3D Imaging of indoor environments with mobile photo-camera

A. O. TRUFASU, A. N. CUTA[a], A. TRUFASU[b]
Universitatea Politehnica Bucuresti, Bucuresti, 050098, Romania
[a]Universitatea Politehnica Bucuresti, Bucuresti, 050098, Romania
[b]Gunnar Optiks,705 Palomar Airport Road, Suite 100, Carlsbad, CA, 92011, California, USA

This paper presents an algorithm for full 3D imaging of indoor environments with mobile shooting camera. Data is acquired by slow-moving photo-camera equipped with two 2D laser range finders. Present approach combines an efficient scan matching routine for camera pose estimation with an algorithm for approximating environments using flat surfaces.

(Received May 4, 2011, accepted August 10, 2011)

Keywords: Stereo-vision, Virtual reality

1. Introduction

The topic of learning 3D models of large objects and small man-made objects has received considerable attention over the past few years, mostly due to vending methods over internet. 3D models are useful for a range of applications. For example, car manufacturers may use 3D models for design and utility studies using virtual reality (VR) technology. 3D models are also useful for robots operating in indoor environments.

And finally, accurate 3D models could be a great supplement to the video game industry, especially if the model complexity is low enough for real-time VR rendering.

In the literature, approaches for 3D mapping can be divided into two categories:

- approaches that assume knowledge of the pose of the sensors, and:
- approaches that does not.

In the present paper, authors were interested in using photo camera for data acquisition; hence authors' approach falls into the second category due to the inherent noise in robot behavior. Authors focused on generating low-complexity models that can be rendered in real-time.

The majority of existing systems also requires human input in the 3D modeling process. Here we are interested in semi automated modeling with less human interaction. Present approach is also related to, which reconstructs planar models of indoor environments using stereo vision, using some manual guidance in the reconstruction process to account for the lack of visible structure in typical indoor environments.

This paper presents an algorithm for generating simplified 3D models of indoor environments. The data for generating these models are acquired by a mobile photo-camera, equipped with two 2D laser range finders. The first laser scans horizontally and is used for focus estimation (like autofocus). The second laser is pointed upwards so that it scans the three-dimensional structures while the photo-camera moves through its environment. To estimate the point of focus of the shooting camera while collecting the data, a 2D scan matching algorithm is used that is reminiscent of the literature on mobile robot mapping. The resulting pre-filtered data is globally consistent but locally extremely noisy. A recursive surface identification algorithm is then employed to identify large flat surfaces, thereby reducing the complexity of the 3D model significantly while also eliminating much of the noise in the sensor measurement. The resulting 3D models consist of large, planar surfaces, interspersed with small fine-structured models of regions that cannot be captured by a flat-surface model.

The models studied in the computer graphics literature are usually assumed to be noise-free; hence, simplifications are only applied for increasing the speed of rendering, and not for the reduction of noise. This has important ramifications, as the noise in the data renders a close-to-random fine structure of the initial 3D models. Secondly, built structure is known to contain large, flat surfaces that are typically parallel or orthogonal to the ground. Such a prior is usually not incorporated in polygon simplification algorithms. Consequently, a comparison with the algorithm presented in illustrates that our approach yields significantly more accurate and realistic-looking 3D models.

2. Concurrent imaging and localization in 2D

2.1 2D Scan alignment

The process of taking 3D photos is really very simple, and the basics can be explained in less than a minute, but to become good at taking and presenting 3D photos take a bit more time, and it's something that really develops with practice.

Taking the Photos

✓ Always take photos in portrait orientation;

✓ Set the camera to full manual and choose the correct exposure and focus;

✓ Camera needs to be oriented to subject and balancing the body from right foot to left one are those two times to shoot first and second photos;

✓ Simply by shifting the weight from one foot to the other, viewpoint is moved by several millimeters, sufficient to get a 3D effect.

Processing

The two photos need to be one and it comes easy using a software called Stereo Photo Maker. It's not the prettiest software, but it does a fantastic job. It can be used in conjunction with a plug-in called Auto Pano, that can analyses the two images and automatically correct for many of the problems that can come from shooting two separate images. This includes tilting and twisting, moving forward or back between shots, and the "keystone distortion" that occurs when you turn the camera to centre the subject for close 3D photos.

Step by step

Drag both photos onto the Stereo Photo Maker, zoom out a bit if need to make it easier to fuse the pair into 3D with the cross-eye technique.
If the 3D effect seems reversed, swap the images correctly for a crossed eye view, auto align to correct for any distortions, fix the 3D images position relative to the 3D window.

The 3D "Window"

The edge of the image is more than just the boundaries of the 3D photograph. In a 3D photo, it is also a "hole" into which you look and through which 3D subjects can appear. A good way to think of the edge of the image is as a literal window in your computer screen. This is one of the reasons why I find a border around both parts of the 3D image helps me, it more clearly defines the edge of the 3D window.

Fig. 1. Two pictures taken from different angles from close up.

Just like a real window, you expect to look through it, and rarely do you expect things seen beyond it to come back through it at you. One nasty optical illusion that can happen with 3D photos is when part of the 3D subject "touches" the window, or worse, appears to overlap it. Have a look at the two examples below.

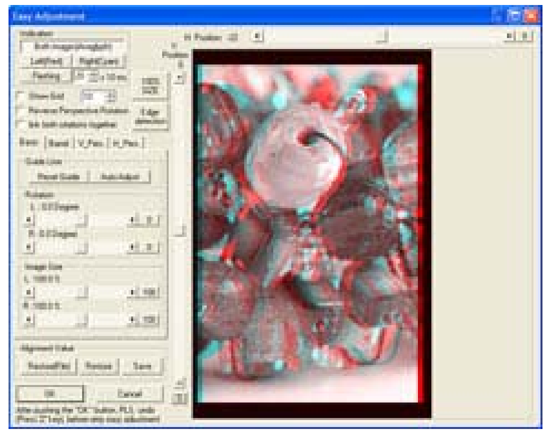

Fig. 2. Stereo photo Maker easy adjustment.

If the object seems to be too far, by software This problem is easily corrected with the "Easy Adjustment" button. Clicking on this will show both images overlapped and tinted red and blue (if you have a pair of red/blue 3d glasses, you can do this process in 3D!). Using the slider above the image, you can adjust the separation of the two images, thus moving them backward and forward in 3D space.

Method presented above has some:

***Advantages*:**

- Each eye's image is captured on a full frame, so the resulting 3D image can be very high resolution
- There is no blurring or ghosting at the edge of the frame, which can be seen in many "beam splitter" attachments where the two views join
- You can take a 3D photo with any lens in your SLR kit, including macro, for extremely close 3D photos
- Many 3D attachments have very limited control over focus and aperture, with the technique you have complete control over the settings
- You can do this with any camera, if you forget to take your 3D attachment or camera, you can still take 3D photos this way.

Disadvantages:

- The most obvious and critical shortcoming is that this method only works with still object that don't move between shots
- Any movement between the two shots will cause a distracting 3D error, so people, animals and even trees in a light breeze will be difficult or impossible to shoot

- You need to take two photos for every 3D image, which takes twice as long, and uses twice as much space
- It is easy to introduce errors such as twisting or tilting the camera between shots which can cause distracting artifacts
- You need to shoot with manual settings so that there's no accidental variation in exposure or focus.

Based on advantages and trying to eliminate or diminish disadvantages authors proposed following procedure.

The first step towards building 3D images with mobile robotics is a 2D pose alignment procedure.

Our algorithm estimates poses in 2D using a real-time scan matching algorithm.

Search in Pose Space

Clearly, when aligning a scan to one or more previously collected scans, the total log-likelihood depends on the pose of the scan, where *pose* refers to the scan's x-y-coordinates together with its orientation. Exploiting the differentiability of our log-likelihood function, scans are aligned relative to previously recorded scans by adjusting the pose in proportion to the negative gradient:

$$< x, y, \theta > \leftarrow < x, y, \theta > \vartheta \frac{\delta E}{\delta < x, y, \theta >} \qquad (1)$$

where E denotes the total log likelihood (a sum over all measurements of a scan), and Θ> 0 is a step-size parameter.

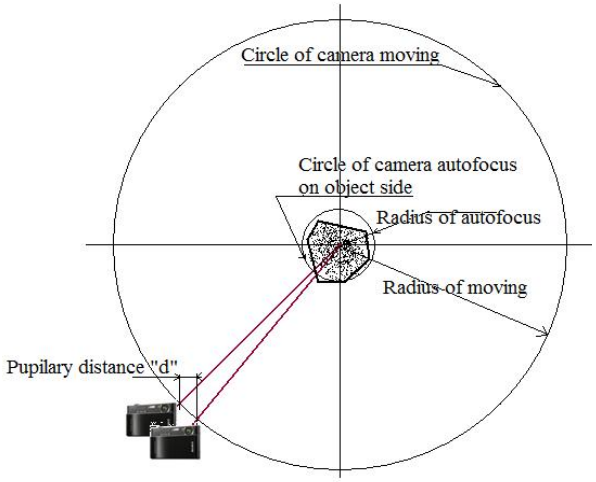

Fig. 3. Aligning two scans.

Fig. 3 shows an example of aligning two scans, using the gradient ascent scan matching routine. In this example, the two scans are initially misaligned by translational errors of 1cm in each coordinate axis, and a rotational misalignment of 3 degrees—these errors exceed practical errors by a factor of 10. After 100 iterations, the scans are aligned with sufficient accuracy for our 3D modeling task.

To perform the scan alignment in real-time, our approach pre-computes occlusion and all necessary distances necessary for calculating gradients using a fine-grained 2-dimensional grid. After this pre-computation, which takes approximately 0.1 seconds, each iteration of gradient ascent requires in the order of 1ms on a low-end PC—which is fast enough to align scans accurately as the robot moves. Fig. 2 illustrates the accuracy of the acquiring two consecutively images.

Generating 'Raw' 3D Images

The 3D images are generated using the upward-pointed laser, as shown in Fig. 3. While the robot traverses and maps an unknown environment in 2D, thereby recovering its pose, the upward pointed laser scans the 3D structure of the environment. We then obtain a polygonal model by connecting consecutive 3D points. To avoid closing wholes, we only create a close surface if the consecutive points are close to each other.

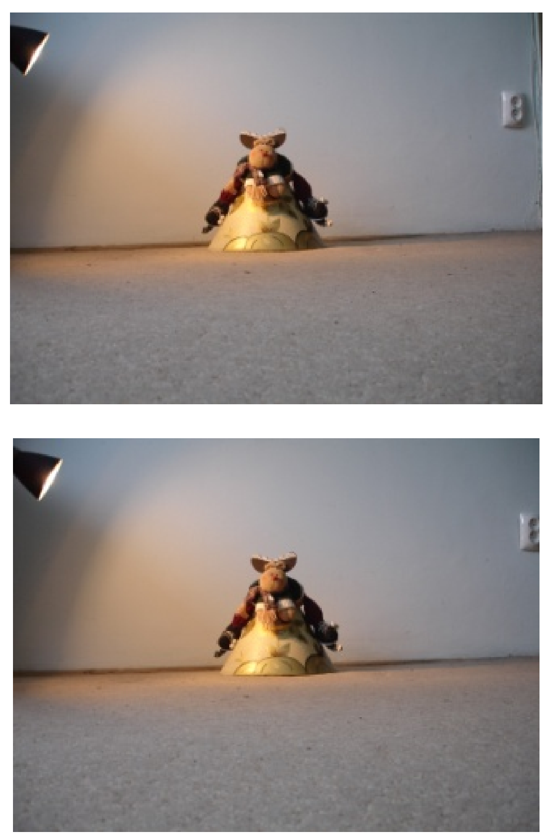

Fig. 4. Up and down images taken with respect the algorithm

Table 1. Angle dependency on radius and pupillary distance.

PD „d"	Focus radius	Angle (rad)	Focus radius	Angle (rad)	Focus radius	Angle (rad)
55	1000	3,1528	2000	1,57	3000	1,050955
56	1000	3,2101	2000	1,60	3000	1,070064
57	1000	3,2675	2000	1,63	3000	1,089172
58	1000	3,3248	2000	1,66	3000	1,10828
59	1000	3,3821	2000	1,69	3000	1,127389
60	1000	3,4394	2000	1,71	3000	1,146497
61	1000	3,4968	2000	1,74	3000	1,165605
62	1000	3,5541	2000	1,77	3000	1,184713
63	1000	3,6114	2000	1,80	3000	1,203822
64	1000	3,6687	2000	1,83	3000	1,22293
65	1000	3,7261	2000	1,86	3000	1,242038
66	1000	3,7834	2000	1,89	3000	1,261146
67	1000	3,8407	2000	1,92	3000	1,280255
68	1000	3,8980	2000	1,94	3000	1,299363
69	1000	3,9554	2000	1,97	3000	1,318471
70	1000	4,0127	2000	2,00	3000	1,33758
71	1000	4,0700	2000	2,035	3000	1,356688
72	1000	4,1273	2000	2,06	3000	1,375796
73	1000	4,1847	2000	2,09	3000	1,394904
74	1000	4,2420	2000	2,12	3000	1,414013
75	1000	4,2993	2000	2,14	3000	1,433121
76	1000	4,3566	2000	2,17	3000	1,452229

As it can be seen from table 1, angles vary significantly with papillary distance and radius too, which imply very accurate measurement of focus distance.

3. Acquiring smooth 3D models

Although the approach described above produces accurate position estimates, the resulting models often lack the appropriate smoothness.

Fig. 5. Stereo images in red/blue overlapping.

Fig. 5 shows, in detail, 3D picture as it appears by overlapping left and right images using stphmkre432 software. As is easily seen, the 3D effect is extremely suggestive. For experiments we did over 300 photos varying all parameters one by one. Whereas some of the ruggedness arises from remaining errors in the pose estimation, the majority of error stems from measurement noise in laser range finders. However, the key characteristic here is that all noise is local, as the scans have been globally aligned by the 2D mapping algorithm. As a result, global structures cannot be extracted by considering small areas of the data. Rather, one has to scan larger fractions of the model in order to find such structures.

Our approach proceeds as follows. It starts with a random point v_1 and its nearest neighbor v_2. It then repeatedly tries to extend the current set of points by considering all other points in increasing distance from this point set. Suppose v_1 is a point such that the papillary distance is 55 mm and radius r = 1000 mm. After a set of points of shooting we changed papillary distance by 1 mm keeping radius the same; at value of 77 mm of papillary distance, radius changed to 1500 mm.

Fig. 6. 3D effect by shooting around the object.

As it can be seen in figure 6 3-D effect is very realistic and most important suitable for any type of vision.

This approach has been implemented and tested using two different platforms both in indoor environments. As pointed out above, robot was equipped with laser-range scanners. Whereas the angular resolution of the laser used on the indoor system was 0.5 degrees.

The resolution of the measured distances is 1cm and the measurement error of these systems lies between 0 and 20cm

The experiment was carried out in one laboratory at the Politehnica University of Bucharest. Here the robot traveled half of circle from wall to wall around object.

Obviously, the quality of our photos is significantly higher than expected. Apparently, our approach provides accurate approximations of the planar structures and computes models with a seriously lower complexity than the QSlim system.

4. Conclusions

We have presented an algorithm for acquiring 3D pictures with mobile robots. The algorithm proceeds in two stages:

First, the robot pose is estimated using a fast scan matching algorithm. Second, 3D data is smoothed by identifying large planar surface regions. The resulting algorithm is capable of producing 3D maps without manual intervention, as demonstrated using data sets of indoor scenes.

The work raises several follow-up questions that warrant future research. Most importantly, the current 3D acquiring pictures are limited to small and medium objects. As a consequence, the resulting way is still fairly complex.

Additionally, an interesting question concerns robot exploration.

The issue of robot exploration has been studied extensively for building 2D maps, but we are not aware of robot exploration algorithms that apply to the full three-dimensional case. This case introduces the challenge that the robot cannot move arbitrarily close to objects of interest, since it is confined to a two-dimensional manifold. Finally, extending this approach to multi-robot and arbitrary outdoor terrain are worthwhile goals of future research.

References

[1] P. Allen, I. Stamos, In Proc. of the IEEE International Conference on Robotics & Automation (ICRA), p.1435, 2000.

[2] R. Bajcsy, G. Kamberova, L. Nocera, In Proc. of the 4th IEEE Workshop on Applications of Computer Vision, 2000.

[3] S. Becker, M. Bove, In Proc. of the SPIE Symposium on Electronic Imaging, San Jose, 1995.

[4] P. Debevec, C. Taylor, J. Malik, In Proc. of the 23rd International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), 1996.

[5] J.-S. Gutmann, W. Burgard, D. Fox, K. Konolige, In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 1998.

[6] S. Hakim, P. Boulanger, F. Blais, In Proc. of the 4th Conference on Optical 3-D Measurement Techniques, 1997.

[7] O. Khatib, The International Journal of Robotics Research **5**(1), 90 (1986).

[8] M. B. L. Iocchi, k. Konolige, In Proceedings of the 2000 International Symposium on Experimental Robotics, Waikiki, Hawaii, 2000.

[9] Y. Liu, R. Emery, D. Chakrabarti, W. Burgard, S Thrun, In Proc. of the International Conference on Machine Learning (ICML), 2001.

[10] F. Lu, E. Milios, Journal of Intelligent and Robotic Systems, 1998.

[11] D. Luebke, C. Erikson, In Proc. of the 24rd International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH), 1997.

*Corresponding author: a_trufasu@yahoo.com