Toward Automatic Reconstruction of 3D Environment with an Active Binocular Head

Chung-Yi Lin∗, Sheng-Wen Shih†, Yi-Ping Hung‡
∗ Department of Computer Science and Information Engineering,
National Taiwan University, Taipei, Taiwan
† Department of Computer Science and Information Engineering,
National Chi Nan University, Nantou, Taiwan
‡ Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan

E-mail: hung@iis.sinica.edu.tw

Abstract

This paper presents a new automatic approach to reconstructing a model for the 3D environment by use of an active binocular head (the IIS head). To efficiently store and access the depth estimates, we propose the use of the inverse polar octree. The depth estimates are computed by using the asymptotic Bayesian estimation method. The path of the local motion required by the asymptotic Bayesian method is determined online automatically to reduce the ambiguity of stereo matching. Some rules for checking the consistency between the new observation and the previous observations have been developed. Our experimental results have shown that the proposed approach is very promising for automatic generation of 3D models which can be used for rendering a 3D scene in a virtual reality system.

1. Introduction

Because of the rapid growth of VR applications in the last few years, automatic generation of 3D models from images has attracted much attention recently. One popular approach to reconstructing 3D models from images is to use the stereo vision techniques. However, it is well known that stereo correspondence is a very difficult problem. Fortunately, when using image-based techniques to render a 3D scene with texture extracted from real images, the quality of the rendered image is more tolerable to inaccurate 3D data. Knowing that image-based techniques do not require highly accurate 3D reconstruction and that the stereo correspondence problem by itself is a difficult problem, our goal is to reconstruct an approximate textured 3D model using a large number of images taken with an active binocular head such that this approximate model can be used to render images which look similar to the real ones when observing from arbitrary viewpoints within certain viewing area.

2. Inverse Polar Octree

With our approach, 3D data are accessed repeatedly, during reconstruction, for examining the consistency of the 3D estimates observed from different views. Hence, we use the voxel-based octree to represent 3D data because it can provide efficient access of the stored 3D data—the complexity of accessing a voxel in an octree is $O(\log N)$. To use the voxel-based 3D representation, we have to first deal with the problem of packing the 3D information contained in the infinite 3D space into the finite memory space of the computer. Our solution to this problem is to take an *inverse spherical transform* before quantizing the estimated 3D data into voxels. The inverse spherical transform is described in the following: (1) Transform the 3D Cartesian coordinates, (x, y, z), to a spherical coordinate system, (ρ, θ, ϕ). (2) For the 3D spherical coordinates, (ρ, θ, ϕ), compute its inverse polar coordinates, $(\frac{1}{\rho}, \theta, \phi)$.

There are two major advantages of taking the inverse spherical transform. The first one is that, after the transformation, all the surrounding 3D objects farther than a minimum distance to the observer, say R_{min}, will be enclosed within a sphere with radius $\frac{1}{R_{min}}$. The second advantage is that, after taking the inverse spherical transform, we can apply a uniform quantization because the estimation error is now bounded [3]. After taking the inverse spherical transform, an octree is created in the spherical coordinate system to maintain the uniform angular resolution. The octree will be referred to as the inverse polar octree(IPO) in this paper.
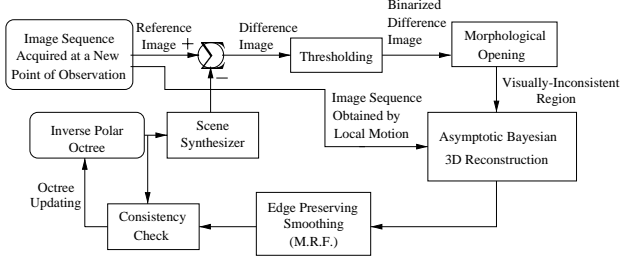
Figure 1. The schematic diagran of our active 3D reconstruction process.

3. Automatic 3D Reconstruction

The schematic diagram of our active 3D reconstruction process is shown in Figure 1. We assume that a well-calibrated active binocular head [5] is available for exploring and reconstructing the 3D environment. For more details of the reconstruction process, please refer to [3].

3.1. Visually-Inconsistent Regions

To save the computation, we apply the asymptotic Bayesian estimation only to *visually-inconsistent regions* defined below. First, an image is synthesized for a given viewpoint according to the current IPO. Next, this synthetic image is subtracted from the corresponding observed image to obtain a difference image. The visually-inconsistent regions are obtained by thresholding the difference image and then applying morphological opening to remove noise. These regions indicate where the depth information are either incorrect or not available

3.2. Depth Estimation

The depth estimation method we used in this work is mainly the asymptotic Bayesian estimation method [2]. which can be thought of as incremental-baseline stereo or multiple-baseline stereo [4]. Instead of directly determining the stereo correspondence using the left and right images, we first move the left camera locally and incrementally in order to compute a rough estimate of the depth of the surface patch. According to our analysis, 50 millimeters of incremental local motion in our setup can reduce the depth uncertainty to an extent such that the search region for stereo correspondence is less than 10 pixels. After rough depth estimation with local motion, our system will use the image taken by the right camera as the new input image of the asymptotic Bayesian estimation process (i.e., a big jump) and perform an exhaustive search in the search region followed by a gradient descent search to further refine the depth

estimate. Once the depth estimates of all the patches in the visually-inconsistent regions are computed with the above process, Markov random fields can be used to smooth the depth map while preserving the depth discontinuity [1].

3.3. Path Planning for Local Motion

When the orientation of an edge is parallel to its corresponding epipolar line, finding stereo correspondence of that edge becomes an ill-posed problem. To minimize the ambiguity in stereo matching, the local motion is selected to be the one having its epipolar lines perpendicular to most of the edges that still have uncertain depth estimates.

3.4. Consistency Check for a New 3D Observation

Since the stereo correspondence may contain some false matching, consistency check before integrating new depth estimates into the existing octree is necessary. We say that the new observation p_{3D} is *compatible* with a previous observation either if p_{3D} has already been "seen" by the previous observation and the color difference between the new observation and the previous observation is less than a threshold, or if p_{3D} is occluded by the previous observation from the viewpoint of the previous observation. If a large proportion of the previous observations are compatible with the new observation, p_{3D}, then p_{3D} is said to be *largely consistent* with the previous observations and can be used to update the IPO; otherwise it is discarded. The updating of IPO concerns with removing the existing data that are not compatible with the new observation, p_{3D}.

4. Experiments

In our experiments, the binocular head is mounted on an X-Y table which is used for simulating a mobile robot platform. In this section, we show how a complex scene in our laboratory can be reconstructed with the active binocular head. The relative position of the objects and the 20 viewpoints chosen in advance are shown in Figure 2. At each viewpoint, the depth value of visually-inconsistent regions is estimated by using the algorithm described in section 3. The progress of the reconstruction is shown in Figure 3 where the virtual camera is set to be slightly overlooking the scene. In Figure 4, two virtual viewpoints were chosen to compare the synthetic image and the observed image.

5. Conclusions

Based on our well calibrated binocular head, we have developed an active stereo vision algorithm which can estimate the 3D depth automatically and can integrate 3D data
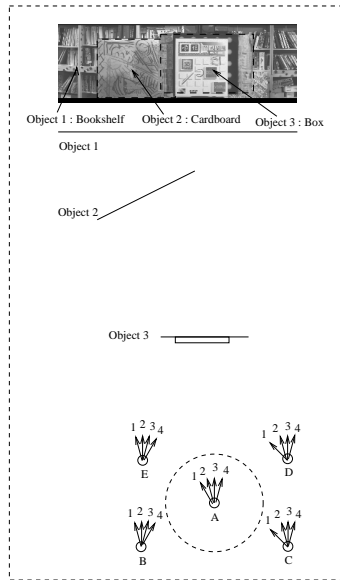
Figure 2. The relative position of the objects and the 20 viewpoints, $A1$, $A2$, .., and $E4$.

obtained from different points of observation. Real experiments have shown that the proposed algorithm is feasible and promising for many applications.

Acknowledgements

References

[1] D. B. Cooper, J. Subrahmonia, Y.-P. Hung, and B. Cernuschi-Frias. *The Use of Markov Random Fields in Estimating and Recognizing Objects in 3D Space Markov Random Fields: Theory and Applications, edited by Rama Chellapa and Anil Jain*. Academic Press, 1993.

[2] Y.-P. Hung, D. B. Cooper, and B. Cernuschi-Frias. Asymptotic bayesian surface estimation using an image sequence. *International Journal of Computer Vision*, 6(2):105–132, 1991.

[3] C.-Y. Lin, S.-W. Shih, and Y.-P. Hung. Toward automatic reconstruction of 3d environment with an active binocular head. Technical Report TR-IIS-98-004, Institute of Information Science, Academia Sinica, Nankang, Taipei, Taiwan, 1998.

[4] M. Okutomi and T. Kanade. A multiple-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(4):353–363, 1993.

[5] S.-W. Shih, Y.-P. Hung, and W.-S. Lin. Calibration of an active binocular head. *to appear in IEEE Transactions on Systems, Man and Cybernetics–Part A:Systems and Humans*, 28(4), July 1998.
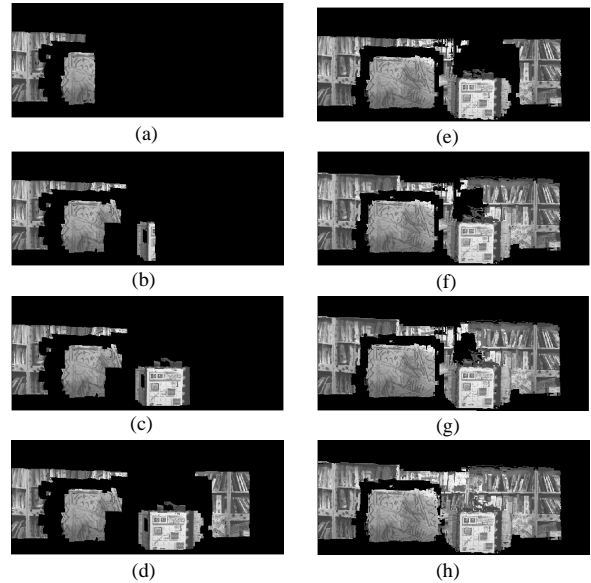
(a)

(e)

(b)

(f)

(c)

(g)

(d)

(h)

Figure 3. The scene is synthesized from a virtual viewpoint slightly overlooking the scene, where (a), (b), (c), (d), (e), (f), (g) and (h) are synthesized by using the IPO reconstructed after taking observations at viewpoints $A1$, $A1$–$A2$, $A1$–$A3$, $A1$–$A4$, $A1$–$B4$, $A1$–$C4$, $A1$–$D4$, and $A1$–$E4$, respectively.
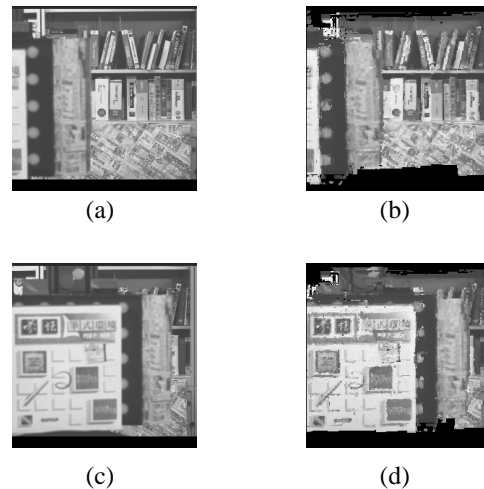

(a)

(b)

(c)

(d)

Figure 4. (a) The real image captured at viewpoint V which is located between viewpoints A and C. (b) The synthetic image generated for V by using the reconstructed IPO (c) The real image captured by viewpoint U located between viewpoints A and B. (d) The synthetic image generated for U by using the reconstructed IPO.