A General Approach for Closed-Loop Registration in AR

Feng Zheng and Ryan Schubert*
The University of North Carolina at Chapel Hill

Greg Welch†
The University of Central Florida and UNC-CH

ABSTRACT

Tracking and augmentation are usually handled in independent consecutive stages in augmented reality (AR). The result is that the real-virtual registration is "open loop"—inaccurate tracking leads to misregistration that is seen by the users but not the system. We propose a general approach to "close the loop" in the displayed appearance by using the visual feedback of registration for tracking. Specifically, a model-based method is introduced to simultaneously track and augment real objects in a closed-loop fashion, where the model is comprised of the combination of the real object to be tracked and the virtual object to be rendered. This method is applicable to paradigms including video-based AR, projector-based AR, and diminished reality.

Keywords: Closed-loop registration, visual feedback, tracking

Index Terms: H.5.1 [INFORMATION INTERFACES AND PRESENTATION (e.g., HCI)]: Multimedia Information Systems—Artificial, augmented, virtual realities; I.4.8 [IMAGE PROCESSING AND COMPUTER VISION]: Scene Analysis—Tracking

1 INTRODUCTION

Registration is a fundamental task in AR. Open-loop AR systems allow for registration errors as they have no mechanism for observing the resulting appearance. In practice, there are a number of static and dynamic error sources [3], resulting in misregistration that goes "unseen" by the system. Other researchers have explored specific cases of using the images of real-virtual registration as feedback into the tracking step to achieve a closed-loop registration system [1, 2]. We propose a related closed-loop approach, with the novelty of being suitable for multiple AR paradigms, that uses the desired augmented imagery (the combination of real and virtual) directly as the goal for a model-based method.

This combined model-based augmentation and tracking offers several advantages. It embodies a closed-loop system that is continuously adjusting parameters to the desired augmented appearance. It does so without the explicit detection and use of features or points in the camera imagery, instead optimizing the parameters directly using any misregistration manifested in the augmented imagery. Our approach can be used by itself or in combination with a conventional open-loop approach by using the open-loop tracking for a coarse pose estimate prior to closed-loop optimization.

2 REAL-VIRTUAL MODEL-BASED REGISTRATION

The user is expected to observe the correct view of the combined appearance of the real and virtual, i.e., the observed image should match an expected image. This suggests a natural formulation of the cost function:

$$\arg\min_{\mathbf{p}} \|\hat{C}(\mathbf{u}) - \hat{G}(W(\mathbf{u};\mathbf{p}))\| \tag{1}$$

where $\hat{C}$ is the image we observe, i.e., the combined appearance of the real and virtual, called the *combined image*, while $\hat{G}$ represents the expected appearance, called the *goal image*. The goal image $\hat{G}$ is the 2D appearance of the goal model, which is the registered combination of the real object to be tracked and the virtual object to be rendered. To minimize the 2D image difference, the goal image is acquired by transforming the goal model using the warping function $W(\mathbf{u};\mathbf{p})$, where $\mathbf{u} = (u, v)^T$ is a 2D column vector containing the pixel coordinates, and $\mathbf{p} = (p_1, \cdots, p_n)^T$ is a vector of parameters for arbitrary spatial transformation, e.g., homography, 6D pose. If the goal model is planar, i.e., both the real object and virtual object are planar, either a homograhy or a 6D pose can be used. On the other hand, we generally use 6D pose parameterization when the goal model is not planar.

For projector-based AR, assuming planar and lambertian surface and no ambient light, the observed combined image $\hat{C}$ can be approximated as a multiplicative modulation of the projector image $\hat{V}$, called the *virtual image*, the surface reflectance $\hat{R}$, called the *real image*, and the cosine angle between the surface normal and projector light:

$$\hat{C}(\mathbf{u}) = \hat{V}(W(\mathbf{u};\mathbf{p})) \cdot \hat{R}(\mathbf{u}) \cdot \cos\theta \tag{2}$$

where the virtual image $\hat{V}$ is warped onto the coordinate frame of the real image $\hat{R}$. The coordinate frames of $\hat{C}$ and $\hat{R}$ are the same. The projector-camera system is assumed to be geometrically calibrated.

Plugging Equation (2) into Equation (1) and using the L2 norm as the error metric, we have

$$\sum_{\mathbf{u}} \|\hat{V}(W(\mathbf{u};\mathbf{p})) \cdot \hat{R}(\mathbf{u}) \cdot \cos\theta - \hat{G}(W(\mathbf{u};\mathbf{p}))\|^2 \tag{3}$$

Thus a nonlinear optimization problem is formulated. Using the log operation to simplify Equation (3) and the Gauss-Newton method to solve for the parameter update, an analytic solution can be obtained.

To extend the proposed method to other AR paradigms, the combined real and virtual image needs to be formulated in a simulated way, not like in projector-based AR which is computed optically. To do this, we consider the relationship as simple addition, i.e., $\hat{C} = \hat{V} + \hat{R}$. Then to compute the virtual image, we can simply subtract the template object image $\hat{T}$ from the goal image $\hat{G}$, i.e., $\hat{V} = \hat{G} - \hat{T}$. An illustration of the various images and computations is shown in Figure 1. With this simplified relationship of the real and virtual, the algorithm can be used without change for different AR paradigms.

3 EVALUATION

We performed three experiments for each mentioned AR paradigm, focusing on planar objects. The method can be readily extended to track and augment non-planar objects if their 3D structure is known. For pose parameterization, we tried both a 2D homography and a 6D pose with twist representation.

3.1 Experiment 1: Projector-Based AR

Similar to [1], in this experiment parts of the expected imagery are printed on the board while the others are projected. We chose to optimize for a 2D homography then extract the 6D pose from it based

---

*e-mail: {zhengf, res}@cs.unc.edu
†e-mail: welch@ucf.edu

(a) goal image $\hat{G}$ (b) template image $\hat{T}$ (c) virtual object

(d) combined image $\hat{C}$ (e) input frame $\hat{R}$ (f) virtual image $\hat{V}$

Figure 1: Illustration of combined image formulation. The goal image $\hat{G}$ in (a) is the 2D appearance of the goal model, which is comprised of the real object or template image $\hat{T}$ in (b) and the virtual object in (c). The virtual image $\hat{V}$ in (f) is computed as the subtraction of $\hat{G}$ and $\hat{T}$. The combined image $\hat{C}$ in (d) is the addition of the input frame $\hat{R}$ in (e) and $\hat{V}$. In the registration process, the goal image $\hat{G}$ is iteratively acquired by transforming the goal model using the current pose estimate, until it matches the combined image $\hat{C}$.

on prior calibration information. We achieved 10 frames/sec with the current implementation. The algorithm successfully converged for the test sequence. Results are shown in Figure 2.

(a) frame 169 (b) frame 180

Figure 2: (a) shows the misregistered view and (b) shows the corresponding registered view after a number of iterations.

Due to the difference in our cost function formulation compared to [1], we project an image for *each* iteration using incrementally estimated pose parameters. This means that the real-virtual optimization (augmentation with lighting) is affected and directly measured optically in the scene space every iteration, as opposed to being simulated. Another difference is that in our optimization we obtained an analytical solution while [1] evaluated the Jacobians numerically. Moreover, even without radiometric calibration, our method worked well and was robust in handling the test sequences.

3.2 Experiment 2: Video-Based AR

(a) 1st iteration (b) 6th iteration (c) 9th iteration

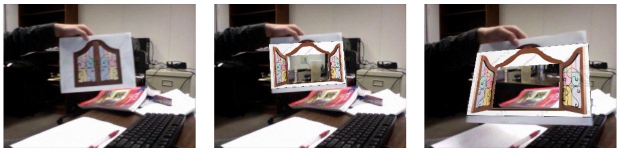
Figure 3: (a) shows the initial misregistered appearance (note the bunny region). (b) shows decreased registration error. (c) shows the converged state with almost no error.

For video-based AR, we optimized for the 6D pose directly since the goal model is not planar, parameterized using the twist representation as in [4]. We tested our approach with two synthetic sequences both of which were accurately tracked and augmented. Figure 3 shows the progression from an intial state with some noise and large registration error to reduced error and finally almost no registration error after nine iterations.

3.3 Experiment 3: Diminished Reality

Diminished reality removes an object or collection of objects and replaces it with an appropriate background image [5]. It can be considered as a real-virtual registration process in which objects are tracked and augmented with virtual content that hides them.

We did a simple proof-of-concept experiment in which we computed the 2D homography between consecutive frames. For a single static camera view with a known static background, we tracked and "camouflaged" a portion of the planar real object in real time, as shown in Figure 4.



(a) real (b) frame 1 (c) frame 106

Figure 4: (a) shows the real object, which is tracked and replaced with background image and also augmented with an "opened" window. Results of two frames are shown in (b) and (c).

4 CONCLUSIONS AND FUTURE WORK

We have presented a general approach for closed-loop real-virtual registration that is suitable for multiple AR paradigms. Although our implementation uses a gradient descent method to iteratively solve for the current pose estimate, other model estimation techniques (e.g., particle filter) could also be used. Future work includes numerical analysis of registration error, comparison with existing methods, and extension to non-planar objects.

ACKNOWLEDGEMENTS

This work was supported by U.S. Office of Naval Research grants N00014-09-1-0813 and N00014-12-1-0052, both titled "3D Display and Capture of Humans for Live-Virtual Training," (Clarke Lethin), and in part by the Singapore Interactive Digital Media program office sponsored BeingThere Centre (Michael Yap).

REFERENCES

[1] S. Audet, M. Okutomi, and M. Tanaka. Direct image alignment of projector-camera systems with planar surfaces. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 303 – 310, june 2010.

[2] M. Bajura and U. Neumann. Dynamic registration correction in video-based augmented reality systems. *IEEE Comput. Graph. Appl.*, 15(5):52–60, Sept. 1995.

[3] R. L. Holloway. Registration error analysis for augmented reality. *Presence: Teleoperators and Virtual Environments*, 6(4):413–432, 1997.

[4] J. Xiao, T. Kanade, and J. F. Cohn. Robust full-motion recovery of head by dynamic templates and re-registration techniques. In *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, FGR '02, pages 163–, Washington, DC, USA, 2002. IEEE Computer Society.

[5] S. Zokai, J. Esteve, Y. Genc, and N. Navab. Multiview paraperspective projection model for diminished reality. In *Proceedings of the 2nd IEEE/ACM International Symposium on Mixed and Augmented Reality*, ISMAR '03, pages 217–, Washington, DC, USA, 2003. IEEE Computer Society.