Augmented reality using range images

Christian Schütz, Heinz Hügli

Institute of Microtechnology, University of Neuchâtel
2000 Neuchâtel, Switzerland

ABSTRACT

This paper proposes range imaging as a means to improve object registration in an augmented reality environment. The addressed problem deals with virtual world construction from complex scenes using object models. During reconstruction, the scene view is augmented by superimposing virtual object representations from a model database. The main difficulty consists in the precise registration of a virtual object and its counterpart in the real scene. The presented approach solves this problem by matching geometric shapes obtained from range imaging. This geometric matching snaps the roughly placed object model onto its real world counterpart and permits to update the virtual world with the recognized model. We present a virtual world construction system currently under development that allows the registration of objects present in a scene by combined use of user interaction and automatic geometric matching based on range images. Potential applications are teleoperation of complex assembly tasks and world construction for mobile robotics.

Keywords: augmented reality, range imaging, geometric matching, virtual world construction

1. INTRODUCTION

One of the major problems in the field of augmented reality (AR) is the registration of virtual and real objects. The object registration problem is elementary to all AR applications since the information which augments the real world information has to be in correspondence with it. For example an aim has to follow the real target as it moves. In this paper, we propose the use of range images and geometric matching to solve this problem for a virtual world construction application. Let us first give an overview of typical characteristics of AR apart from the registration problem.

In general, AR is defined by following characteristics: combination of real and virtual data which are registered and updated in real time. Real data is acquired with a sensor as for example a CCD camera whereas virtual data stands for the additional information provided by a database or other knowledge. In the case of video images the virtual data may be a wireframe object representation which enhances the real object boundaries or instruction text which assists the operator during some task. The virtual data used in AR allows to augment the operator's perception and understanding of a scene in applications as medical visualization, orientation, guidance and assembly.

In contrast to virtual reality (VR), AR does not focus on realistic scene rendering but on accurate matching of the virtual and real world objects. Usually AR aims not at a seamless integration of virtual objects in a real world as done in some famous movies, but enhances the senses of the user with information he can not perceive otherwise.

There exists several techniques to combine the virtual and real data in one representation. Although AR can also be applied to audio or other information, we limit the following considerations to visual data. We can classify the available devices which combine real and virtual images or views into two broad categories: optical or video technology. The optical approach typically uses see-through displays that place an optical combiner in front of the user's eyes. This provides a relatively simple and effective way to add visual information from the virtual world to the human view [5, 11]. However, since the virtual world is generated in general with a simple camera model and projected by a monitor system the combination with the real world considering its different lightening conditions is problematic and the virtual object in the resulting representation may be of low contrast and is not necessarily in focus with its corresponding real world object [1]. The video approach provides an escape to this problem: Here, digital hardware is used to combine the real world image scanned by a

Christian Schütz, Heinz Hügli, "Augmented reality using range images", SPIE Photonics West, The Engineering Reality of Virtual Reality, San Jose, Proc. SPIE, Vol. 3012- IV, Feb. 1997

CCD camera and the rendered virtual image. Video or even stereo displays present the resulting image to the user [8, 9].

An other advantage of the video approach is the possibility to verify the registration since both the real and virtual representation are available to the computer. In optical systems the virtual information is projected onto the real world by estimating the user's view point with head position tracking devices. If there is no sensing of the real world, the virtual world generation runs in open loop mode and there is no mean to verify the registration result. Video technologies however allow to compare the real and virtual image by image processing methods: A closed loop control of the object registration provides more accuracy.

Once the real world image is available not only the verification but also the calculation of the registration can be performed with image processing techniques. Proposed systems use different types of image features to locate objects and the registration methods range from manual to fully automatic [3, 5, 8, 9]. For the following investigations, we split the task referred generally in AR as registration into the following sub-problems: localization, recognition and matching. The object localization and recognition problems are discussed in section 3. Section 4 presents the matching problem in more details.

The application we aim at is virtual world construction with an AR interface as described in section 2. Real-time system answer is of minor concern since the scene is static. However, an accurate matching is necessary to get a correct representation of the real world. We propose a new semi-automatic approach which combines human perception for object identification with computer matching accuracy. We use a range finder to measure the scene geometry and a geometric matching algorithm performs the accurate model matching, as described in section 5.

2. VIRTUAL WORLD CONSTRUCTION

The construction of virtual worlds has several application fields. Virtual worlds are used for instance in architecture to represent building geometry or in virtual reality robotics to program assembly tasks. Once the virtual representation of a real world is constructed the operator's view is not limited any more to the camera observing the real world. The virtual world provides a copy of the real world where the user has free inspection and interaction possibilities. There are other advantages as for example in a virtual robotics environment where the real robot is only used for the final task execution which saves energy and robot resources [14].

AR provides a comfortable way to construct virtual worlds in an interactive manner. The operator constructs the virtual world object by object by using some point and click interface while the growing virtual world is continuously registered to the real world. AR provides an intuitive way to see the progression of the construction. For example, the complete interior of a building can be constructed by using a mobile robot moving around and sending video images from the different rooms. The remote operator identifies walls and doors and places the corresponding objects on top of the real image. The computer registers then the virtual rooms with the real world and allows the operator to verify the result.

We can distinct different levels of real world knowledge [9] ranging from simple to complex and from fully modeled to completely unknown. They lead to different reconstruction methods. With some method simple unknown polyhedral worlds are constructed interactively by drawing lines along the object edges and relying on grouping the complete object is formed. Other methods focus on the construction of block worlds where cubes are aligned manually [8]. Methods to construct virtual worlds built of more complex objects rely on a model database. Here, the object types present in the real scene are known from a model database, as for example in assembly or medical tasks where the model information is accessible from a CAD database or computer tomograph scans. The choice of the representation strongly influences the degree of object complexity.

For our application, we assume that the objects are modeled and complex. We dispose of a object database containing objects represented by their geometry which may be of any shape. Objects are registered to the real world which allows a stepwise construction of the complete scene by adding one object after the other. The problems which occur during registration are discussed in the following sections 3 and 4.

3. LOCALIZATION AND RECOGNITION PROBLEM

During registration the object which has to be augmented with some virtual data has first to be localized in the real world. Depending on the application there exist several techniques to perform this task. Visual marks put on the object are easily located applying traditional segmentation techniques to the video image [5]. More flexible systems which allow several objects to be present in the scene search for geometric features in the video image. Features as for example object edges are detected and grouped into one object representation [3].

As soon as we deal with objects of more complex shape the definition of pertinent and reliably detectable features becomes more difficult [4]. The object localization is even more difficult if the objects are placed on a complex background formed by a pile of objects or textured material. We presented a localization method dealing with objects on complex background using range change detection which however needs observations over several video acquisitions and a controlled environment [13].

If the knowledge database contains several types of objects the correct model has to be determined for every located object in the scene. This object recognition is a quite difficult computer vision task especially if the objects are of arbitrary shape and when the database contains a large number of different models.

To avoid the localization and recognition problem, interactive systems for the virtual world construction have been presented recently [8, 9]. Here the operator locates and identifies the object and its type which results in a large flexibility since any object type can be treated. We will use this approach for object localization and recognition since the use of human perception has the following advantages: high score in estimating kind and rough pose of an object. A high quality interface is crucial for successful and ergonomic operator task completion. Stereo vision systems provide the necessary depth perception and allow the operator to manipulate and locate objects successfully in 3D space.

4. MATCHING PROBLEM

Accurate matching of the virtual and real world is crucial for a successful virtual world construction. There exist a lot of methods to match the virtual object to the real world. Many rely on visual features like points or line segments. As pointed out in the previous section, visual marks are easily detectable and a unique disposing on the object leads directly to the matching transformation after some correspondence establishment and error minimization [5]. Similar methods apply when the matching is based on other image features as edges for example [3].

The pose alignment of 3D objects by matching image features fails in presence of edgeless, smooth objects. In order to cope with any object form, we propose in the following section the use of geometric matching based on range images. This method uses directly the measured 3D shape information of the scene and does not rely on visual features. Therefore, the matching is very accurate and only limited by the range finder resolution.

Once the virtual object is matched it has to be projected on the real world image. In virtual world construction the observed objects are static and the camera movements well controlled. Therefore the camera position can be estimated or is even known exactly. Camera calibration as used in photogrammetry and robotics determines the necessary transformation matrix between the virtual world and the camera image space.

5. GEOMETRIC MATCHING USING RANGE IMAGES

We present a system for the construction of virtual worlds by registration of objects available in a database with a real view of the world. Being semi-automatic the system combines the advantages of interactive and automatic systems discussed in the previous sections. It assigns the object detection and identification task to the operator and uses the computer for fine pose matching.

5.1 Video and range data

Our system uses a video camera and a range finder to sense the real world. We use a range finder system working on the principle of space coding with projected stripe pattern and triangulation fabricated by ABW in Germany. This system measures the intensity and depth information of the observed scene. The depth measurements are stored in a 2D array called range image. Fig. 1 shows the intensity and the range image for a typical scene containing two tape dispenser parts. Object points which are near to the camera are bright in the range image.

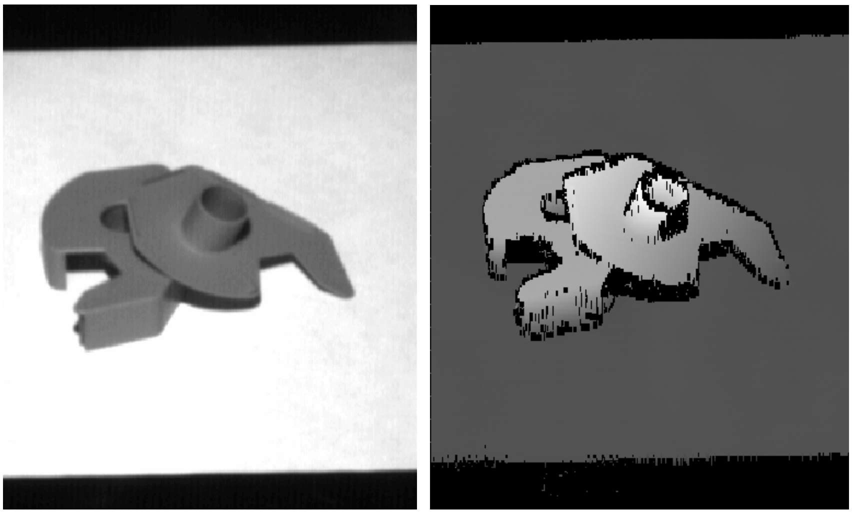

Fig. 1. Intensity and range image of two tape dispenser parts

Since the camera is calibrated, the 3D space coordinates for every surface point in the scene can be determined from the range image. A 3D representation of the real world may be obtained by applying a local triangulation method to the range image. Every range image pixel is compared to its neighbors and connected via triangles if the neighbor value is valid [12]. The resulting triangulated surface can then be combined with the intensity image and observed in a scene renderer as shown in Fig. 2.

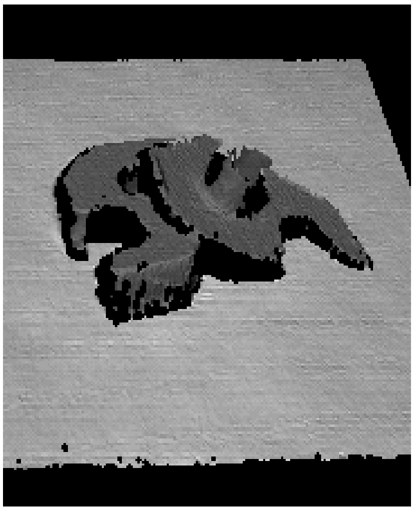

Fig. 2. Triangulated scene surface

The use of a range finder has several advantages compared to a stereo vision system. The knowledge of the depth information helps to combine the virtual and real world objects in a correct manner. A virtual object placed behind a real world object will disappear since the render software has access to the 3D information of the scene. This is not the case for uncalibrated stereo vision systems often used in AR applications.

5.2 Interactive localization

As we mentioned before, human perception is used for the object localization and recognition task. In the current system the user selects the correct model type from the database and places the model object roughly at the corresponding place in the scene.

To generate the model database, two different techniques have been used. The first one uses 3D geometric databases of different commercial CAD packages. The second technique relies on the use of a range finder where acquisitions from different view angles are fused into one object representation.

Fig. 3 shows the 3D interface presented to the operator. The interface has been programmed using the OpenInventor library running on a SiliconGraphics Indigo2 workstation. The scene is rendered and displayed on a stereo monitor and can be observed through stereo glasses.

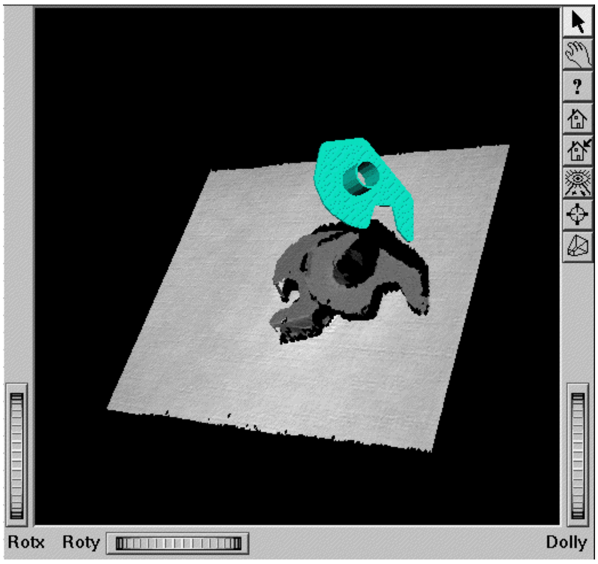

Fig. 3. User interface for object localization combining real and virtual objects

A space mouse allows to change the position and orientation of the virtual object. Therefore, it can be easily manipulated in all six degrees of freedom. Furthermore, the camera view can be changed which allows to inspect the scene from different view angles and improves object localization.

5.3 Automatic matching

Once the operator has oriented and placed the identified model at the approximate location in the real world presented on the monitor, the computer then automatically does the accurate matching using a geometric matching algorithm. This algorithm transforms the roughly placed model to a pose which fits best the measured data in the real scene. It calculates the optimal rigid transformation by minimizing some distance measure between the two objects. The error calculation uses the 3D information provided by the model and the one measured with the range finder.

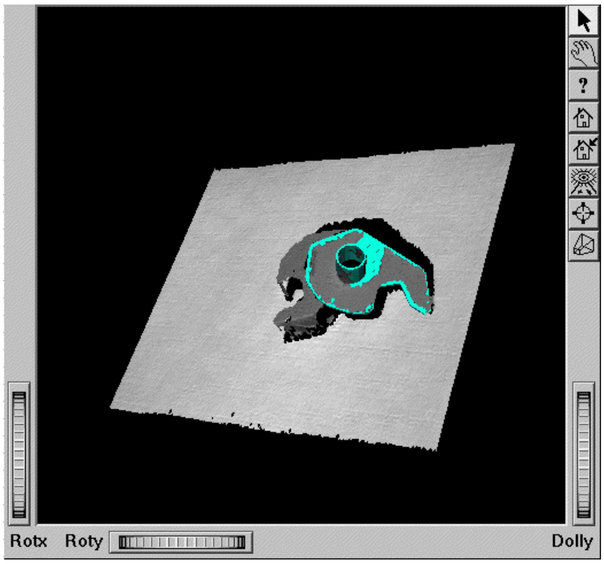

Fig. 4. Matched virtual object

Registration is illustrated in Fig. 4 which shows the matched virtual object added to the real scene after the geometric matching process. The augmented scene shows a step of the virtual world construction process. The final virtual world after the construction process is shown in Fig. 5. This virtual world representation can now be used for example for the task programming of a robot doing an assembly task of the two tape dispenser parts [14].

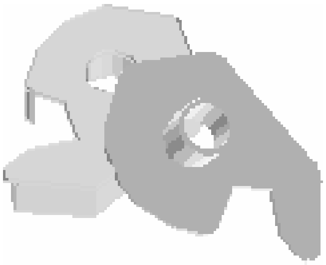

Fig. 5. Constructed virtual world

The geometric matching performed by the system is based on an iterative closest point (ICP) algorithm [2] which matches two surfaces. The algorithm iteratively minimizes the mean square distance of nearest neighbor points between the two surfaces. In our case one surface is the virtual model and the other corresponds to the measured scene.

The ICP algorithm converges successfully only for a limited range of initial misalignments. However, this range appears to be comfortably large and does not demand for a precise alignment from the operator. Previous experiments showed successful matching for two object surfaces if the two surfaces are placed one on top of another with a difference in orientation of up to +/- 60° around the vertical axis and +/- 50° around a horizontal axis [6]. Therefore, a rough pose estimate of the virtual object is expected to be sufficient, but further experiments have to prove that fact.

The original ICP algorithm holds for the matching of two surfaces where one surface is a subset of the other. However, it should allow the matching of two surfaces which are each subsets of larger surfaces because in our application both surfaces, the scene and the virtual object, contain parts not present in the other one. Therefore, a modification of the ICP algorithm is necessary as proposed by other authors [7, 10] for similar problems. We currently do research on an appropriate adaptation of the ICP algorithm for our application.

6. CONCLUSIONS

We present a system for the construction of virtual worlds by mean of an AR interface. It uses a semi-automatic registration approach where the operator locates and identifies objects and the computer calculates the registration. Human perception gives great flexibility while the computer provides accuracy.

The innovative aspect of the system concerns the use of range images which give access to the depth information of the scene geometry. This 3D data is used for better world rendering and improved matching.

Several benefits are expected from this new use of range images in AR:
– It allows the user to concentrate on object identification and classification.
– Since the matching algorithm uses directly the measured 3D points of the object surfaces, there is no need of feature segmentation and objects of any form can be matched very accurately.
– Shape differences can be made visible using different surface colors [10].

ACKNOWLEDGMENTS

This research has been funded by the Swiss national science foundation under project number 2100-43530.

REFERENCES

1. R. Azuma, “A survey of augmented reality,” *Proceedings of SIGGRAPH*, ACM, Los Angeles, 1995.
2. P.J. Besl and N.D. McKay, “A Method for Registration of 3-D Shapes,” *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, vol. 14, no. 2, pp. 239-256, 1992.
3. M. Dhome, A. Yassine, and J.M. Lavest, “Pose estimation of a robot arm from one graylevel image,” *Workshop on Computer Vision for Space Applications*, Antibes, 1993.
4. M. Hebert, J. Ponce, T.E. Boult, A. Gross and D. Forsyth, “Representation for computer vision,” *Report of the NSF/ARPA workshop on 3D object*, CUNY Graduate School and University Center, 1994.
5. W. A. Hoff, K. Nguyen, T. Lyon, “Computer vision-based registration techniques for augmented reality,” *Proceedings of SPIE Intelligents Robots and Computer Vision*, Boston, vol. 2904, pp. 538-548, 1996.
6. H. Hugli, Ch. Schutz, “How well performs free-form 3d object recognition from range images?,” *Intelligent Robots and Computer Vision XV, Algorithms, Techniques, Active Vision and Materials Handling*, SPIE, Boston, vol. 2904, pp. 66-74, 1996.
7. B. Krebs, P. Sieverding and B. Korn, “Correct 3D Matching via a Fuzzy ICP Algorithm for Abitrary Shaped Objects,” *DAGM96 - Mustererkennung*, Heidelberg, 1996.
8. D. Maman, P. Fuchs, F. Nashashibi and C. Laurgeau, “Utilisation des techniques de la realite virutelle pour la modelisation interactive en teleoperation,” *Proc. 5th Int. Conf. on Interface to real and virtual worlds*, Montpellier, pp. 51-64, 1996.
9. P. Milgram, A. Rastogi, J.J. Grodski, “Telerobotic control using augmented reality,” *Robot and Human Communication (RO-MAN) 95*, Japan, 1995.
10. V. Moron, P. Boulanger, T. Masuda, T. Redarce, “Automatic inspection of industrial parts using 3-D optical range sensor,” *Proceedings of SPIE Videometrics*, Philadelphia, vol. 2598, pp. 315-325, 1995.
11. B. Peuchot, A. Tanguy and M. Eude, “Virtual reality as an operative tool during scoliosis surgery,” *Proc. 5th Int. Conf. on Interface to real and virtual worlds,* Montpellier, pp. 149-157, 1996.
12. M. Rutishauser, M. Stricker, M. Trobina, “Merging Range Images of Arbitrarily Shaped Objects,” *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Seattle, pp. 573-580, 1994.
13. Ch. Schutz and H. Hugli, “Change detection in range imaging for 3D scene segmentation,” *Proceedings of SPIE Europto*, Besancon, vol. 2786, pp. 98-106, 1996.
14. Ch. Schutz, N. Natonek, Ch. Baur, H. Hugli, “3D Vision in a Virtual Reality Robotics Environment,” Proceedings of SPIE Telemanipulator and Telepresence Technologies, Boston, vol. 2901, 1996.
15. Z. Zhang, “Iterative point matching for registration of free-form curves and surfaces,” *International Journal of Computer Vision*, vol. 13, no. 2, pp. 119-152, 1994.