3D Corner Detection from Room Environment Using the Handy Video Camera

Ryo HIROSE †, Hideo SAITO † and Masaaki MOCHIMARU ‡

† : Graduated School of Science and Technology, Keio University, Japan {ryo, saito}@ozawa.ics.keio.ac.jp ‡ : Digital Human Research Center, AIST, Japan

Abstract Fatal accidents of young children in the house are serious problems in Japan in recent days. Especially a lot of children accidentally get injured by falling down to the sharp corners of furniture. Even though there are a lot of sharp corners in the indoor environment, it is not easy to be noticed by the parents' viewpoints. Therefore we propose a method for automatic 3D corner detection from the indoor environment just by capturing the scene with a handy digital video camera, so that the parents can easily find such sharp corners in their rooms for avoiding the accidental injure.

1 Background

Fatal accidents of young children in the house such as degradation, choking by mistake, drowning at bathrooms etc. - are serious problems in Japan. The numerous efforts to prevent such accidents have been researched [1, 2]. We particularly focus on the degradation accidents, because a lot of young children accidentally get injured by falling to sharp corners of furniture. Even though there are a lot of sharp corners in the indoor environment as shown in Fig.1, it is not easily to be noticed by the parents' viewpoints.



Fig.1 Example of Sharp Corners Existing in the Room Environment

For avoiding the accidental injury by the degradation, it is important to create the safe environment for children to help prevent injury when they fall down on the floor or fall from the high place. However their parents can not easily recognize the potential risk behind the room environment, because they believe such kind of corners are not dangerous for their children.

The purpose of this research is automatic detection of sharp 3D corners in the indoor environment by capturing the scene with a handy digital video camera, so that the parents of children can easily find such sharp corners in their rooms for avoiding the accidental injure of their children. The spatial distribution of such sharp corners in the room can also be used for behavior analysis of children. In this paper, we propose a method for automatic 3D corner detection from the room environment using the captured video sequence by a handy video camera. The points that are recognized as shape corners are finally displayed onto the input video sequence. Fig.2 shows the overview of our system.



Fig.2 System Overview

2 Proposed Method

2.1 Outline

Fig.3 shows the outline of the proposed method. The system is divided into two phases, "Constructing Feature Point Database Process" and "3D Corner Detection Process". The first process detects the candidates of the sharp corners in the scene. The second process is employed to distinguish the sharp corners from the candidates.



Fig.3 Outline of Proposed Method

2.2 Constructing Feature Point Database

This section describes the first stage of our method, which creates the database of 3D feature points. This database keeps information related to a number of the reconstructed feature points. For each feature point, the following information is preserved: the recovered 3D position of the feature point, the positions and orientations of the camera frames that track the feature point, and the 2D positions of the feature point in the camera frames.

To construct the feature point database, 3D reconstruction of the environment is required. The framework of estimation of 3D coordinate is presented in Fig.3 (Left Side).

We utilize the AR Tag marker [3, 4] which is the fiducial marker system that consists of a library of markers and the computer vision software for detecting these markers from input images. The camera position and orientation in each frame can easily be obtained by utilizing the AR Tag marker.

On the other hand, the natural feature points are

detected and tracked by using the KLT (Kanade-Lucas-Tomasi) Feature Tracker [5]. In addition to the forward-tracking method, the back-tracking method is also utilized to reduce the miss-corresponding points.

Then 3D positions of each feature point can be estimated by using the structure from motion techniques [6]. If the 3D reconstruction is failed, such feature points are not registered in the feature database and they are rejected from the candidates of the 3D corner points.

2.3 3D Corner Detection

This section describes the method to detect the 3D corners from whole feature points registered in the feature database.

The key idea for distinguishing the feature points in 3D is to compute the local texture fluctuation of the reconstructed feature points.

Fig.4 indicates the difference of the visual appearance around the feature point depending on the camera position and orientation. Fig.4(a) shows the positions of the two kinds of the feature points existing either on the planar area or on the sharp corner. A red circular point indicates the feature point extracted from the planar texture on the object's side face. A black crossed line indicates the 3D corner located at the object's vertex. When the camera moves around the objects in the direction of the arrow in the figure, the texture images around these feature points are changing as shown in Fig.4(b) and (c).



Fig.4 Difference of Visual Appearance

- (a) Positions of the target feature points.
- (b) Local texture around the circular point.
- (c) Local texture around the crossed line.

In case that a feature point exists on the planar area, the local texture doesn't change much as

shown in Fig.4(b). Furthermore, these small differences of visual appearance can be approximated by the affine transformation. In contrast, in case of a 3D feature point such as object's vertices, the local texture significantly changes as shown in Fig.4(c). These changes of the visual appearance can not be approximated by any planar transformation.

From such observation, we can have an assumption as follows. If a feature point exists on the planar area, the local texture fluctuation must be small. On the other hand, if a feature point is located at the 3D corner, the local texture fluctuation must be large. Consequently, we can distinguish the sharp 3D feature points by checking difference of the visual appearance around the feature points based on such assumption.

To calculate the fluctuation of local textures around the feature points, we consider the search planar window in 3D space to compare the local texture for same areas. The search planar window is constructed at a reference frame by projecting the local texture image around the feature point to the plane which is vertical to the viewing direction of the feature point from the camera. The local texture around the feature point is projected onto the search planar window in every frame. If the local shape of the feature point is planar, the texture fluctuation on the search planar window will be small. Finally the local texture fluctuation is calculated and 3D corners can be distinguished by finding the feature points whose local texture fluctuation is large.

2.3.1 Creating Search Planar Window

The search planar window is defined in the reference frame as shown in Fig.5.

First, the normal vector of the feature point is defined to determine the search planar window. This is assumed the unit vector with the direction from the feature point to the camera position. The plane of the search planar window on is defined as the vertical plane to this normal vector.



Fig.5 Constructing the Search Planar Window

Next the local texture size on the image plane is defined. The local texture size is changed depending on the distance between the projection center of the camera and the 3D position of feature point, because the search planar area size in 3D is constant. This local texture image is back projected onto the world coordinate system by the camera position and orientation in reference frame. Thus the search planar window is constructed.

2.3.2 Projection Local Texture Image onto Search Planar Window

Once the search planar windows of each feature point are constructed in reference frame, the local texture image around the feature point in every frame is projected onto the search planar window by obtained projection matrix as shown in Fig.6(a). Fig.6(b) shows both the position of the local texture image on the image plane (Blue area) and the projected image onto the search planar window (Red Area).



2.3.3 3D Corner Distinction

3D sharp corners are selected from all feature points by considering the fluctuation of the local

texture around each feature point based on our assumption. The fluctuation is evaluated by the correlation values of the local textures on the search planar window between the reference frame and the other frames. The zero-mean normalized cross-correlation R_{ZNCC} can be calculated in Eq.(1).

$$R_{ZNCC} = \frac{\sum (I_{1 < x, y >} - \bar{I}_1)(I_{2 < x, y >} - \bar{I}_2)}{\sqrt{\sum (I_{1 < x, y >} - \bar{I}_1)^2 \sum (I_{2 < x, y >} - \bar{I}_2)^2}}$$
(1)

After the calculation of the correlation values against all candidates of sharp corners, the local texture fluctuations are checked by using the variance of the correlation values. If the variance is large, the feature point is determined as the 3D corner. On the other hand if the variance is small, the feature point is determined as a feature point existing on the planar area.

3 Experiments

We have conducted the several experiments in the room environment to validate the proposed method. We employ one AR Tag marker to obtain the camera position and orientation. The world coordinate system is attached to this marker coordinate system.



(c) Scene.3

Fig.7 The Experimental Room Environments

We applied our proposed method to the following environments. Scene.1 and 2 are same room environments as shown in Fig.7(a) and (b), but these two sequences are different from the

shooting date and time and the camera path of captured video sequences are also substantially different. Scene.3 is captured in a reception room environment as shown in Fig.7(c). These input video sequences are captured with a handy video camera, SONY DV Handy Cam DCR-TRV-900, whose resolution is 720×480 pixels.

Fig.8 indicates the input images and the feature points detected by KLT Feature Tracker. In this experiment, input video sequence consists of 60 frames in which 60 feature points are detected. The blue rectangular points indicate the feature points whose 3D coordinate can not be calculated, and the red circular points indicate the candidates of sharp corner whose 3D coordinate can be reconstructed. The total number of the candidates of sharp corner is 39 points to which the 3D corner detection procedure is applied.



Fig.8 Detected Feature Points (Scene.1)

Fig.9(a) shows the example of the feature point existing on the planar area. In this case, the feature point is detected from the texture of the book cover. Fig.9(b) shows the example of the 3D sharp corner. Fig.9(c) and (d) indicate the result graphs of the correlation value fluctuation for Fig.9(a) and (b), respectively. The fluctuation of the feature points on planar area become small while the one of the feature points on 3D sharp corners become large. In this way, by considering the result graphs, we can distinguish the 3D corners from all candidates



Fig.9 Correlation Value Change Graph (Scene.1)

- (a) Example of feature point on planar area
- (b) Example of 3D corner point
- (c) The result graph of (a).
- (d) The result graph of (c).

of sharp corner.

Our system detects six 3D sharp corners out of 39 candidates of sharp feature points. These points recognized as the 3D corners are displayed onto the input video sequence as shown in Fig.10.

Next, we applied the proposed method to Scene.2 and Scene.3. The input image and the detected feature points are shown in Fig.11 and Fig.13. The white crossed lines and the red circular points respectively indicate the unregistered feature points and the registered feature points which 3D coordinates can be reconstructed successfully. The detected 3D corners are superimposed onto the input video sequences as shown in Fig.12 and Fig.14. Scene.2 consists of 45 frames and 26 feature points out of 35 detected feature points are registered in the feature database. At last 8 feature points are recognized as the 3D corners. Scene.3 consists of 60 frames. And 19 feature points out of 45 detected feature points are registered in the feature database, and 8 feature points are recognized as the 3D corners.



Fig.10 Detected 3D Corners (Scene.1)

4 Conclusion

An automatic method for 3D corner detection from the room environment using the captured video sequence by a handy video camera is proposed in this paper.

The proposed method utilizes the presupposition of the visual appearance difference, which is that the local appearance around the 3D corner such as the object's corner is drastically changed depending on the camera movement. We create the feature database by reconstructing the detected feature points at first. Then the search planar spaces are created in the world coordinate system. By comparing the fluctuation of projected search windows, we can determine whether the 3D corner or not.

References

- [1] Tokyo Fire Department, http://www.tfd.metro.tokyo.jp/hp-seianka/kojiko/
- [2] Safe Kids Network Japan, http://safekids.ne.jp/childaccidentreport/
- [3] http://www.artag.net
- [4] Fiala Mark. AR Tag Revision1. A Fiducial Marker System Using Digital Techniques. In National Research Council Publication 47419/ERB-1117, November 2004.
- [5] Stan Birchfield. Derivation of Kanade-Lucas-Tomasi Tracking Equation. in Technical Report, 1997.
- [6] Richard Hartley and Andrew Zisserman. Multiple View Geometry in Computer Vision. Cambridge University Press, second edition, 2002.

Hirose, Ryo: received the B.E. degree in the department

of information and computer science from Keio University, Japan, in 2005. He has been a graduate student in Keio University since 2005. His research interests include computer vision and augmented reality. Saito, Hideo: received B.E., M.E., and Ph.D. degrees in Electrical Engineering from Keio University, Japan, in 1987, 1989, and 1992, respectively. He has been on the faculty of Department of Electrical Engineering, Keio University, since 1992. Since 1997 until 1999, he had been a visiting researcher of the Robotics Institute, Carnegie Mellon University. Since 2006, he has been an Professor of Department of Information and Computer Science, Keio University.

Mochimaru, Masaaki: is the deputy director at the Digital Human Research Center of the National Institute of Advanced Industrial Science and Technology (AIST). He received a Ph.D. from the graduate course of Biomedical Science, Keio University in 1993. His interests are related in modeling of human body shape and motion.



(a) Frame 10 (b) Frame 30 (c) Frame 45 Fig.11 Detected Feature Points by KLT Feature Tracker (Scene.2)



(b) Frame 30 Fig.12 Detected 3D Corners (Scene.2)

(c) Frame 45



Fig.13 Detected Feature Points by KLT Feature Tracker (Scene.3)



(b) Frame 20 Fig.14 Detected 3D Corners (Scene.3)