Walking Motion Recognition System by Estimating Position and Pose of Leg Mounted Camera Device Using Visual SLAM

Katsunori Tai
Kwansei Gakuin University
2-1 Gakuen, Sanda, Hyogo, JAPAN
fzp62983@kwansei.ac.jp

Yasuyuki Kono
Kwansei Gakuin University
2-1 Gakuen, Sanda, Hyogo, JAPAN
kono@kwansei.ac.jp

Abstract
This paper presents the method for recognizing the walking motion of the user with estimating position and pose of leg mounted camera devices by employing Visual SLAM. Approximate displacement of the position and orientation of the sensor terminal can be estimated using output values of the acceleration sensor, the gyro sensor, and the geomagnetic sensor. It is also considered that the position and orientation of the terminal can be estimated by the image captured by the camera mounted on each terminal. The system records surrounding scenery of the user with each camera device attached to each part of the user's leg, and estimates the 3D position and orientation of each camera device by using Visual SLAM. Finally, we develop the walking training support system that does not require a special environment.

Author Keywords
Localization; Sensor; Camera; SLAM; Android; Rehabilitation.

ACM Classification Keywords

H.5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous; H.5.2. Information interfaces and presentation (e.g., HCI): User Interfaces

Introduction

In this research, we have developed the system for recognizing the walking motion of user by estimating the 3D position and orientation of the camera terminals attached to the user's legs. Recently, as society ages, various rehabilitation technologies that support the healthy living of elderly people attract attention. Causes for elderly people to need nursing care include fractures and joint diseases, and these are often come from deterioration of physical function due to aging and deterioration of walking ability. For elderly people, walking training to prevent deterioration of walking ability is necessary for maintaining health condition and improving the quality of life. However, special walking training requires expert guidance and special environment such as training facilities. Our goal is a walking training support system that does not require a special environment. The rough flow of the entire system is shown in Figure 1. In order to recognize the walking motion of the user, we attach camera terminals to each part of the user's legs, and the system estimates the position and orientation of the terminal with the camera images photographed by each terminal and various sensor output values. In this paper, we mainly describe the recognition method of walking motion of the user.

Figure 1 Flow of the entire system

Related Works

Many methods for estimating the position and orientation of the camera from the camera images have been proposed. Hamada et al. [1] proposed a method to estimate the 3D position and orientation of a mobile camera by matching feature points extracted from images captured with both environmental cameras and a mobile camera. In this method, it is necessary to prepare environmental cameras that shoots the measurement range and the position and orientation estimation of the mobile camera is performed only within the space of the range that can be shot by the environmental camera. In this research, we aim to realize better recognition with only user camera without using environmental cameras. Some researchers proposed a method of estimating the displacement of the terminal by integrating the acceleration sensor values, but this method has problem that the error of the estimated position accumulates together with the time of measurement. Furthermore, in the method of this study, the orientation of the terminals attached to each part of the leg changes according to the movement of the leg.

Figure 2 Installation image of the system

Therefore, it is necessary to perform coordinate transformation from the local coordinate system based on the terminal to the world coordinate system and then to integrate the acceleration sensor values on each coordinate axis in order to estimate the displacement of the terminal based on the acceleration sensor values. Singhal et al. [2] succeeded in reducing this error by carrying out filtering processing with the high pass filter and the low pass filter on the acceleration sensor values and then performing coordinate transformation. However, it is difficult to correct the error to such an extent that the walking motion in the indoor can be recognized by this method. In this research, we use Visual SLAM and Google VR API for estimating the position and orientation of the terminal, thereby realizing robust walking motion recognition with high accuracy.

Hardware Setup

In this research, we employ Android smartphones (ASUS ZenFone 2 ZE 551 ML) as terminal for recognizing the walking motion of the user because it is relatively easy to realize the hardware requirements required by the system. This smartphone can shoot movies with resolutions up to 1,920×1,080 (Full HD), and it equips with an acceleration sensor, a geomagnetic sensor, and a gyro sensor. The walking motion of the user is recognized by estimating the three-dimensional position and orientation of the smartphone with the camera image and the output value of various sensors. The installation image of the proposed system is shown in Figure 2. Total of eight smartphones is attached to each part of both of the user’s legs. In order to fix each smartphone to the user’s legs, we made dedicated wearing belt. This wearing belt is made up of a nylon material belt and a flip cover for fixing the smartphone. The length of the belt can be adjusted by an adjuster attached to the belt so as to match the diameter of the user’s leg and the attachment site. A unit composed of a smartphone and a wearing belt is hereinafter referred to as "measuring unit". Figure 3 shows the measurement unit and its mounting image.

Walking Motion Recognition

In this research, the system estimate the position and orientation of each terminal by using the image around the user photographed by the camera terminal attached to each part of the user's leg and the output value of various sensors equipped with the terminal (Figure 4). Visual SLAM [3] is used to estimate the position of the camera from the images captured by the terminal attached to the user's right waist. SLAM (Simultaneous Localization and Mapping) is a general term for technologies that realize high-precision and high-speed self-location estimation by simultaneously performing environment map creation and self-location estimation. Among them, which using image information obtained from cameras and the like is called Visual SLAM. Examples of Visual SLAM include PTAM [4] by Klein et al. and LSD-SLAM [5] by Engel et al. In this research, we use ORB-SLAM [6] by Mur-Artal et al. to estimate the position and orientation of the camera. The ORB-SLAM is a feature point-based Visual SLAM that operates with a monocular camera and it adopts ORB feature detection as a method of detecting feature points in images. Using the ORB-SLAM, the system estimate the position and orientation of the

Figure 3 Measuring unit and its installation example

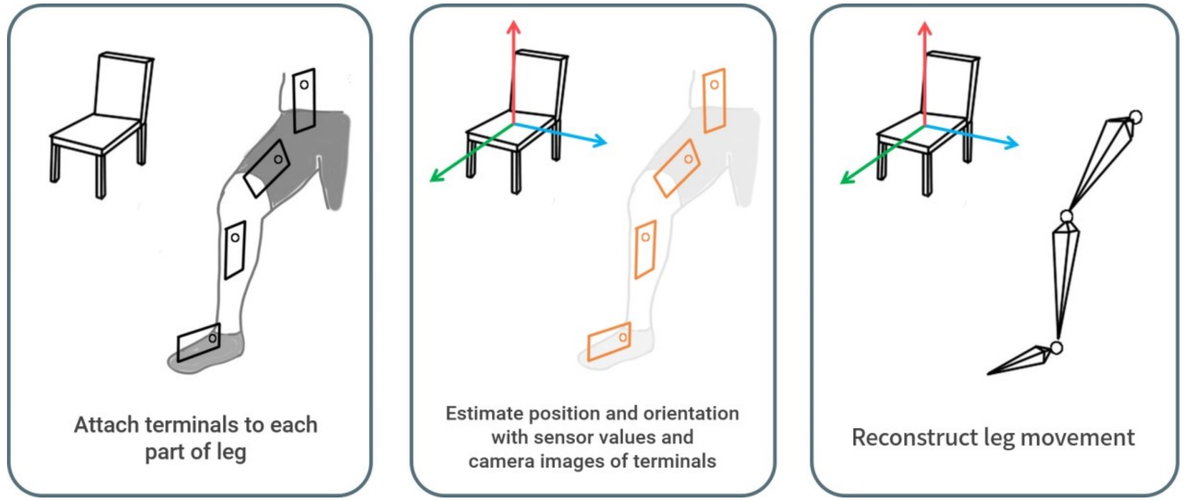

Figure 4 Schematic of measuring system

waist from the camera image shot by the camera terminal attached to the user's right waist. In addition, orientation estimation by the Google VR API is performed at each terminal. Google VR API is API for VR content developed by Google Inc. It can estimate pose using output values of acceleration sensor, geomagnetic sensor and gyro sensor. The azimuth angle of the terminal can be calculated from the geomagnetic sensor, and the tilt angle and the rotation angle can be calculated with both acceleration sensor values and gyro sensor values. The gyro sensor can acquire angular velocities of three axes of the terminal, and by integrating those with time, the rotation angle of each axis of the terminal can be estimated. However, this estimation contains a drift error, and it accumulates in proportion to the measurement time. This error can be corrected by using the estimated value of the terminal orientation derived from the gravitational acceleration obtained from the acceleration sensor. The system estimates the motion of the user's legs based on the self-position and orientation estimation result of each terminal and restores it as walking motion. For calibrating the system, the user assumes a stiffly erect posture as shown in Figure 5(A), and sensor readings from each smartphone terminal are recorded as initial posture values. After the calibration, the global position and posture of each terminal is continuously obtained by (1) calculating the differences of the sensor readings from the initial ones, and (2) consulting the current global position and posture of the terminal at user's waist calculated by Visual SLAM as in Figure 5(B). Then, the system reconstructs the walking motion of the user by

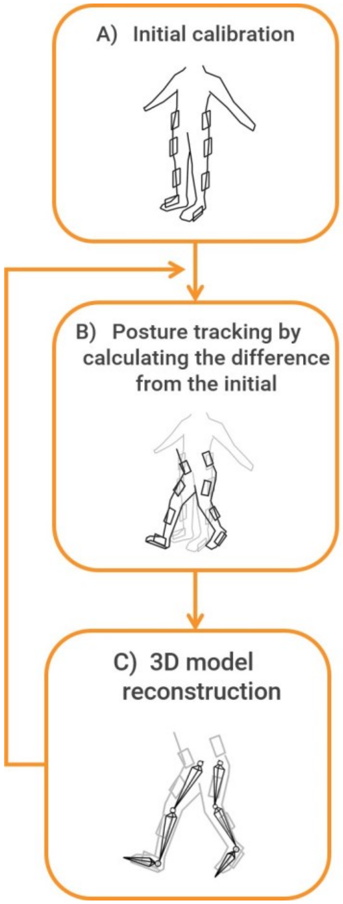

Figure 5 System flow

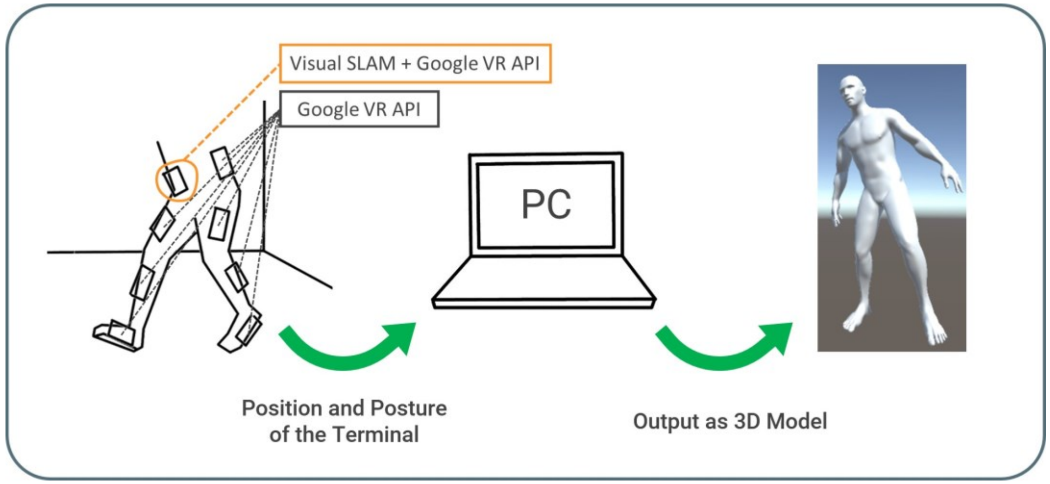

Figure 6 Schematic of demonstration system

reflecting the global position and posture of each terminal in the 3D model (Figure 5(C)).

Demonstration System Implementation

In this research, the system estimate the position and orientation of the terminal attached to each part of the user's leg by ORB-SLAM and Google VR API. We install an Android application for orientation measurement on each terminal. This application can transmit the orientation information of the terminal outputted by the Google VR API to the PC via Bluetooth and transmit the camera image photographed by the camera of the terminal to the PC via Wi-Fi. As a result, the orientation information of each terminal and the camera image photographed by the terminal attached to the right waist of the user are transmitted to the PC. The position of the user's right waist is estimated by performing ORB-SLAM processing on the image transmitted from the terminal. Then, from the estimated position and orientation information, it is restored as the motion of the leg starting from the position of the waist of the user. The motion of the leg is output as a 3D model by the demonstration system implemented on the PC (Figure 6). This demonstration system is implemented by Unity. The PC prepares a virtual socket for wireless communication for the number of installed terminals, and orientation information of the terminal received from each socket is reflected on each part of the leg of the 3D model. On the PC, Linux is installed as a virtual OS, and ORB-SLAM is operated on a tool called ROS (Robot Operating System). Then, the position of the waist of the user is estimated by applying ORB-SLAM processing to the received camera image.

Figure 7 System operation example

Concluding Remarks

In this research, we present the system to recognize the walking motion of user by estimating the position and orientation of the camera terminals attached to the user's legs. We also developed a demonstration system that restores the user's walking motion by estimating the position and orientation of the terminal attached to each part of the user's legs using Visual SLAM and Google VR API. Now, we have not verified the accuracy of walking recognition by this system yet. Therefore, we will verify the recognition accuracy of walking motion by this method by simultaneously conducting walking motion recognition by this method and walking motion recognition by optical motion capture system and improve the system more. We also aim to develop a system to support walking training of users by recognizing the walking motion of the user using the system proposed in this paper and guiding the walking exercise to the user.

References

1. Shuhei Hamada, Itaru Kitahara, Yoshinari Kameda, Yuichi Ohta. "Pose Estimation of a Mobile Camera by Using Sift Features in Environmental Camera Images". 12th Annual Conference of the Virtual Reality Society of Japan (VRSJ'07), pp. 514-517, September 2007.
2. Toshak Singhal, Akshat Harit, D N Vishwakarma. "Kalman Filter Implementation on an Accelerometer Sensor Data for Three State Estimation of a Dynamic System". International Journal of Research in Engineering and Technology (IJRET), vol. 1, no. 6, 2012.
3. Andrew J. Davison. "Real-Time Simultaneous Localization and Mapping with a Single Camera". Ninth IEEE International Conference on Computer Vision (ICCV'03), pp. 1403-1410, October 13-16, 2003.
4. Georg Klein, David Murray. "Parallel tracking and mapping for small AR workspaces". Sixth IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR'07), pp. 1–10, November 2007.
5. J. Engel, T. Schöps, D. Cremers. "LSD-SLAM: Large-Scale Direct Monocular SLAM". In European Conference on Computer Vision (ECCV), 2014.
6. Raúl Mur-Artal, J. M. M. Montiel, Juan D. Tardós. "ORB-SLAM: A Versatile and Accurate Monocular SLAM System". IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, October 2015.