5K−7

Reliable Localization of a "Flying-TV-Camera" in 3D for Shape Reconstruction

D. Ritter, Y. Matsumoto, K. Sugimoto, T. Arakawa
Tokyo Information & Communication Research Center
SANYO Electric Co., Ltd.

1) Introduction

Three-dimensional (3D) data is not used any more only in specialized applications like in reverse engineering or CAD, but is as well wide spreaded in 3D telecommunication, multimedia, internet catalogs or video games. Therefore our goal is to develop a fast, robust and skill-free 3D sensor (delivering shape and texture of objects) which can be operated easily by everybody without having a special hardware. We introduce a method which allows to get the 3D-surface of objects just by taking some single 2D-images (e.g. using a still video camera) of arbitrary viewpoints around the object. First the position of the camera belonging to each image of the sequence is computed relative to the object, then the 3D surface can be evaluated.

2) Overview over basic methods for shape reconstruction

To get the all-around surface of 3D-objects, several "input" images have to be combined. A new input image has to be adjusted to the already existing ones which are called the "reference" images. We may distinguish the different methods using 2D camera data (e.g. video images) or optical 3D sensors [1]. Let us assume first to use 2D input images which are easy to achieve but need more complex algorithms to compute the 3D-shape. Then we can separate between reference based or object based algorithms. Reference based methods use either an external reference, e.g. a computer controlled turntable [2] which allows to compute the 3D-surface, or a pad with known 3D geometry to localize the camera relative to the object [3]. This information can be used e.g. with a stereo algorithm [4] to get the desired 3D shape. Using an object based method, we can adjust the input image to a partially known 3D-object using the shading information [5]. Even without a 3D information, 2D images (with unknown camera position) can be fitted to compute the 3D shape [6]. Instead of such template based methods, feature based algorithms can be applied as well as used in photogrammetry.

Compared to 2D data, 3D (range) input images delivers a much higher precision, but need a more complex sensor hardware. Such sensors can be mounted on robots, then their positions are known and melting the data can be done easily. Instead of such reference based methods, more flexible object based algorithms can be used [7]. Feature oriented algorithms are first extracting markers or intrinsic object features which are used to compute the relative location between the 3D sensor positions. Then the range images can be transformed into a common coordinate system. Template based registration techniques can be used as well, they don't need special features and can adjust the range images automatically if their locus don't differ too much.

3) The algorithm

With respect to computing speed, accuracy and required hardware we decided to use 2D video images as input, an algorithm for a feature based camera localization (in 3D) using a reference pad, and a following shape from silhouette or stereo algorithm. This seems to be the best trade off for our purpose. In this paper we want to focus on the camera localization. The object to be digitized must be placed onto a reference pad or reference solid. Now the camera, taking the single images, has to be localized relatively to the pad. A model based localization algorithm is used for it where the pad is stored as a 3D-model first. Attached 3D-features like points, lines or center of ellipses are used for its representation. Therefore the pad features, which can be seen in the input video image, have to be extracted first. To localize the camera, the pad features of the 2D input image and the 3D reference model have to be combined. Two problems must be treated: first, the correspondence of features. We are using a Hough strategy to select the correct combinations and to ensure a robust evaluation even for many false detected or missing features. Second, the computation of the location parameters is solved using the locus of corresponding features in the input image and the reference model. In practice it is necessary to make use of a perspective projection model. An implicit indexing is applied to restrict the possible feature combinations. At least a matching technique is implemented to verify the location parameters. To increase the accuracy (which is limited by the finite Hough-tables which indicate the location parameters), in a following step an algorithm for nonlinear-least-square fitting is used. All in all this leads to a robust and reliable localization of the camera even using point-like features without any characteristics.

4) Design of the experiment and results

Fig.1 shows the 3D-model of the reference pad as stored in "zero position" in the computer. The position of its features (center of circles) are marked with crosses. To speed up the algorithm and to get a more reliable localization, we use features with some different colors (named 1,2,3,4). The object to be digitized has to be placed onto this pad. Fig.2 shows a snapshot of the input scene from an arbitrary camera position, where the feature positions are extracted in a preprocessing step. Combining the loci of input and reference features leads to the camera location in all 6 dimensions (rotation α,β,γ, translation x,y,z). The result can be seen in Fig. 3, where the reference pad is transformed with the detected camera location parameters and is overlaid with the camera image to check the result. In this experiment, the accuracy of the rotation angles is 0.075deg, the lateral shift 0.04mm and the axial shift 1.8mm. Usually we use about 30 input images for our shape from silhouette algorithm.

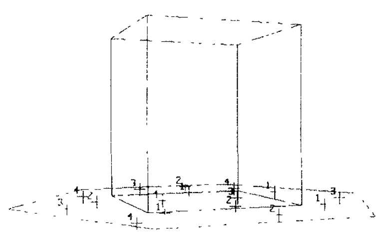
Fig.1

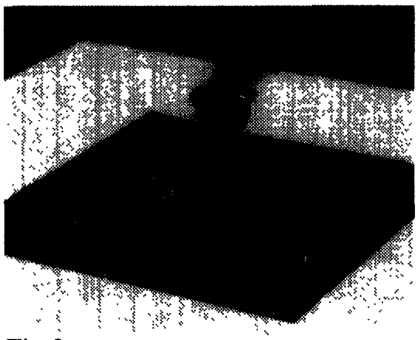
Fig.2

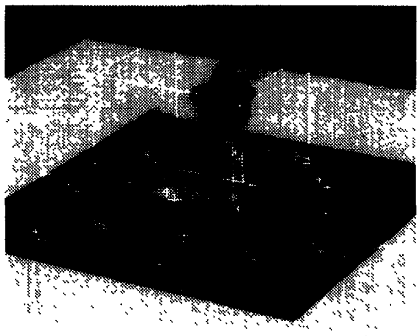
Fig.3

5) Conclusion

Several experiments showed that the localization of the camera is robust even for incomplete and wrong detected features. The remaining location error depends on the accuracy of the feature extraction, therefore a subpixel interpolation is used to get the center of the features. The location uncertainty does not accumulate because a reference pad is used with known 3D geometry. A following shape from silhouette algorithm [2] makes use of this camera location parameters to compute the complete all-around-shape and texture of objects. The only thing to do is "walking" around the object, taking some images and start the fully automated algorithm. Usually the shape from silhouette algorithm requires a constant background. That means that the object has to be placed in a black box while getting the 2D-image sequence, or that the object (together with the reference pad) must be rotated in front of the camera. In future we will add a stereo algorithm to avoid this disadvantage. Further, this will allow us to get the 3D shape not only of relatively small objects but of an interior or building too.

6) References

[1] R.G. Dorsch, G. Haeusler, J.M. Herrmann, "Laser triangulation: fundamental uncertainty in distance measurement", Applied Optics, Vol.33, No.7, March 1994

[2] Y. Matsumoto, H. Terasaki, K. Sugimoto, T. Arakawa, "A Portable Three-dimensional Digitizer", IEEE Proc. Int. Conf. on Recent Advances in 3-D Digital Imaging and Modeling, Ottawa, 1997

[3] D. Ritter, "Feature Based Object Recognition and Localization in 3D-Space Using a Single 2D Image", http://www.physik.uni-erlangen.de/Optik/haeusler/people/dr/orl/orl_e.html, 1995

[4] S.B. Marapane, M.M. Trivedi, "Multi-Primitive Hierarchical (MPH) Stereo Analysis", IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol.16, No.3, March 1994

[5] C.-E. Liedtke, R. Koch, "Modelling of 3D scenes from the Analysis of Stereoscopic Image Sequences", http://www.tnt.uni-hannover.de/org/papers/1993/, 1993

[6] E. Steinbach, S. Chaudhuri, B. Girod, "Robust Estimation of Three-Dimensional Motion and Structure of Multiple Objects from Image Sequences", http://www-nt.e-technik.uni-erlangen.de/~steinb/publications/publications.html, 1996

[7] D. Ritter, "3D-matching of range images to synthesize the whole object surface", http://www.physik.uni-erlangen.de/Optik/haeusler/people/dr/matching/matching_e.html, 1996