An OpenPose Based Method to Detect Texting while Walking

Hiroshi Hanaizumi* and Haruhi Misono

Hosei University, 3-7-2 Kajino-cho, Koganei, Tokyo 184-8584, Japan

*hana@hosei.ac.jp

Abstract

A method was proposed to detect so called “texting while walking” from positions of human body 25 keypoints estimated by the OpenPose for a person in a video scene. The method was based on the fact that movements of human joints had some spatial correlation structures. Walking behaviors with or without using a smartphone were separated by the difference between the correlation structures. We evaluated the difference using a canonical correlation analysis applied to two covariance matrices calculated from pose vectors for both walking behaviors, respectively. The canonical score well separated the texting while walking from the normal one. The proposed method was successfully applied to actual video scenes for three persons’ three types of walking behaviors.

Keywords: pose vector, spatial correlation structure, canonical correlation analysis, smartphone, walking behavior.

1. Introduction

With the spread of mobile phones and the increasing of smart applications, so called texting while walking, that is, using smartphone while walking has come to be seen everywhere. In places where many people gather, such as a station, using smartphone while walking may cause an accident such as collision with another person or injury in some cases, such as falling from the platform to the track. For avoiding those dangers, we should detect person using smartphone while walking, and warn to stop the using. There have been many methods and/or systems for detecting the texting while walking. Most of them, however, needed some hardware devices such as accelerometer attaching bodies⁽¹⁻²⁾ for detecting the behavior through a neck tilt angle. The Kinect was also used for recognizing the walking behavior in former methods⁽³⁾. On the other hand, unique methods have developed for estimating person pose from a

monocular camera image⁽⁴⁻⁶⁾. Here, we propose an Openpose based method for detecting the texting while walking. The proposed method separates the problem walking from normal one only by using positions of human body 25 keypoints estimated by the Openpose⁽⁷⁻⁸⁾. The basis of the method is the fact that human joints do not move randomly but have some spatial correlation among them. We extract the correlation structure into covariance matrices for both normal walking and one with using a phone, respectively. A canonical correlation analysis enhances the difference between the both walking behaviors. The 1st canonical score is used as an index for the detection.

2. Principle and Procedures

2.1 OpenPose⁽⁸⁾

The OpenPose is a software system developed at Carnegie Mellon University. The OpenPose estimates positions of human body 25 keypoints in an image or video captured by a monocular camera. Table 1 indicates those keypoints. The body keypoint detector was trained using

Table 1. Human body keypoints

Number and Key-point			
No.	Key	No.	Key
0	Nose	13	L_Knee
1	Neck	14	L_Ankle
2	R_Shoulder	15	R_Eye
3	R_Elbow	16	L_Eye
4	R_Wrist	17	R_Ear
5	L_Shoulder	18	L_Ear
6	L_Elbow	19	L_BigToe
7	L_Wrist	20	L_SmallToe
8	Mid_Hip	21	L_Heel
9	R_Hip	22	R_BigToe
10	R_Knee	23	R_SmallToe
11	R_Ankle	24	R_Heel
12	L_Hip		

deep learning. The keypoint positions are provided through the json file as coordinates x and y on the image with the degree of confidence c .

2.2 Pose Vector and Covariance Matrix

We apply the OpenPose to every scene in video capturing a walking behavior. We extract the coordinates x and y for all body keypoints from the generated json file. For the k -th scene coordinates x_i^k and y_i^k ($i = 0, 1, \dots, m-1$), we define a vector $\mathbf{q}_k$ as

$$\mathbf{q}_k = (x_0^k \ y_0^k \ x_1^k \ y_1^k \ \dots \ x_{m-1}^k \ y_{m-1}^k)^T. \quad (1)$$

where, subscripts mean keypoint number, and superscript t transpose. We cancel the person movement by subtracting coordinates of the MidHip position (number 8) from those of the elements, and define pose vector $\mathbf{p}_k$ as

$$\mathbf{p}_k = (x_0^k - x_8^k \ y_0^k - y_8^k \ x_1^k - x_8^k \ y_1^k - y_8^k \ \dots \ 0 \ \dots \ x_{24}^k - x_8^k \ y_{24}^k - y_8^k)^T. \quad (2)$$

Using all N scenes in the video, we calculate the covariance matrix Σ and call it behavior matrix as

$$\Sigma = PP^T, \quad (3)$$

where, P is a matrix with the pose vectors in columns as

$$P = (\mathbf{p}_0 - \boldsymbol{\mu} \ \mathbf{p}_1 - \boldsymbol{\mu} \ \dots \ \mathbf{p}_k - \boldsymbol{\mu} \ \dots \ \mathbf{p}_{N-1} - \boldsymbol{\mu}), \quad (4)$$

and

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{k=0}^{N-1} \mathbf{p}_k. \quad (5)$$

Now, human behaviors are coded into the behavior matrix. We consider that correlation structure of the matrix represents some features of the behaviors such as normal walking or texting while walking. Thus, we use the OpenPose as a behavior coder with detecting humans in a scene. Next, we extract the information from the matrix and evaluate the difference between two behaviors.

2.3 Canonical Correlation Analysis and Score

The proposed method detects a special behavior by separating it from another one. The separation is reduced to quantitatively evaluate the difference between two covariance matrices. Let Σ_1 be the behavior matrix for a normal walking, and Σ_2 a special walking to be detected. We calculate a matrix Σ_0 as

$$\Sigma_0 = \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2}, \quad (6)$$

where, a spectral decomposition is used for the calculation. The special correlation structure in Σ_2 against to Σ_1 is emphasized by the normalization. For integrating the

emphasized differences into a score, we need a projection vector. Let L be a rotation matrix which diagonalizes the matrix Σ_0 as

$$L^T \Sigma_0 L = L^T \Sigma_1^{-1/2} \Sigma_2 \Sigma_1^{-1/2} L = \begin{pmatrix} \lambda_0 & & & \\ & \lambda_1 & 0 & \\ & 0 & \ddots & \\ & & & \lambda_{m-1} \end{pmatrix}, \quad (7)$$

where, $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{m-1}$ are the eigen values of the matrix, and L consists of the eigen vectors having the same order with the eigen values as

$$L = (\mathbf{l}_0 \ \mathbf{l}_1 \ \dots \ \mathbf{l}_{m-1}), \quad (8)$$

and

$$(\mathbf{l}_i, \mathbf{l}_j) = \begin{cases} 1 & (i = j) \\ 0 & (i \neq j) \end{cases} \quad (9)$$

Thus, the projection vector is obtained as the first row of matrix B as

$$B = (\mathbf{b}_0 \ \mathbf{b}_1 \ \dots \ \mathbf{b}_{m-1}) = \Sigma_1^{-1/2} L, \quad (10)$$

and the score is estimated by an inner product as

$$s_k = (\mathbf{p}_k - \boldsymbol{\mu}, \mathbf{b}_0), \quad (11)$$

for the k -th scene pose vector. Once we obtain the projection vector, we can evaluate the behavior using a scene.

2.4 Correcting positions of the Key-Points

The coordinates of human body keypoints obtained by the OpenPose include positional fluctuation based on recognition uncertainty, and in some cases, the right and left legs may be swapped or one may not be recognized. Some of the slight fluctuations are statistically smoothed in applying the canonical correlation analysis, but severe misrecognitions should be corrected. Some of them are found by the degree of confidence. We correct those keypoint coordinates by using a linear interpolation as

$$\begin{cases} x_i^k = \frac{1}{2}(x_i^{k-1} + x_i^{k+1}) \\ y_i^k = \frac{1}{2}(y_i^{k-1} + y_i^{k+1}) \end{cases} \quad (0 \leq i \leq 24), \quad (12)$$

where, coordinates having low confidence in the k -th scene are replaced with the mean of those in both the $(k-1)$ -th and the $(k+1)$ -th scenes. The correction works for isolated misrecognition, but does not work for sequential ones. We currently correct those remaining low confidence coordinates manually.

2.5 Procedures of the Proposed Method

In applying the proposed method, various walking behaviors are firstly captured into video. Then videos are decomposed into sequential images, and the OpenPose is applied to all the images. Coordinates of the body key-points are extracted from the json file corresponding to the image. After checking and correcting the coordinates, they are used to produce the projection vector and to detect the special walking behavior.

3. Experiments and Discussion

In order to evaluate the performance of the proposed method, we applied the method to eight videos captured for three persons with three types of walking as listed in Table 2. Camera was placed so that its optical axis was perpendicular to a wall, and persons walked parallel to the wall as shown in Fig. 1.

Table 2. Eight video data processed

Person \ Walk	A	B	C
Normal	V ₁	V ₄	V ₇
Phone	V ₂	V ₅	V ₈
Phone + Bag	V ₃	V ₆	-

Fig. 1 Camera and walking track.

Images decomposed from the videos have 640 x 480 pixels. Those images were processed by a Windows 10 based PC. Figures 2 show some decomposed images from outputs of the OpenPose, where a person walked without smartphone (a), with smartphone in both hands (b), and with both smartphone and bag. We found not a few misrecognitions in the keypoints coordinages as shown in Figs.3. One frequent of them was reduction, that is, both legs were recognized as one leg. Other remarkable ones were leg swaps. Keypoint coordinates on both legs often swapped. Besides, right arm behind the body was some-

Fig. 2 Examples of captured data with OpenPose outputs.

Fig. 3 Legs were recognized as one leg (a), and legs swapped in sequential two scenes (b) and (c).

times not recognized. We detected those misrecognitions by using the degree of confidence and corrected them by using Eq.(12). Those not automatically detected were manually corrected. Now, we applied the method to the eight videos.

3.1 Results and Discussion

Figure 4 shows an example of the position correction of the keypoints due to the misrecognition. In Fig.4, blue line shows y coordinate of the original position of the left elbow in a walking behavior and the red one those after correction. The misrecognition affects the mean value of

Fig. 4 An example of the position correction (left elbow).

the data, which may add unexpected bias components to the covariance matrix and distort the correlation structure. For getting the higher performance, we have to do the position correction carefully.

After correcting the misrecognitions, we applied the proposed method to some pairs of the data listed in Table 2. Figures 5 show the 1st canonical scores for three persons'

Fig. 5 The 1st canonical scores calculated for normal walking (red line) and one with using a phone (blue) for three persons.

walking with and without using a smartphone. Actually, we calculate covariance matrices Σ_1 using pose vectors estimated from V_1 and Σ_2 from V_2 , respectively, for person A, and obtained the canonical scores for both sets of pose vectors in a stable walking region. For other persons, we applied the method with the same manner. In those results, the scores were separated well from each other. Although there is some overlap, we can detect the texting while walking with using an appropriate threshold. In the point of view of score fluctuation, the result for person C gave us the best separation. This means that the OpenPose had less misrecognition for person C. In fact, we confirmed that keypoint swaps in left and right legs were small for person C. The keypoint misrecognition rate may change depending on the clothes worn by a person. We consider that keypoints for persons A and B probably contained misrecognitions which could not be corrected.

Next is the detection of walking while using a smartphone in the right hand and with a bag on the left arm. Figures 6 show the scores obtained for persons A and B.

Fig.6 Scores calculated for normal walking (red line) and one with bag and using a phone (green) for two persons.

Red line indicates the score for normal walking and green for one with bag and phone. Although, there are some overlaps in the result for person B, scores are well separated for the person A.

Finally, we tried to evaluate the generality of the method. We obtained canonical scores for three pose vectors estimated from V_1 , V_2 and V_3 , respectively, for person A as the inner products with the projection vector b_0 calculated from V_8 and V_7 for person C. Figure 7 shows those scores, where red line indicates normal walking, blue walking with phone, and green with both bag and phone. Although red and blue lines were overlapped, green one was clearly separated from the others. This means that some types of walking with using smartphones can be detected by using an appropriate projection vector. We also consider that this result was affected by quality of data (misrecognition rate), person size, and other unknown parameters. For solving those problems, some methods will be required to correct the misrecognition for improving data quality and to normalize raw data for getting the score free from person size and his/her walking direction.

Fig. 7 Canonical scores of three walking behaviors of person A by using projection vector for person C.

4. Conclusions

We proposed a method to detect so called “texting while walking” from coordinates of human body 25 key-points estimated by OpenPose. The method was based on the fact that human joint points cooperatively move according to a behavior. The movement of them had a unique spatial correlation structure on the behavior. We extracted information on the correlation structure as a covariance matrix of pose vectors. Finding the difference between normal walking and one with using smartphone was

reduced to emphasizing the difference between the two covariance matrices. We achieved the process by applying a canonical correlation analysis to those covariance matrices. The 1st canonical score well separated the walking behaviors. The method was successfully applied to three persons with three types of walking behaviors.

Our project has just started and there are some problems to be solved in order to apply the method more stably with sufficient accuracy. As the OpenPose may yield misrecognitions such as legs swap, developing a new algorithm for finding and correcting the misrecognitions is the first subject for a future study. The second one is to develop a normalizing algorithm for generalizing the proposed method. The third one is to expand the method to detecting other human behaviors.

References

- (1) M.T.Nguyen, Q.K.Dang, Y.S.Suh and Y.J.Chee, “A Method for Estimating Text Neck While Walking Using a 3-axis Accelerometer,” Proc. Int. Conf. BSN2017, pp.23-26 (2017)
- (2) Shikishima, K.Nakamura and T.Wada, “Detection of Texting While Walking by Using Smartphone’s Posture and Acceleration Information for Safety of Pedestrians,” Proc. Int. Conf. ITST2018, pp.1-6 (2018)
- (3) S.Kaenchan, P.Mongkolnam, B.Watanapa, and S.Sathienpong, “Automatic Multiple Kinect Cameras Setting for Simple Walking Posture Analysis,” Proc. Int. Conf. ICSEC2013, pp.245-249 (2013)
- (4) L. Pishchulin, et.al., “Deepcut: Joint subset partition and labeling for multi person pose estimation,” Proc. Int. Conf. CVPR2016, pp.4929-4937 (2016)
- (5) E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, “Deepcut: A deeper, stronger, and faster multi-person pose estimation model,” pp.34-50, Proc. Int. Conf. ECCV 2016 (2016)
- (6) Tomas Simon, Hanbyul Joo, Iain Matthew, and Yaser Sheikh, “Hand Keypoint Detection in Single Images using Multiview Bootstrapping,” Proc. Int. Conf. CVPR2017, pp.4645-4653 (2017)
- (7) Shin-En Wei, Varun Ramakrishna, Takeo Kanede, and Yaser Sheikh, “Convolutional Pose Machine,” Proc. Int. Conf. CVPR2016, pp.4724-4732 (2016)
- (8) Zhe Cao, Tomas Simon, shin-En Wei, and Yaser Sheikh, “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields,” Proc. Int. Conf. CVPR2017, pp.1302-1310 (2017)