ISSN 2278-3091
Volume 8, No.1.4, 2019
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse4781.42019.pdf
https://doi.org/10.30534/ijatcse/2019/4781.42019

Identification of Photo-taking behaviors using Optical Flow Vector

Yuhi Kaihoko, Phan Xuan Tan, Eiji Kamioka
Shibaura Institute of Technology, Graduate School of Engineering and Science,
Course of Electrical Engineering and Computer Science, Japan
Email: {ma18028, tanpx, kamioka}@shibaura-it.ac.jp

ABSTRACT

The rate of smartphone ownership has significantly increased all over the world year by year. According to the statistical data by Japanese government, more than 90% of people aged between 20 and 30 own smartphones as of 2017. Smartphones are very useful and it is easy for people to communicate with each other, taking pictures and sharing the pictures on SNS (Social Networking Sites). However, there exists an important social problem related to taking pictures, namely, unintended appearance in photos. When someone is taking a photo in a public place, other people may appear in the photo unintendedly due to the lack of photographer's moral, resulting in a privacy risk of the photographed persons. To avoid such a situation, most of existing studies perform image processing to the photo image, e.g. superimposing pixelated or blurred images around the faces. This is a passive approach for the photographed persons conducted at the photographer side, and thus, there still exists a privacy risk. In this research, an active approach conducted at the photographed person side is proposed, aiming at detecting photo-taking behaviors by smartphone. In the proposed approach, the photographer's behaviors, which show someone is about to take a photo, are focused on. It is assumed that a photographed person (user) wears a small camera like a "life log camera" and monitors his/her surroundings. The final goal of this research is to detect whether the person is about to take a photo or not, based on the video data analysis. In this paper, we analyze the characteristics of photo-taking behaviors using Optical Flow technique, referring to the movement of arms and/or hands of human. The result of evaluation experiments reveals an interesting feature distribution and shows that the detection accuracy of photo-taking behaviors is 67%.

Key words: Photo-taking behaviors, Smartphone, Optical Flow, Privacy

1. INTRODUCTION

According to data of Ministry of Internal Affairs and Communications [1] the ownership rate of smartphones in Japan is 56.8% overall, and the rate is over 90% in 2017 for those under 40. Therefore, almost all people have a smartphone and use for accessing the Social Network Service (SNS). It means that people can easily take photos and post them on the SNS. However, the moral of posting photos on the SNS and the privacy of unintended appearance in photos has been increasingly a big issue. One of the problems for posting photos of other persons via SNS without their permission is so-called photo harassment as shown in Nikkei [2], leading inconveniences, for example, it might disclose the location of photographed person.

As human being, there is no problem for us to recognize the human photo-taking behaviors. However, it is very difficult for computers to detect/recognize human, smartphone, camera and so on. Recently, image processing technologies is increasing by many researchers and machine leaning techniques. For example, YOLO proposed by Redmon et al. [3] is performed real time object detection with high accuracy. Nevertheless, it is not yet perfect for behaviors recognition. There are numerous approaches that passively avoid the unwanted appearance in the photos. In other words, these solutions focus on the photographer side. For example, Frome et al. [4] proposed privacy protection for all people in Google Street View. This method is effective to use public situation, e.g. upload to Internet and protect all person privacy. However, it is no meaning to hide all person when use on private situation. That's why their face is hidden by this system when they post their pictures via SNS. Meanwhile, there are few works in the active prevention of appearance in photos from photographed person side. Yamada et al. [5] proposed method to avoid the appearance in photos physically using privacy visor that uses infrared rays. Thereby, the user needs to attach the privacy visor with power supply and cable for protect the privacy. However, it might cause a large burden for the user. Therefore, in this research, a novel method is proposed in order to prevent appearance in photos from the photographed person side, resulting in a protection of the photographed person's privacy. The proposed method leverages Optical Flow technique to detect photo-taking behavior, where Optical Flow vector Norm and Angle based video analysis is focused on. In this paper, we collect six persons data-set of photo-taking behaviors through the experiment. From the results, we could find interesting characteristics and common features on Optical Flow vector information. Analyzing the results, we defined thresholds of photo-taking behaviors. Finally, photo-taking behaviors detecting accuracy can be got 67% by using that thresholds.

In this paper, we talk about related work as the first. Next, methodology of this study. Then, evaluation for this research. In this section, we provide and explain about experiment and its results. Then, we talk about discussion. Finally, conclusion of this paper and taking about feature work.

2. RELATED WORKS

Kaihoko et al. [6] proposed a method to prevent the unintended appearance in photos through the detection of behavior of taking photo from photographer side. This method is mainly concentrated on smartphone detection by considering its geometrical features. The authors indicated that this approach is effectiveness due to its dataset-agnostic characteristic, even if the overlap of hands on holding a smartphone. However, photo-taking behavior was not taken into account in this study.

Tsai et al. [7] proposed the Optical Flow based analysis of human behavior-specific system. Although this proposal did not relate to the detection of photo-taking behaviors, this is very useful for us to proceed with this research. The main proposal is certain specific behaviors are detected by using Optical Flow. The specific behaviors are smoking, drinking, phoning and other in this paper. They used face recognition, object (handheld) detection and their method. Then, detect the behaviors. However, these behaviors are very simple ones. There is only limited motion on each behavior. Meanwhile, photo-taking behaviors/motions are more complicated than that. When the photographer will take photos, he/she uses one hand or two hands. Moreover, the smartphone orientation is in landscape or portrait. This paper doesn't refer to detect many types of behaviors. In particular, photo-taking behaviors and many motions in one behavior. Therefore, we need to consider this thing.

3. METHOTOLOGY

3.1 Assumptions

For the feasibility of the proposal, in this study, let assume that a photographed person (user) wears a small camera like "Life log camera" and monitors surroundings. Then, based on the monitored video data, the indications of whether the other persons are trying to take photos of him/her or not, are made. These indications are decided through the consideration of the following criteria:

- Whether the camera is directed to him/her
- Whether he/she is in the frame of camera

The notifications will be sent to the user once the above criteria are confirmed. For instance, the user will receive a warning by vibration of a wearable devise such as smartwatch.

3.2 Photo-taking Behaviors Detection

As the nature of human being, when a person tries to take pictures using smartphone, his/her arms and/or hands parts must be moved. Therefore, in this paper, an Optical Flow-based method is proposed in order to detect photo-taking behavior. Optical Flow is a popular technique in computer vision, especially for analyzing the motion and tracking an object. Technically, Optical Flow visualize motion pattern of an object. By estimating optical flow differences between consecutive video frames, the Optical Flow vector of object is obtained as shown in Figure 1, depicting an object (rocket) movement. The numbers this figure indicates the frame number on moving rocket video, meanwhile black arrow shows Optical Flow displacement vector.

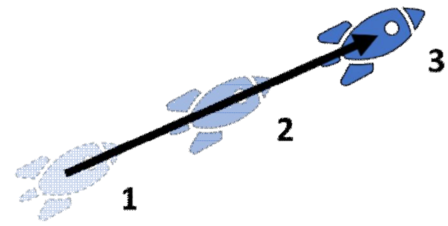


Figure 1. Image of an object movement with Optical Flow vector

Besides of that, Optical Flow comes up with several assumptions:

(1) The pixels intensities of an object don't change between consecutive frames.

(2) Neighbor pixels have similar motion.

There are numerous methods for calculating Optical Flow value. one of them is called Lucas-Kanade method which estimates non-dense Optical Flow feature. In addition, a method to calculate the dense Optical Flow was proposed by Farnebäck [8]. In this paper, we will utilize the latter one as a tool to calculate Optical Flow value because the dense information is useful to analyze the behaviors. Thereby, the proposed Optical Flow-based approach in detecting photo-taking behavior can be broken down into following steps:

(a) Collect video data of photo-taking behaviors

(b) Detect human part by object detection and moving part by Optical Flow

(c) Analyze the video and obtain the features (Norm and Angle of Optical Flow)

4. EVALUATION

4.1 Experiments

A. Experimental Setup

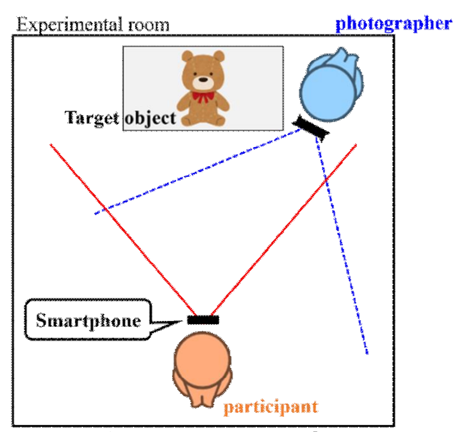


Figure 2. Experimental setup

This experiment purpose is twofold. First, whether Optical Flow technique is possible to detect photo-taking behaviors or not. Second, whether photo-taking behaviors characteristics is obtained by Optical Flow information or not. Therefore, the experimental setup was established as shown in Figure 2 where the orange person is a participant (user) and the blue person is a photographer. In this experiment, we focus on the movement of photo-taking behaviors to investigate the characteristics. Hence, this experiment must be conducted in a static situation as the initial step. There were 6 subjects who are currently university students took part in this experiment as participants. Each participant took pictures of target object by a smartphone, while the photographer was recording videos to capture the participants' photo-taking behaviors. In this experiment, we accepted any taking photos style of the participant. This experiment was conducted in a small meeting room with sufficient light condition. The photographer who was another student, used the camera of iPhone 5s to take videos with output frame rate is 30 *fps*.

B. Data Acquisition and Analysis

The obtained video data was analyzed throughout image processing technique. In particular, a program written in Python that relied on OpenCV library and TensorFlow object detection API, was utilized for processing the video data. Python version is 3.6.5. Accordingly, the data acquisition and analysis were performed as following steps:

a) Detect the human part
b) Focus on human part and obtain its coordinates
c) Estimate Optical Flow value on only human part
d) Estimate Norm and angle (value of cosine) of Optical Flow vector
e) Collect data in (c) and visualize its

We focus on only (c) ~ (e) steps in this paper because TensorFlow object detection API cannot always detect the human. In this case, we could not obtain a human part and the coordinates in every frame. Hence, the human part was defined manually. In addition, we remove the noise that is camera shake information in this experiment from a preliminary experiment in order to analyze Optical Flow data. The noise value was defined by us as less than Optical Flow norm is two.

In order to estimate Optical Flow value in human part, the dots as shown in Fig. 3 was utilized, which was defined in each video frame. In this experiment, the dots are depicted as a sampling grid with 16 by 16 of size for each video frame. Technically, Optical Flow vector shows arrow's direction in the video image at each dot. Fig. 3 shows an example of Optical Flow where the rocket is moving from lower left (rocket shown in dash line) to upper right (black rocket). Then, the orange arrows indicate Optical Flow vector at each dot. The information of those vectors at all dots was collected as time series data.

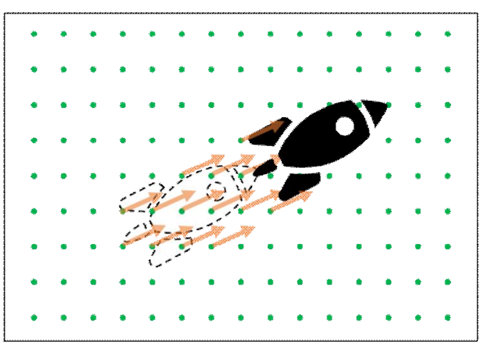


Figure 3. Example of sampling grid of Optical Flow

In this experiment, we can obtain the original coordinates (start point) of all dots and Optical Flow vector coordinate (end point). Initially, each point's coordinates could be defined as follows:

$$\text{startpoint}(n) = (x_n, y_n), \qquad \text{endpoint}(x_{Fn}, y_{Fn})$$

Where, n means the number of pixels in images. In addition, unit vector for x axis is defined in order to calculate angle shown in below:

$$\boldsymbol{e}_x = (1,0), \qquad ||\boldsymbol{e}_x|| = 1$$

Therefore, the norm and angle (value of cosine) is calculated by these things. The norm and angle must be calculated by below equations:

$$||\boldsymbol{F}_n|| = \sqrt{F_{xn}^2 + F_{yn}^2} \tag{1}$$

Where,

$$\boldsymbol{F}_n = \begin{bmatrix} x_{Fn} - x_n \\ y_{Fn} - y_n \end{bmatrix} = \begin{bmatrix} F_{xn} \\ F_{yn} \end{bmatrix}$$

$$\cos\theta = \frac{\boldsymbol{F}_n \cdot \boldsymbol{e}_x}{||\boldsymbol{F}_n|| \cdot ||\boldsymbol{e}_x||} = \frac{x}{\sqrt{F_{xn}^2 + F_{yn}^2}} \tag{2}$$

It should be noted that in our experiment, video image condition was rotated left by 90 degrees to analyze the video data by computer program. Therefore, the coordinate needed to be rotated left by 90 degrees. The linear transformation in terms of rotation left by 90 degrees shown in below:

$$(x, y) \xrightarrow{f} (-y, x) \qquad f\text{: Rotate left by } 90°$$

The angle value must be modified according to the above linear transformation with equation (2). Thereby, the angles value is defined as cosine value. Even though it is easy to understand angle value with degrees expression, cosine expression should be obtained periodic value. Because the cosine function is a periodic function.

C. Evaluation Methods

Initially, we define the starting frame is when the user initially moving focus on just beginning to take photos behaviors for detecting photo-taking behaviors. Therefore, the beginning motion is extracted from experimental results. Then, we determine the starting frame and ending frame of photo-taking behaviors through a preliminary experiment. In this time, starting frame is defined by when moving arms with a smartphone in front of their body. Ending frame is defined by when there is no motion of the photographer. Finally, the obtained data is based on these frames data in order to analyze the behaviors.

The results were evaluated by relation between Optical Flow angle (cosine value) and number of frame (time) with norm information, in addition to the distribution of correlation between Optical Flow norm and angle. From these data, thresholds/conditions were determined by Optical Flow common characteristics from obtained six persons dataset. The determined thresholds were then tested with the data-set obtained from the other six subjects. For evaluation purpose, the accuracy of detection was calculated as follow:

$$\text{Accuracy} = \frac{\text{Passed the thresholds/conditions subjects}}{\text{all subjects (in this case 6)}}$$

4.2 Results

A. Experimental Results

Initially, the table 1 shows all video data information, e.g. time, the number of dots, frame and so on. Accordingly, the attribute of "Time" indicates the duration of video file, "Dots" indicates defined number of dots in sampling grid of each video frame, whereas, "Frame" indicates the number of frames in each video.

Table 1. Summary of all video data information for data-set

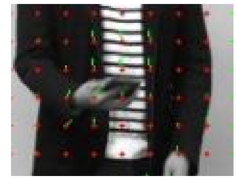
Video No.	Time [sec]	Dots	Frame
1	11	192	330
2	16	216	503
3	11	200	352
4	26	180	797
5	24	200	744
6	10	200	304

Figure 4 shows an actual image of participants who are participating in the experiment. In particular, Figure 4-(1) shows original image of conducting experiment. Figure 4-(2) shows state of estimating Optical Flow in whole image. In this figure, the human part which is covered by red dots is focused on. Meanwhile, Figure 4-(3) illustrates the Optical Flow vectors on moving part of human part, which is presented by the green arrow, in other words, Optical Flow arrow/vector.

Figure 5 shows transition of Optical Flow vector's norm with time series data. Vertical axis indicates norm value and horizontal axis is number of frames. In addition, Figure 6 shows transition of Optical Flow vector angle (cosine) with time series data. Vertical axis indicates cosine values and horizontal axis is number of frames. All depicted results are taken from subject No.1. As shown in Figure 5, the color line shows transition of vector norm on each dot. In this case, for the subject No.1, there are 192 lines. In addition, the motion part can be seen clearly from this figure. There is no cosine value when the Norm value is 0 form equation (1) and (2). That's why some point is not shown in line but only a dot in Figure 6.

(1) Condition of conducting experiment

(2) Calculating Optical Flow in whole image



(3) Focusing on Optical Flow vector in moving part

Figure 4. Photos of Experimental environment and Analyzing process

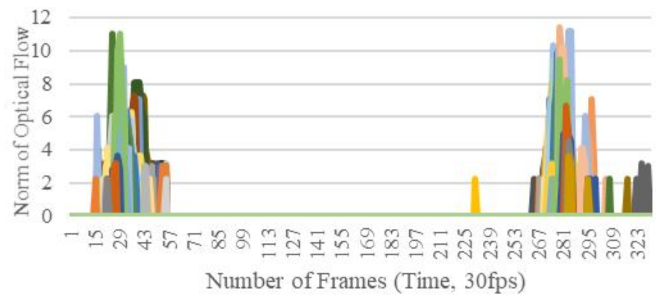


Figure 5. Transition of Optical Flow vector norm (taken from subject 1)

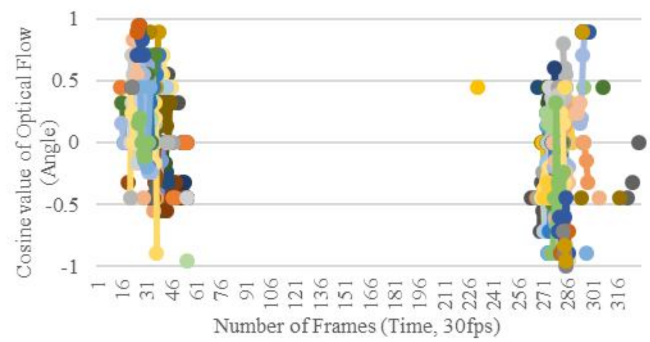


Figure 6. Transition of Optical Flow vector angle (cosine) (taken from subject 1)

B. Evaluation Results

Table 2 shows the result of start-end frame and motion time on each video.

Table 2. Summary of all video data information from preliminary experiment for data-set

Video No.	Time [sec]	All Frame	Starting Frame	Ending Frame	Motion time [sec]
1	11	330	25	92	2.2
2	16	503	79	130	1.7
3	11	352	47	121	2.4
4	26	797	77	133	1.8
5	24	744	209	253	1.4
6	10	304	104	146	1.4

In fact, people tend to take photo by putting the camera in front of their faces or chests. Therefore, Optical Flow information will be extracted from the specific areas of face and chest of each subject.

In this study, using the real time data and Optical Flow data, common characteristics is found out from six person-dataset. The analysis of Optical Flow information is twofold. First, considering the correlation between number of frames and the angle. Second, considering the correlation between vector angle and norm.

Figure 7 illustrate the relation between the angle (defined by cosine value) and the number of frames. In this figure, color bar on each graph right side indicates the value of Optical Flow vector's norm where white color represents the smallest value. The red circles indicate results of specific area of face and chest as mentioned above. Throughout this figure, we found the characteristic to determine photo-taking behavior, that is to say, the value of zero of cosine and its continuity. Once cosθ is equal to zero, the vector orthogonally moves up with the theta θ of 90 degree. In other words, the arms or/and hands are vertically moving up. Accordingly, we found such the tendency in major part of monitored data. Therefore, such the characteristics will be then utilized to classify photo-taking behavior. Specifically, we determine the thresholds which is the value of zero of cosine and its continuity as over 10 frames in a specific area in this experiment. From data of No.2, 3, 4, 5 and 6.

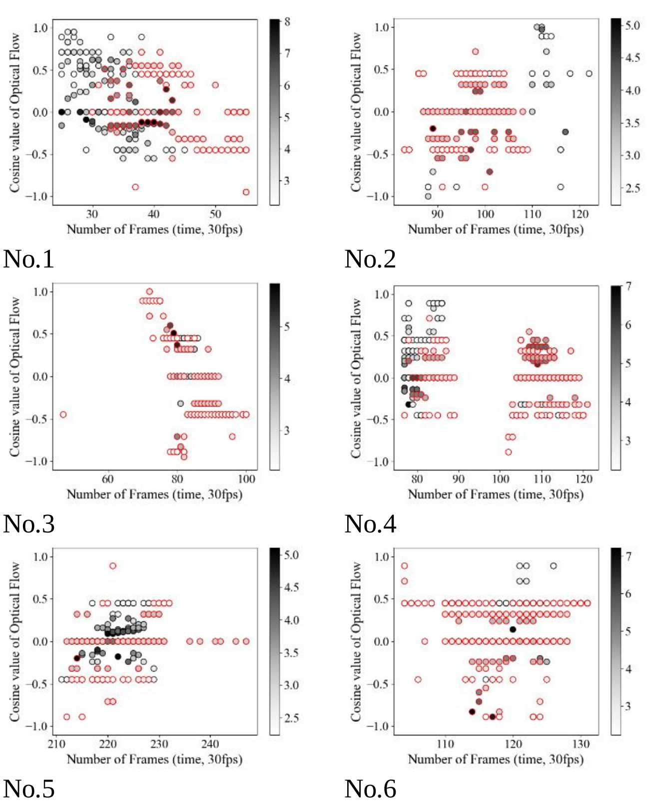


Figure 7. Frame-Angle relation which respects to the norm value in each video of each subject. The vertical axis is cosine value and horizontal axis is number of frames

Figure 8 depicts the distribution of correlation between Optical Flow norm and angle, which is estimated by using Kernel Density Estimation (KDE) approach. In this figure, the curves on upper and right side shows histogram with KDE. Meanwhile, the contour lines in graphs shows distribution and frequency of obtained data points. Accordingly, frequency becomes higher once the blue color becomes thick. As shown in this figure, although the distribution in 6 cases are not all the same, there is a clear tendency can be found in cases of 1, 2, 4 and 5. It can be seen that the high frequency is highly distributed around the area with a cosine value of zero. Therefore, this characteristic will be taken into account as a factor to determine the behavior of taking photo.

Based on the above characteristics, in order to classify the behavior of taking photo, we conducted the experiment with the same scenarios, but other six subjects. The obtained video data of this experiments is shown in Table 3 where photo-taking time is not much different from the one in previous experiment. The experimental results are now evaluated based on frame-angle relation and the distribution of correlation of vector norm and the angel.

Figure 9 illustrate that result of frame-angle relation respect to Optical Flow vector's norm value. According this figure, the cosine value of zero can be found continuously within results of five subjects (No.8, 9, 10, 11 and 12).

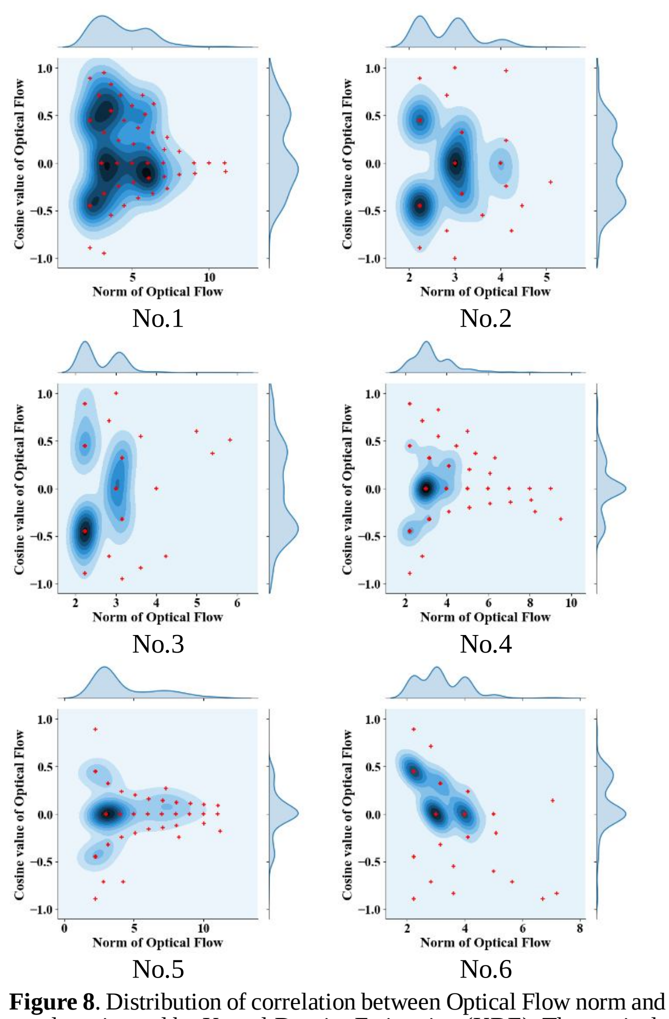


Figure 8. Distribution of correlation between Optical Flow norm and angle estimated by Kernel Density Estimation (KDE). The vertical axis is cosine value of angle and horizontal axis is the norm of Optical Flow vector

Table 3. Summary of all video data information

Video No.	Time [sec]	Dots	All Frame	Strating Frame	Ending Fraem	Motion time[sec]
7	13	192	394	34	128	3.1
8	17	200	518	55	84	0.97
9	10	200	323	27	89	2.1
10	17	198	537	99	160	1.8
11	13	162	402	114	159	1.5
12	27	200	828	134	178	1.4

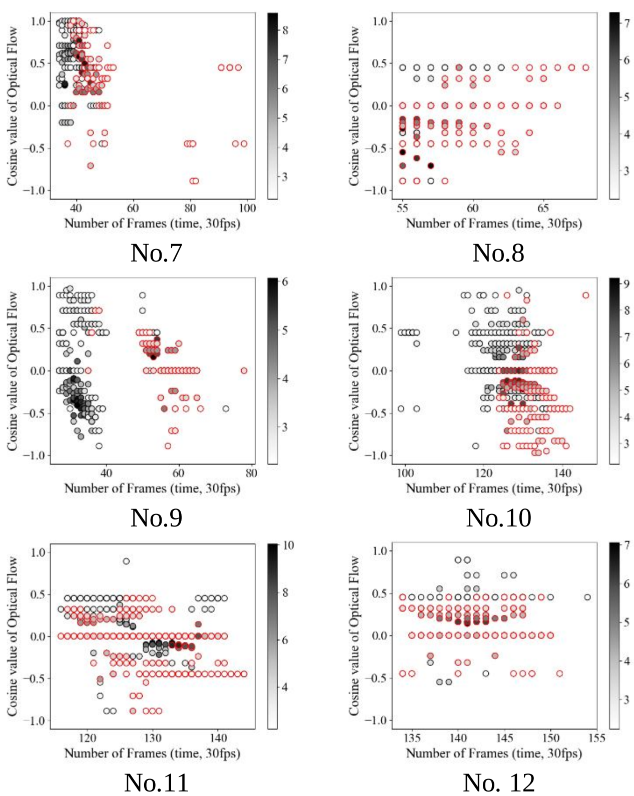
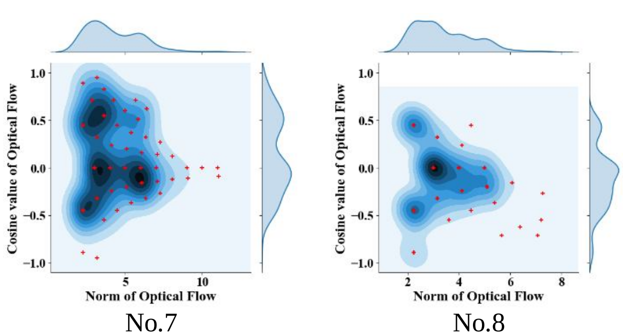


Figure 9. Time-Angle relation with norm's color bar and specific area

Figure 10 shows distribution of correlation between Optical Flow vector's norm and angle which is estimated by KDE. It can be seen that the bell curves of this distribution appear in visualized results of four subjects (No.8, 10, 11 and 12). Hence, that accuracy of detection of photo-taking behavior is around 67%.



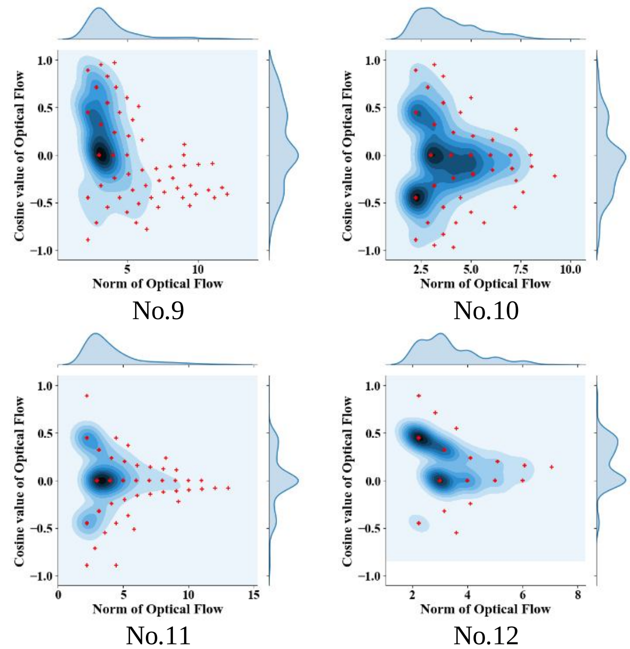


Figure 10. Distribution of correlation between Optical Flow norm and angle estimated by Kernel Density Estimation (KDE)

5. DISCUSSION

From the results, photo-taking behaviors must be identified by setting thresholds based on the time-series information of cosine value and distribution of Optical Flow norm and angle. In this section, the identification method and its accuracy are discussed.

Initially, a threshold of cosine value with considering time series data was investigated. The threshold was determined by not only continuity of cosine value but also considering specific area which is in front of face and chest area. It is important to focus on a specific area because whole motion includes a lot of noise information, e.g. moveless parts, camera shake and unrelated motion. Hence, these noise information need to be removed to effectively and easily analyze the data. As the result, common features have been obtained in the experiment. In addition, the characteristics of photo-taking behaviors are differentiated from other behaviors by focusing on a specific area. Here, thresholds are defined continuously over ten frames on a specific area, that is to say, it is around 0.3 sec. It seems to be a little bit difficult to identify photo-taking behaviors from only this threshold because 0.3 sec is very short time. However, this threshold is determined based on the time and angle information, hence, it does not always happen. Therefore, this threshold can be regarded as an effective one. This is not enough for detecting photo-taking behaviors. Hence, another condition is discussed here. That is a distribution between Optical Flow norm and angle. We can see a characterized shape like a bell curve in each graph of Fig. 8 and 10. Actually, the results from eight out of 12 subjects, show almost the same distribution by photo-taking behaviors. This is very interesting and a significant common feature. As a result, we can say that photo-taking behaviors have a significant relation with both Optical Flow norm and angle distribution, and the action time. However, these tendencies need to be broken down as numerical indicators by optimal analysis.

Using these conditions, we can get the accuracy of photo-taking behaviors detection, which is around 67%. This accuracy is not really high. Actually, in this experiment, we did not specify the photo-taking behaviors, meaning that each subject behaved as he/she likes, hence, the number of samples for each behavior is not large. Nevertheless, the accuracy is 67% which is quite prospective. We believe that the accuracy will be drastically improved when more experiments are performed with various behaviors.

As a future work, we need to consider more complicated scenarios, which requires photo-taking behavior to be classified from other types of motions and behaviors. The other Optical Flow techniques are also needed to be investigated for higher detection accuracy. In addition, in this paper, we did not deeply discuss how to differentiate photo-taking behaviors from other similar behaviors. Therefore, the characteristics of other behaviors should be investigated comparing with photo-taking behaviors.

6. CONCLUSION

In this paper, we propose an approach that relies on Optical Flow technique in detecting photo-taking behaviors. Thereby, such the behaviors can be detected with accuracy of 67%. In this study, we practically found that the frame-angle relation and the distribution of correlation of Optical Flow vector's norm and angle play very important roles. In photo-taking behavior, while the former one is represented by the continuous moving up of Optical Flow vector with the degree value of 90, the latter one shows the highest distributed area of high norm value respect for cosine value of 0. Although the accuracy of this proposed method is just about 67%, it is still a potential and effective approach in detection photo-taking behavior. In order to improve the accuracy, a larger dataset need to be obtained in short future. In addition, other behaviors will be analyzed for clarifying the difference in each motion.

REFERENCES

1. Ministry of Internal Affairs and Communications, **2017 WHITE PAPER Information and Communications in Japan**, Trend in individual smartphone ownership rate, 2017, ch.1, Figure1-1-1-2, p.2.
2. Nihon Keizai Shinbun, **Photo-hara ni goyojin SNS mudantoko toraburuni (in Japanese) [Beware in Photo harasment, to be trouble in SNS without permission]**, 2018, October 10, Retrieved 2019/01/12, https://www.nikkei.com/article/DGXMZO36299280101 02018CC0000/
3. Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi, **You Only Look Once: Unified, Real-Time Object Detection**, 2016 IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, June 2016, pp.779-788. https://doi.org/10.1109/CVPR.2016.91
4. Andrea Frome, German Cheung, Ahmad Abdulkader, Marco Zennaro, Bo Wu, Alessandro Bissacco, Hartwig Adam, Hartmut Neven and Luc Vincent, **Large-scale privacy protection in Google Street View**, Proceedings on IEEE Computer Vision (ICCV2009), 2009, pp.2373-2380.
5. Takayuki Yamada, Seiichi Gohshi and Isao Echizen, **Evaluation of privacy visor to prevent face recognition from camera image** (in Japanese), The Institute of Electronics, Information and Communication Engineers Technical Report Order Information, Vol.112, No.420, EMM2012-92, 2013, pp.7-12. https://doi.org/10.1109/ICCV.2009.5459413
6. Yuhi Kaihoko, Manami Kanamaru, Yoshihiro Nagano, Phan Xuan Tan and Eiji Kamioka, **Prevention of Unintended Appearance in Photos through Detection of Taking Photo Behavior by Photographer** (in Japanese), The Institute of Electronics, Information and Communication Engineers Technical Report Order Information, Vol.118, No.305, EMM2018-34, November 16, 2018, pp.69-74.
7. Hsin-Chun Tsai, Chi-Hung Chuang, Shin-Pang Tseng, Jhing-Fa Wang, **The Optical Flow-based analysis of human behavior-specific system**, 2013 1st International Conference on Orange Technologies (ICOT), 2013, pp.214-218. https://doi.org/10.1109/ICOT.2013.6521195
8. Gunnar Farnebäck, **Two-Frame Motion Estimation Based on Polynomial Expansion**, 13th Scandinavian Conference on Image Analysis, SCIA 2003 Proceedings, Sweden, June-July2003, pp.363-370. https://doi.org/10.1007/3-540-45103-X_50