Model-Based Visual Tracking for Outdoor Augmented Reality Applications

Reinhold Behringer, Jun Park*, and Venkataraman Sundareswaran
Rockwell Scientific Company (RSC)
{rbehringer, vsundar}@rwsc.com, jpark@cs.hongik.ac.kr

Abstract

Outdoor Augmented Reality (AR) applications rely on hybrid tracking (GPS, digital compass, visual) for registration. RSC has developed a real-time visual tracking system that uses visual cues of buildings in an urban environment for correcting the results of a conventional tracking system. This approach relies on knowledge of a CAD model of the building. It not only provides motion estimation, but also absolute orientation/position. It is based on the "Visual Servoing" approach, originally developed for robotics tasks. We have demonstrated this approach in real-time at a building on the NRL campus This poster shows the approach and results. The concept can be generalized to any scenario where a CAD model is available. This system is being prepared for integration into the NRL system BARS (Battlefield Augmented Reality System).

1. Introduction

Augmented Reality (AR) applications require a high registration precision [1]. In an outdoor scenario, GPS and digital compass provide position and orientation data, but their accuracy is not sufficient for a seamless and well-aligned overlay onto the user's view. Visual tracking (with a user-worn camera) can improve the tracking precision by using real-world (natural or man-made) features of the environment (e.g., [2]) and provide for example a correction of the drift of inertial trackers [3]. By using the knowledge of exact location of features in the environment, the complete 6 DOF registration can be obtained from visual tracking. RSC has implemented such a system, to be used in the NRL system BARS [4].

2. The Visual Servoing Approach

Assuming a known initial position/orientation of a camera, its motion in 6 DOF can be obtained by the "visual servoing" approach [5]. This algorithm is based on a minimization of the error in the forward projection of image features (known from a CAD model). The algorithm determines the optimal motion vector in 3D (translation and rotation) from the displacement of the 2D image features in order to provide a transition into the best match of measured 2D features and predicted features, based on the internal assumption of position/orientation. Its application for tracking in an AR system has been described in [6] and is patented (US 6,330,356 B1). It is independent on the type of employed image processing and generally can be used with any 2D feature tracking, as long as the CAD model includes these features.

3. Employed Image Processing Techniques

Due to the real-time nature of AR, the emphasis on feature tracking algorithms must be on fast and efficient execution. We have developed a set of image processing routines for simple extraction of edges (gradient scan) and corners, the latter based on the Kanade-Lucas-Tomasi (KLT) tracker [7]. These routines are also suitable to be ported onto a low-power wearable platform. In the case of a specific demonstration environment, we chose the windows of a building as well trackable features.

4. Control Structure

The initialization assumes that the visual tracking module is embedded in a hybrid tracking system, which provides an approximate initial position and orientation. If the visual tracking is operated in stand-alone mode, the user can manually perform the initial feature correspondence by "dragging" features in the video image onto the overlaid (and not registered) CAD model. Once the correspondence is set, the position and orientation of the camera (resp. the user) are calculated by the visual servoing algorithm. If the camera is moved, new features coming into the viewing area are predicted, based on the perspective projection of the internal CAD model. At least 4 arbitrary features must be tracked for the algorithm to produce robust 6 DOF tracking. For best and mathematically stable results, at least one of the features should not be coplanar to the others.

5. Implementation

The software is written in Microsoft Visual C++ for execution on a PC, running Windows 2000™. It is comprised of a set of ActiveX controls which allow a sharing of software components among applications. The following ActiveX controls are employed:

---

* now working at Hongik University, Seoul, South Korea

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 **IEEE**

Video capture can be performed using virtually any commercially available camera (USB, PCI capture board, etc.), supporting VideoForWindows drivers.

Socket communication allows a platform-independent access of the tracking functionality as a server. A set of functions has been written for integration in a client application, using the GNU CommonC++ library for cross-platform compatibility.

6. Results

For developing the algorithms, we have used video footage from our partners in this project at NRL from a campus building. We also built a 3D scale model for live and arbitrary capture. The software in this setup is running on a Pentium-4 CPU with 1.4 GHz with an Osprey video capture board, which allows video capture at 30 fps into a 640x480 pixel sized image from DV input or analog composite/S-video. The CPU load for this video capture procedure on our system is around 54%. The image processing algorithms do not add a significant load.

6.1. Real-Time Live Demonstration

During June 3-7, 2002, we have successfully demonstrated real-time registration with this approach at the NRL campus in Washington DC. The setup for this demo was a Sony Vaio Laptop with a Pentium III CPU, running 750 MHz. The video capture format was chosen to 320x240 pixel at 15 fps, due to limitations of the connected USB camera. The CPU load here was 80% for the complete system, including video capture and image processing..

6.2. Feature Detection and Tracking

A fast capture rate proved to be important for reliable tracking, without the displacement getting too large. 30 fps in a 320 pixel wide video provided the capability to track during a rotation rate of approx. 40 deg/sec in stand-alone mode. In the anticipated hybrid mode (with tracking through compass / gyro), this rate describes the additional correction that the visual tracking can provide. We experimented with tracking of up to 28 features simultaneously, without any significant increase of the computational load (up to 60% CPU utilization).

6.3. Tracking Accuracy

The theoretical precision of visual tracking can be one pixel; that is between 0.05 and 0.1 deg, depending on the optics of the camera lens. However, a realistic error assessment shows that a precision of 0.5 deg can be achieved. Error sources are incorrect camera calibration, lens distortion, and possible incorrect CAD model data. In Figure 1, the resulting overlay of a building CAD model is shown, as calculated by our approach. The left picture shows the convergence during the Visual Servoing activity, the right picture shows the matched overlay.

Figure 1.Left: Convergence during Visual Servoing. Right: Final result of converged overlay.

7. Summary and Conclusion

We have shown the feasibility of the visual tracking for AR applications. Improvements need to be made on the robust feature tracking / image processing. Suitable would be a dynamic template matching, which automatically adapts itself to changing viewpoint and lighting conditions.

8. Acknowledgements

Part of this work was funded by ONR, contract N00014-99-3-0018.

9. References

[1] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent advances in Augmented Reality." *IEEE Computer Graphics and Applications*, Vol. 21, No. 6, pp. 34-47. IEEE Computer Society, November 2001.

[2] Ferrari, V., Tuytelaars, T., and van Gool, L., "Markerless AR with a real-time affone region tracker." *ISAR 2001*, New York, October 2001, pp. 87-96.

[3] Satoh, K., Anabuki, M., Yamamoto, H., and Tamura, H., "A hybrid registration method for outdoor Augmented Reality." *ISAR 2001*, New York, October 2001, pp. 67-76.

[4] Julier, S., Baillot, Y., Lanzagorta, M., Brown, D., Rosenblum, "*BARS: Battlefield Augmented Reality System.*" NATO Symposium on Information Processing Techniques for Military Systems, 9-11 October 2000, Istanbul, Turkey.

[5] B. Espiau, F. Chaumette, P. Rives, "A New Approach to Visual Servoing in Robotics." *IEEE Transactions on Robotics and Automation*, Vol. 8, No. 3, pp. 313-326, 1992.

[6] V. Sundareswaran, and R. Behringer, "Visual Servoing-based Augmented Reality." *First IEEE Workshop on Augmented Reality (IWAR'98)*, San Francisco, November 1998.

[7] Carlo Tomasi and Takeo Kanade, "Detection and Tracking of Point Features." Carnegie Mellon University Technical Report CMU-CS-91-132, April 1991.

Stereo Augmentation of Simulation Results on a Projection Wall by Combining Two Basic ARVIKA Systems....271
S. Noelle

Figure 1: Augmentation of a crashed aluminum tube with the corresponding simulation results by using a transparent mesh visualization

Figure2: Augmentation of a crashed aluminum tube with the corresponding simulation results using a transparent visualization

Figure 3: Augmented crashed tube

Figure 4: Augmented entire tube

Model-Based Visual Tracking for Outdoor Augmented Reality Applications....277
R. Behringer, J. Park, and V. Sundareswaran

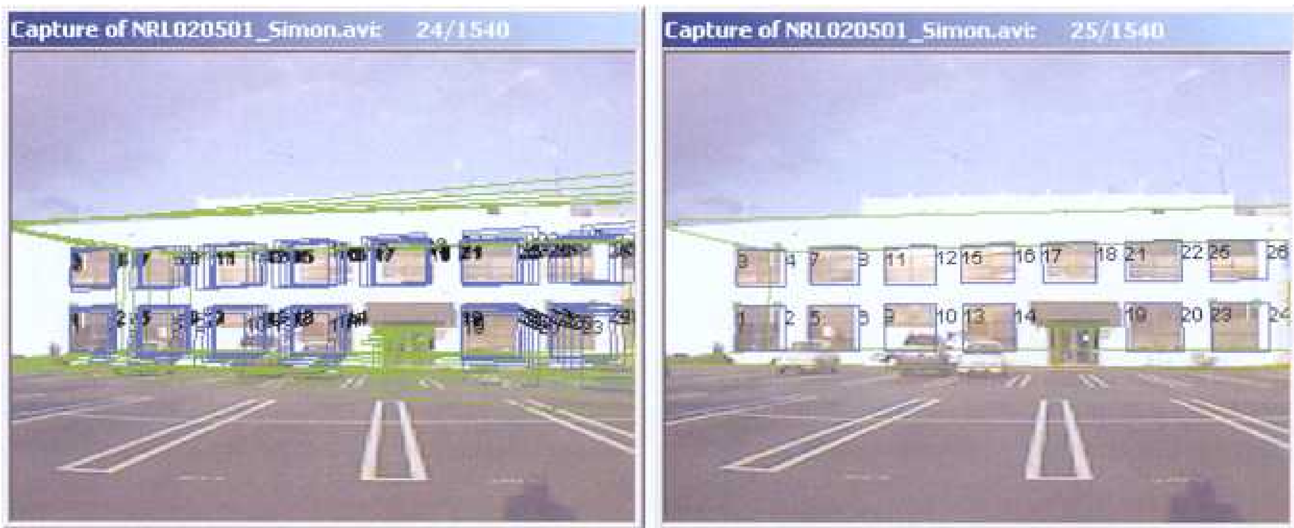

Figure 1. .Left: Convergence during Visual Servoing. Right: Final result of converged overlay.

Proceedings of the International Symposium on Mixed and Augmented Reality (ISMAR'02)
0-7695-1781-1/02 $17.00 © 2002 **IEEE**