Weihong Deng, and JunPing Du. "Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2852-2861. 2017.
11. Frank Pallas, Max-Robert Ulbricht, Lorena Jaume-Palasí, and Ulrike Höppner. 2014. Offlinetags: a novel privacy approach to online photo sharing. In CHI ’14 Extended Abstracts on Human Factors in Computing Systems (CHI EA ’14). Association for Computing Machinery, New York, NY, USA, 2179–2184. DOI:https://doi.org/10.1145/2559206.2581195
12. Shu, Jiayu, Rui Zheng, and Pan Hui. "Cardea: Context-aware visual privacy protection for photo taking and sharing." In Proceedings of the 9th ACM Multimedia Systems Conference, pp. 304-315. 2018.
13. Li, Ang, Qinghua Li, and Wei Gao. "Privacycamera: Cooperative privacy-aware photographing with mobile phones." In 2016 13th Annual IEEE International Conference on Sensing, Communication, and Networking (SECON), pp. 1-9. IEEE, 2016.

14. P. Aditya, R. Sen, P. Druschel, S. Joon Oh, R. Benenson, M. Fritz, B. Schiele, B. Bhattacharjee, and T. T. Wu, "I-Pic: A Platform for Privacy-Compliant Image Capture," in Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '16. New York, NY, USA: ACM, 2016, pp. 235–248. [Online]. Available: http://doi.acm.org/10.1145/2906388.2906412
15. L. Zhang, K. Liu, X.-Y. Li, C. Liu, X. Ding, and Y. Liu, "Privacy-friendly Photo Capturing and Sharing System," in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ser. UbiComp '16. New York, NY, USA: ACM, 2016, pp. 524–534. [Online]. Available: http://doi.acm.org/10.1145/2971648.2971662
16. Zhe Cao, Tomas Simon, Shih-En Wei, Yaser Sheikh, "Realtime Multi-person 2D Pose Estimation Using Part Affinity Fields," Proc. of Int'l Conf. on Computer Vision and Pattern Recognition, pp.1302-1310, 2017.
17. B. C. McCarthy and A. Feis, "Rogue NYPD cops are using facial recognition app Clearview." https://nypost.com/2020/01/23/ rogue- nypd- cops- are- using- sketchy- facial- recognition- app- clearview/ (Accessed on July 15, 2020)
18. S. Ribaric, A. Ariyaeeiniaand, and N. Pavesic. De-identification for privacy protection in multimedia content: A survey. Signal Processing: Image Communication, 47:131–151, 2016.
19. I. Kitahara, K. Kogure, and N. Hagita, "Stealth vision for protecting privacy," Proc. Int'l Conf. Pattern Recognition, pp.404–407, 2004.
20. Wu, Yifan, Fan Yang, and Haibin Ling. "Privacy-protective-gan for face de-identification." arXiv preprint arXiv:1806.08906(2018).
21. Yamada, T.; Gohshi, S.; Echizen, I. 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 (SMC), Manchester, UK, 13–16 October 2013; pp. 1572–1577.
22. A. Perez, S. Zeadally, L. Matos Garcia, J. Mouloud, and S. Griffith, "FacePET: Enhancing Bystanders Facial Privacy with Smart Wear- ables/Internet of Things," Electronics, vol. 7, no. 12, p. 379, 2018.
23. L. Zhang, K. Liu, X.-Y. Li, C. Liu, X. Ding, and Y. Liu, "Privacy-friendly Photo Capturing and Sharing System," in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ser. UbiComp '16. New York, NY, USA: ACM, 2016, pp. 524–534. [Online]. Available: http://doi.acm.org/10.1145/2971648.2971662
24. P. Aditya, R. Sen, P. Druschel, S. Joon Oh, R. Benenson, M. Fritz, B. Schiele, B. Bhattacharjee, and T. T. Wu, "I-Pic: A Platform for Privacy-Compliant Image Capture," in Proceedings of the 14th Annual International Conference on Mobile Systems, Applications, and Services, ser. MobiSys '16. New York, NY, USA: ACM, 2016, pp. 235–248. [Online]. Available: http://doi.acm.org/10.1145/2906388.2906412
25. M. Dimiccoli, J. Marin, and E. Thomaz,"Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation," Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies, vol. 1, no. 4, pp. 1–18, 1 2018. [Online]. Available: https://doi.org/10.1145/3161190
26. Yuhi Kaihoko, Tan Phan Xuan, Eiji Kamioka, "Identification of Photo-taking behaviors using Optical Flow Vector," International Journal of Advanced Trend in Computer Science and Engineering, Vol. 8, No. 1.4, 2019; pp. 306-312.
27. Tsakanikas, Vassilios, and Tasos Dagiuklas. "Video surveillance systems-current status and future trends." Computers & Electrical Engineering 70 (2018): 736-753.
28. False Acceptance Rate (FAR) and False Recognition Rate (FRR) in Biometrics, https://www.bayometric.com/false-acceptance-rate-far-false-recognition-rate-frr/ (accessed on July 15, 2020)
29. ZDNet, "Americans spend far more time on their smartphones than they think", https://www.zdnet.com/article/americans-spend-far-more-time-on-their-smartphones-than-they-think/ (accessed on August 28, 2020)
30. Bhowmik, Mrinal Kanti, Kankan Saha, Sharmistha Majumder, Goutam Majumder, Ashim Saha, A. Nath Sarma, Debotosh Bhattacharjee, Dipak Kumar Basu, and Mita Nasipuri. "Thermal infrared face recognition—a biometric identification technique for robust security system." Reviews, refinements and new ideas in face recognition 7 (2011).