

Semi-automatic calibration of 3D camera images - Monitoring activities made easy

Bart Jansen

Department of Electronics and Informatics
Vrije Universiteit Brussel
Pleinlaan 2, Brussel, Belgium
Email: bjansen@etro.vub.ac.be

Rudi Deklerck

Department of Electronics and Informatics
Vrije Universiteit Brussel
Pleinlaan 2, Brussel, Belgium
Email: rpdekler@etro.vub.ac.be

Abstract—A major practical problem when implementing home monitoring using 3D camera technology is the calibration process which serves to relate the camera coordinate system to a coordinate system interpretable by the caregiver. In this paper, we propose a calibration algorithm which can relate both coordinate systems without performing actual measurements of the position of reference points in the room. The calibration tool is constructed such that it can be operated by any caregiver and hence eases the deployment of the home monitoring appliances. Our method (semi)automatically detects a plane corresponding to the ground floor in the three dimensional images. Random points from that plane are used to calculate rotation angles in a least squares manner. Experiments are performed to investigate how the obtained results depend on the parameters of the algorithm.

I. INTRODUCTION

Over the last decade, a wide range of technologies were developed for supporting the life of the chronically ill or elderly in their home environment. Advances range from health monitors over fitness aids to amusement robotic toys. For instance for monitoring the health of the subject in its home environment, many research artifacts were turned into commercial products nowadays on the market: wireless blood pressure sensors, temperature sensors, glucose sensors, ... and many other medical sensors can be used to remotely monitor important health variables. Tools were also developed for monitoring activity levels of patients, as it is believed that activity levels also contain medically relevant information. Most tools for monitoring activities use accelerometers[1] or infrared sensors for deciding whether a person is in a specific area. Few tools use camera technology, as it is not always easy to obtain reliable spatial information[2], [3]. For instance, stereo vision is computationally expensive and has poor results in uniform regions. With the advent of 3D cameras[4] based on the *time-of-flight* principle, it is possible to overcome these problems.

In previous work, we have argued that 3D cameras are very well suited for activity monitoring of elderly in their home environment. However, in order to obtain spatial information which is expressed in an intuitive coordinate system (as for instance defined by the ground floor and the walls in a corner of the room), rather than in camera coordinates, a reliable calibration method is required. Good calibration relies on a

set of reference points for which coordinates in both the camera coordinate system and the target coordinate system are known. Practically speaking, this means that the position of the reference points in the room must be measured by the installer of the monitoring system at home. Although this approach is theoretically fine, in practice this is prohibiting the deployment of such a monitoring system in a nursing home or the home environment, as it requires a lot of measuring whenever the position of the camera changes.

In section two, our camera based activity monitoring approach is explained. In section three, the calibration problem and its traditional solution are explained. In section four, our semi-automatic calibration routine is discussed and in section five some experimental results are provided. Section six elaborates on automatic ground floor detection and section seven is the conclusion of this paper.

II. CAMERA BASED MONITORING

In previous work (e.g. [5]), we have introduced a system for monitoring activities of elderly using a 3D camera¹. The camera provides an intensity image and spatial information of the measured pixels. Using image processing techniques as filtering, background subtraction and ellipse fitting, the position of the monitored subject - expressed in a coordinate system defined by the camera - can be derived. Using an appropriate calibration method (cf. supra), this position is transformed into coordinates defined by the room. Hence, using the 3D camera, we know the position of the subject in the room, including its height above the ground. Using this information, we can derive features with medical relevance, such as (1) the distance walked per time unit, (2) the time the subject stays in bed, in a chair or elsewhere, (3) the number of walking episodes and (4) the number of transitions between predefined locations such as the chair, the bed, ...

Moreover, we have shown that the spatial information is of direct relevance for the detection of fall detection and other emergency situations. By applying a learnable context model, the amount of false positives can be reduced.

¹SwissRanger SR-3000™

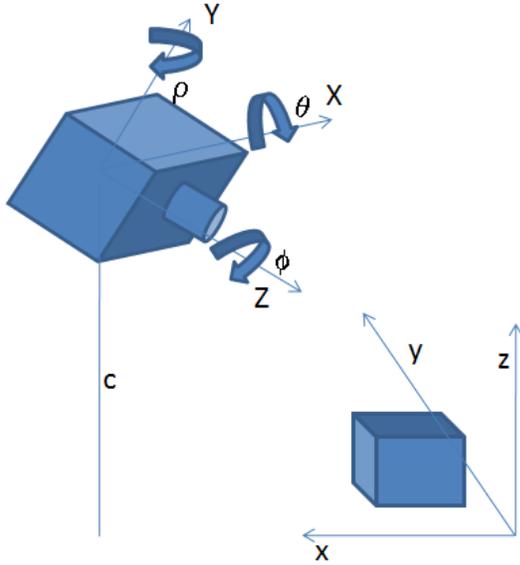


Fig. 1. The calibration problem in its pure form: how to relate points expressed in camera coordinate system (X, Y, Z) to points expressed in the world coordinate system (x, y, z) ?

III. CALIBRATION

In order to relate points in camera coordinate systems to room coordinate systems, a rotation and a translation have to be estimated. Typically, one collects a set of reference points known in both coordinate systems. They can be used to calculate a rotation and a translation for which the error on the reference set is minimal. Given $P_i = (X_i, Y_i, Z_i)$ in the camera coordinate system and $p_i = (x_i, y_i, z_i)$ in the room coordinate system, we look for the matrix M such that $MP \approx p$ where M is composed of a 3-by-3 rotation matrix R and a translation vector \vec{t} .

$$\begin{pmatrix} R & \vec{t} \\ 000 & 1 \end{pmatrix}$$

By rearranging terms, we have

$$A \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ \dots \\ m_{32} \\ m_{33} \\ m_{34} \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ \dots \\ x_n \\ y_n \\ z_n \end{pmatrix}$$

where A takes the form

$$\begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 \\ \dots & \dots \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 \end{pmatrix}$$

Using for instance a Single Value Decomposition, the least squares solution M^+ can be calculated.

IV. PSEUDO-AUTOMATIC CALIBRATION

This approach outlined above for mapping any P_i onto p_i is straightforward, but requires the reference points to be known in both coordinate systems. Below, we will outline a pseudo-automatic calibration method which can be performed by any layman using an intuitive graphical user interface and which does not require to measure reference points in the room. The proposed method is limited in the sense that it assumes that the camera is mounted in vertical position, which means that ρ and ϕ are zero (see fig. 1). In a first step, we provide a method for estimating θ , which is the angle between the XZ and xy plane (see fig. 1). Currently, we do not need to estimate the translation between the origins of both spaces, because our applications are targeted toward measuring distances in a given plane, which do not require absolute positions.

In our calibration method, we ask the human operator who is shown the intensity and depth image of the room in the calibration tool, to click on a set of k points which all are part of the ground floor of the room. All those points P_i share the property that after rotating them over the angle θ , their Z coordinate equals to c . So, we are looking for the angle θ which minimizes the following expression:

$$\sum_i^k \|x_i \sin \theta + z_i \cos \theta - c\|^2 \quad (1)$$

The angle θ can be recovered by solving the following equation:

$$\begin{aligned} (A^2 - E^2) \tan^4 \theta - 2(AC + BD) \tan^3 \theta + \\ (C^2 - 2A^2 - D^2 - E^2) \tan^2 \theta + \\ 2(AC - DE) \tan \theta + \\ A^2 - D^2 = 0 \end{aligned}$$

after applying the following substitutions:

$$\begin{aligned} A &= \sum_i x_i z_i \\ C &= \sum_i (x_i^2 + z_i^2) \\ D &= c \sum_i x_i \\ E &= c \sum_i z_i \end{aligned}$$

Only one of the solution provides a non complex (real) number in the first quadrant.

V. EXPERIMENTS

In this section, four experiments are described which serve to investigate the reliability of this approach. Three of the four experiments were performed using generated data, while the fourth experiment was performed using the camera.

The experiments with generated data serve to perform controlled parameter studies which are unfeasible with a real

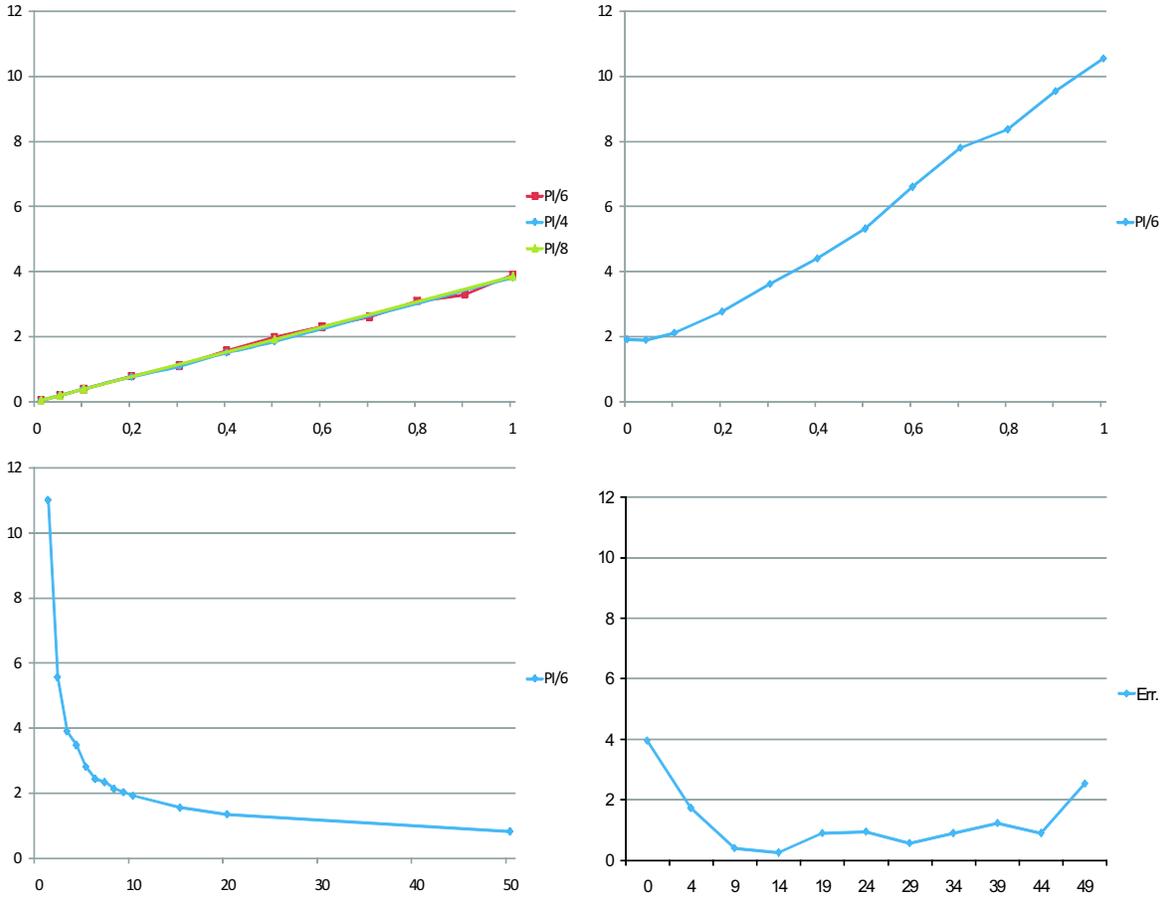


Fig. 2. Top left: the calibration error in degrees for different amounts of noise on the data points. Top right: the calibration error in degrees for different amounts of noise on the parameter c . Bottom left: the calibration error in degrees for different numbers of datapoints. Bottom right: the calibration error in degrees for different tilt angles of the 3D camera.

camera. 100 Random points on the ground plane are generated and rotated over a given camera angle. Those rotated points are fed into the calibration method and the calculated angle $\hat{\theta}$ is compared with the real rotation angle θ .

In an initial experiment, it is investigated how the accuracy of the proposed method scales with the quality of the reference points. Each reference point is corrupted with gaussian noise ($\mu = 0$) where σ is varied over the x-axis. This is investigated for several angles θ . Results are shown in top left graph of figure 2 and show that (1) the error scales linearly with the amount of noise and (2) the error does not depend on the actual angle θ . The second experiment serves to investigate the importance of a reliable measurement of the parameter c , which is the height of the camera above the ground. In order to investigate this, the actual value of c is corrupted with different amounts of gaussian noise ($\mu = 0$, σ is varied over the x-axis). Results are shown in the top right graph. The results clearly show the importance of an accurate measurement of this parameter. In the third experiment, it is investigated how the error in estimating the rotation angle scales with the number of reference points k . Results are shown in the lower left graph. The graph shows that the error is decreasing

exponentially with the number of points, such that certainly more than 10 points are required to obtain good results.

The last experiment is performed using the actual 3D camera. It was mounted on a pan/tilt unit² and the pan/tilt unit was steered toward the different tilt angles it supports. Images of a simple room were taken and a user indicated six different points on the ground floor. The difference between the calculated angle and the angle reported by the pan / tilt device is plotted in the lower right figure. Results show that for the supported tilt angles, calibration errors are around 1 degree, except for the outer points. This might be caused by inaccuracies in the pan/tilt device at its outer positions.

VI. SEMI-AUTOMATIC GROUND FLOOR DETECTION

The method as described above eases the calibration process in an important manner: rather than measuring a set of reference points in the room, one has to make one actual measurement (the height c of the camera above the ground floor) and one has to click a set of points on the ground floor. The lower left graph of figure 2 however shows that the calibration error decreases exponentially with the number

²TrackerPod™

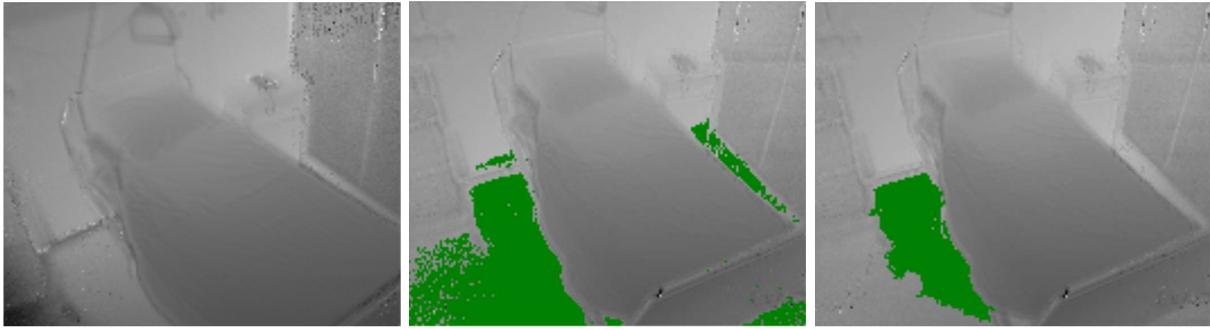


Fig. 3. On the left: the depth image of a typical hospital room. On the middle: the plane computed from three reference points (in green). On the right: the region computed from one seed point (in green).

of points. In order to obtain the best performance, more than twenty points on the ground floor have to be collected. This can further be reduced to only three points to be clicked by the operator. From the three points clicked by the operator, the equation of the plane through them can be calculated. For all points in the image, their distance to that plane is then calculated. Whenever the distance is less than 10 cm, the points are labeled as belonging to the ground.

Alternatively, only one point has to be clicked if that point is taken as a seed for a classical region growing process. A neighbouring pixel p' was grouped to the same region as pixel p if the difference in brightness is less than 10 (where the difference between white and dark is 255). An example of the performance of both approaches in finding the groundfloor of a typical hospital room in the depth images is shown in figure 3.

VII. CONCLUSION

We have provided a calibration method which allows for the estimation of the camera angle of a 3D camera, without measuring the position of reference points. This method acts as an important speedup in the installation of a 3D camera based activity monitoring tool in the patient's home environment.

With the calibration method we propose, only a single actual measurement must be performed (i.e. the height of the camera above the ground). In an intuitive graphical user interface, the operator has to click a set of points in the image which lay on the ground floor of the room. As the calibration error decreases exponentially with the number of selected points, the ground floor is extracted automatically from the images, given one up to three seed points.

REFERENCES

- [1] J. T. Cavanaugh, K. L. Coleman, J. M. Gaines, L. Laing, and M. C. Morey, "Using step activity monitoring to characterize ambulatory activity in community-dwelling older adults," *Journal of the American Geriatrics Society*, vol. 55, pp. 120–124, 2007.
- [2] H. Nait-Charif and S. McKenna, "Activity summarisation and fall detection in a supportive home environment," in *International Conference on Pattern Recognition (ICPR)*, Cambridge, 2004.
- [3] S. McKenna and H. Nait-Charif, "Summarising contextual activity and detecting unusual inactivity in a supportive home environment," *Pattern Analysis and Applications*, vol. 7, no. 4, pp. 386–401, 2004.
- [4] T. Oggier, M. Lehmann, R. Kaufmann, M. Schweizer, M. Richter, P. Metzler, G. Lang, F. Lustenberger, and N. Blac, "An all-solid-state optical range camera for 3d real-time imaging with sub-centimeter depth resolution," in *SPIE Proceedings*, St. Etienne, 2003, pp. 5249–65.
- [5] B. Jansen and R. Deklerck, "Context aware inactivity recognition for visual fall detection," in *Proceedings of the first international conference on pervasive computing technologies for healthcare 2006*, Innsbruck, Austria, 2006.