

Research Article

Estimation of Self-Posture of a Pedestrian Using MY VISION Based on Depth and Motion Network

Joo Kooi Tan^{1,*}, Tomoyuki Kurosaki²¹Faculty of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata, Kitakyushu, Fukuoka 804-8550, Japan²Graduate School of Engineering, Kyushu Institute of Technology, 1-1 Sensuicho, Tobata, Kitakyushu, Fukuoka 804-8550, Japan

ARTICLE INFO

Article History

Received 10 November 2019

Accepted 18 May 2020

Keywords

Posture
posture analysis
pedestrian
MY VISION
depth and motion network

ABSTRACT

A system is proposed that performs gait analysis of a pedestrian to improve a walk posture and at the same time to prevent a fall. In the system, a user walks with a chest-mounted camera. His/her walking posture is estimated using a pair of images obtained from the camera. Normally it is difficult to estimate the camera movement, when the parallax of the image pair is small. Therefore, the system uses a convolutional neural network. Optical flow and camera movement, and depth images are estimated alternately. Satisfactory results were obtained experimentally.

© 2020 The Authors. Published by Atlantis Press B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).

1. INTRODUCTION

In recent years, an increasing aging rate has become a serious problem worldwide. Due to the fatal fall of elderly people over the age of 65, the elderly need care. However, there are relatively few young people who can care for elderly people and there is a shortage of human resources. Therefore, elderly people need to take measures by themselves to prevent falls. It is more convenient and useful to measure their own walking posture by themselves, rather than measuring their walking posture from another person's viewpoint using special facilities. This paper proposes a method of analyzing one's walking posture from chest-mounted camera images. The goal is to help elderly people improve their walking posture by the feedback of analyzed posture in order to prevent a falls.

Researches on the acquisition of human walking posture generally called motion capture have been done for the purpose of creating animation motion and virtual reality games in many cases. Many studies have proposed the installation of a single camera [1], multiple cameras [2,3] or a time of flight (ToF) camera [4]. There are two main methods for estimating a posture of a pedestrian. One is to attach markers to human joints and estimate their positions by convolutional neural network [5], for example. A study has also been proposed in which multiple cameras are attached on a subject (human) limbs and joints and estimates the walking posture using the geometric relationship between the cameras [6]. However, it is a rather large system and no one can use it easily.

This paper proposes a method of analyzing one's walking posture from a video provided by a monocular chest-mounted camera, which we call a MY VISION system. The camera motion of chest

mounted camera is estimated using Depth and Motion Network (DeMoN) [7]. The obtained camera motion data is corrected based on a walking motion. The difference between each walking motion pattern is quantified from the experiment considering multiple walking patterns, and the effectiveness of the proposed MY VISION system is evaluated by comparing each roll, pitch and yaw angle of walking patterns to the angles obtained from a 9-axis sensor which provides ground truth. This method is distinguished from existing methods by adopting a monocular self-mounted camera. It is also claimed that the proposed system can be used easily by elderly people with a small, light-weight camera.

2. MATERIALS AND METHODS

This section describes a method of estimating the rotation angle of a chest-mounted camera for posture estimation. DeMoN is used for estimating the camera rotation angles.

2.1. Camera Motion Estimation

Two images are fed into the neural network shown in Figure 1. The two image frames are taken out every n frames on a video. The relative angle from the initial camera posture is obtained by calculating the product of the relative angles of the output at times t and $t + 1$. In addition, trend removal is performed for the purpose of removing accumulated errors. The left network in Figure 2 predicts the optical flow and its reliability, and the right network predicts the depth image, camera motion and surface normal. The bootstrap net only forwards the fed image pairs to the left network of the iterative net, while the iterative net also inputs depth images, camera motion, and

*Corresponding author. Email: etheltan@cntl.kyutech.ac.jp

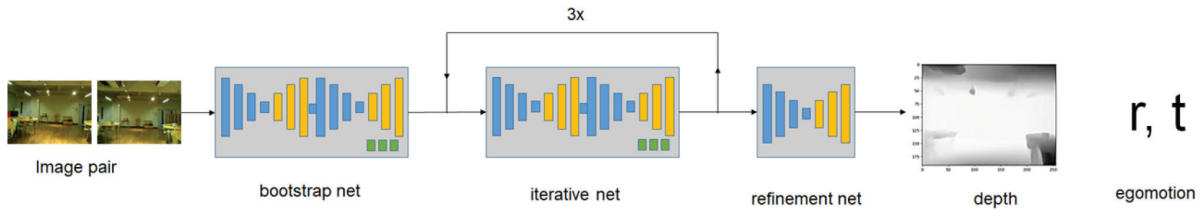


Figure 1 | Overview of the used network.

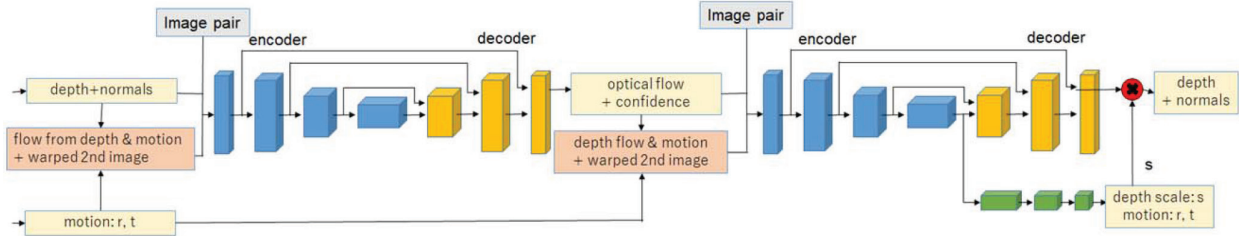


Figure 2 | Schematic of the networks used in the bootstrap net and in the iterative net.

surface normals output from the right network in Figure 2 (i.e., the iterative net inputs its own outputs by feedback, other than the input from the bootstrap net). The iterative net is executed three times to improve the estimation accuracy of the optical flow, depth image and camera motion at the same time. The camera motion is referred to by translation T and rotation R . In this paper, the camera motion R is used to analyze the walking posture of a person.

The refinement net shown in the right network of Figure 1 generates depth images from low resolution (64×48) to high resolution (255×192), as shown in Figure 1. This corrects the incorrect depth prediction from the edge of the depth image.

2.2. Correction Based on Camera Motion

Trend removal calculates the sum of squares with a plausible line for the data by simple regression analysis, and calculates the least square approximation line that minimizes it. This is done by finding the difference between the obtained least squares approximation line and the $y = 0$ line and subtracting the difference from the data. In addition, since the initial angle is not 0 due to the trend removal, correction based on the camera motion is performed. The straight line connecting the start point and the end point of the data from which the trend has been removed is used as the straight line before correction. Also, the starting point being 0, the end point is obtained from the camera motion obtained by inputting the first and the last image pairs of the video to the network, and the straight line connecting them is taken as the corrected straight line. Once the difference between these two lines are found, it is subtracted from the trend-removed data.

2.3. Indicators for Walking Analysis

The extrema for the corrected data is found. Then, in order to consider only extreme values corresponding to significant posture changes excluding camera vibrations, only those extreme values whose absolute value of the difference between successive extreme

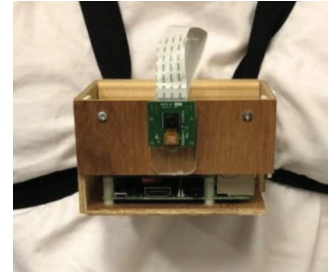


Figure 3 | Experimental equipment (a camera and a 9-axis sensor).

values is greater than or equal to a threshold value are used. The average of the obtained maximum and the local minimum value is used as an index of gait analysis. Particularly, we focus on a forward inclination angle index as a factor of balancing ability of a person. The index of forward inclination angle or swayback angle (i.e. the pitch angle of walking posture) is expressed by Eq. (1), where max and min are the obtained maximum value and the local minimum value, respectively, and t ($t = 1, 2, \dots, T$) is the number of steps.

$$\theta(t) = \frac{1}{2}(\max(t) + \min(t)) \quad (1)$$

3. EXPERIMENT

This chapter describes the evaluation method and results of the posture estimation experiment using the proposed method.

3.1. Experimental Environment

In the experiment, we fabricated and used the device shown in Figure 3 that is equipped with a camera and a sensor, and is mounted on the chest. As experimental data, some images were taken by the chest-mounted camera in an outdoor environment. At the same time, the Raspberry Pi was used to obtain the values of a 9-axis sensor consisting of acceleration, gyro, and geomagnetism.

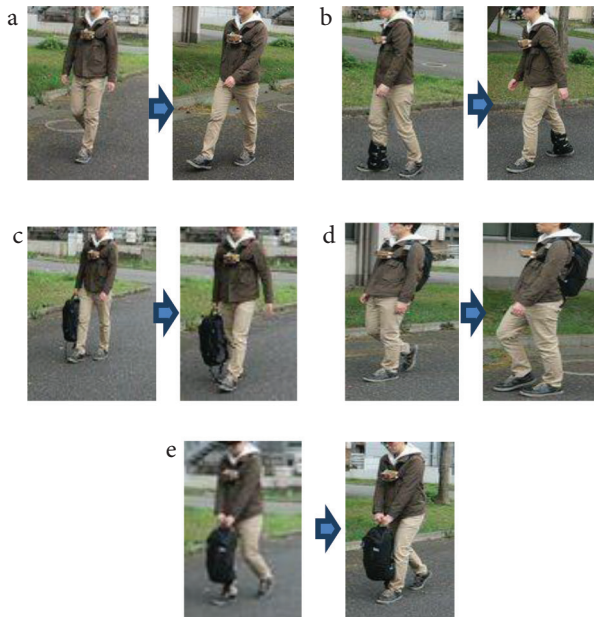


Figure 4 | Walking posture: (a) Normal walk, (b) drag leg walk with a weight attached to the right foot, (c) walk with a heavy bag on the right hand, (d) swayback walk with a heavy bag on the back, (e) forward bending walk with a heavy bag in both hands.

Experiments were performed with five different walking patterns by six persons (all in their 20 s). The details of the walking patterns are described as follows, and they are illustrated in [Figure 4](#).

- (i) Normal walk,
- (ii) Drag leg walk with a weight attached to the right foot,
- (iii) Walk with a heavy bag on the right hand,
- (iv) Swayback walk with a heavy bag on the back,
- (v) Forward bending walk with a heavy bag in both hands.

3.2. Experimental Method

The video obtained from the camera is chosen images every 10 frames and fed into the neural network. Then the camera rotation angle between the two input images is derived. By accumulating these angles, the camera rotation angle from the initial frame to the current frame is obtained. The true value is obtained by integrating the values from the 9-axis sensor using the RTQF algorithm (an inertial measurement unit library from RTIMULib library) [8]. However, as the sampling interval of the sensor is normally different from that of the camera, the sensor information at the time closest to the camera sampling time is used in the experiment. The initial posture of the sensor is set to 0°.

3.3. Experimental Results

[Table 1](#) shows the average accuracy of the five types walking posture before the correction based on trend removal and walking posture, whereas [Table 2](#) shows the average accuracy after the correction

Table 1 | Average accuracy of the walking posture estimation without using trend removal correction (RMSE) (°)

Walking patterns	Roll	Pitch	Yaw
Walking posture			
Normal	2.76	4.50	7.53
Drag leg	7.78	11.09	8.23
Walk with a heavy bag	6.93	7.21	6.80
Swayback	4.96	15.15	15.97
Forward bending	5.87	20.41	18.31

Table 2 | Average accuracy of the walking posture estimation after using trend removal correction (RMSE) (°)

Walking patterns	Roll	Pitch	Yaw
Walking posture			
Normal	1.33	1.71	2.43
Drag leg	4.85	5.18	7.07
Walk with a heavy bag	3.80	2.79	3.84
Swayback	2.41	2.73	4.85
Forward bending	3.35	4.39	3.55

Table 3 | RMSE of the estimated index value by [Eq. \(1\)](#) (°)

Person's walking postures	Pitch (°)					
	A	B	C	D	E	F
Normal walk	3.76	2.77	6.69	2.28	1.07	1.54
Forward bending walk	1.64	4.97	3.74	7.52	2.36	3.10
Swayback	3.16	8.63	8.04	1.43	2.20	3.70

based on trend removal and walking posture. The accuracy of the estimation is evaluated using Root Mean Square Error (RMSE). This index is expressed by [Eq. \(2\)](#), where N is the number of data, x_i is an estimated value and X_i is a true value.

$$\text{RMSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - X_i)^2} \quad (2)$$

[Table 3](#) shows the RMSE of true values obtained from the 9-axis sensor and the estimated value of the index of forward inclination angle [see [Eq. \(1\)](#)], or the index of angle of normal walking, forward bending walking, and swayback walking, respectively, by six persons.

4. DISCUSSION

The results in [Table 2](#) are more accurate than the results in [Table 1](#), indicating that correction based on trend removal and camera motion is effective. For example, in [Table 2](#), when comparing the values of roll and pitch of the walking patterns, the average values of roll and the average of pitch differ approximately 0.2°. On the other hand, in [Table 1](#), the average values of roll and the average of pitch differ about 6.0°. Looking at [Table 1](#), RMSE of the roll and pitch values of swayback and forward bending walking patterns differ more than 10°. In [Table 2](#), one can see that the results are related to walking patterns. In the table, the roll, pitch and yaw of the normal walking pattern are between 1° and 2.5°. However,

drag leg walking pattern is between 4° and 7° . This is probably because ‘normal’ is good for the left and the right balance, but ‘drag leg’ is biased toward either side because of drag. Figure 4c and 4e show that walk with a heavy bag and forward bending walk are apparently similar, but Table 2 shows that the pitch angle of the latter pattern is double higher than a former angle. From these results, there is a relation between the walking pattern and the rotation angle of the chest-mounted camera. Therefore, it is considered possible to improve the walking posture by analyzing the motion of the chest-mounted camera. From Table 3, we find that the index of forward inclination angle of the normal walk of persons A and C are larger than other persons. This indicates that the posture of A and C has more tendency to be stoop. On the other hand, comparing the value of normal walk, forward bending walk and swayback walk posture, person B’s posture has the tendency to swayback larger than other persons.

5. CONCLUSION

In this paper, we proposed a method of analyzing a pedestrian’s walking posture using a MY VISION system which processes the images provided from a chest-mounted camera. The walking posture data obtained by estimating the camera motion using DeMoN was processed by removing the trend. To evaluate the accuracy, a device containing a 9-axis sensor and a camera was manufactured and worn on the chest of a subject. The 9-axis sensor and the camera were synchronized and the data provided from them were analyzed. Experiments using five walk patterns by six persons were conducted in an outdoor environment and the results showed the effectiveness of the proposed method.

CONFLICTS OF INTEREST

The authors declare they have no conflicts of interest.

AUTHORS INTRODUCTION

Prof. Dr. Joo Kooi Tan



She is currently with Department of Mechanical and Control Engineering, Kyushu Institute of Technology, as Professor. Her current research interests include three-dimensional shape/motion recovery, human detection and its motion analysis from videos. She was awarded SICE Kyushu Branch Young Author’s Award in 1999, the AROB Young Author’s Award in 2004, Young Author’s Award from IPSJ of Kyushu Branch in 2004 and BMFSA Best Paper Award in 2008, 2010, 2013 and 2015. She is a member of IEEE, The Information Processing Society, The Institute of Electronics, Information and Communication Engineers, and The Biomedical Fuzzy Systems Association of Japan.

ACKNOWLEDGMENT

This work was supported by JSPS KAKENHI Grant Number 16K01554.

REFERENCES

- [1] D. Mehta, S. Sridhar, O. Sotnychenko, H. Rhodin, M. Shafiei, H.P. Seidel, et al., VNect: real-time 3D human pose estimation with a single RGB camera, *ACM Trans. Graph.* 36 (2017), 44.
- [2] S. Ishikawa, J.K. Tan, H. Kim, S. Ishikawa, 3-D recovery of a non-rigid object from a single camera view employing multiple coordinates representation, 2013 2nd IAPR Asian Conference on Pattern Recognition, IEEE, Naha, Japan, 2013, pp. 946–950.
- [3] J.K. Tan, S. Ishikawa, Deformable shape recovery by factorization based on a spatiotemporal measurement matrix, *Comput. Vis. Image Understand.* 82 (2001), 101–109.
- [4] V. Ganapathi, C. Plagemann, D. Koller, S. Thrun, Real time motion capture using a single time-of-flight camera, 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, IEEE, San Francisco, CA, USA, 2010, pp. 755–762.
- [5] B. Bodenheimer, C. Rose, S. Rosenthal, J. Pella, The process of motion capture: dealing with the data, *Computer Animation and Simulation ’97. Eurographics*, Springer, Vienna, 1997, pp. 3–18.
- [6] T. Shiratori, H.S. Park, L. Sigal, Y. Sheikh, J.K. Hodgins, Motion capture from body-mounted cameras, *ACM Trans. Graph.* 30 (2011), 31.
- [7] B. Ummenhofer, H. Zhou, J. Uhrig, N. Mayer, E. Ilg, A. Dosovitskiy, et al., DeMoN: depth and motion network for learning monocular stereo, 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Honolulu, HI, USA, 2017, pp. 5622–5631.
- [8] RTIMULib2, Available from: RTIMULib/RTIMULib2: <https://github.com/RTIMULib/RTIMULib2> (last accessed September, 2020).

Mr. Tomoyuki Kurosaki



He obtained B.E. and M.E. from Kyushu Institute of Technology. His research interests include computer vision, human posture estimation and neural network.