

AR for the Masses: Building a Low-Cost Portable AR System from Off-the-Shelf Components

Hsiang-Ting Chen*

Chun-Fa Chang†

Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan, R.O.C.

Abstract

To create the illusion that a virtual object coexists with physical objects and its environment is always an important goal in the research of augmented reality. Though there are already many commercial products on the market, they are too expensive, too cumbersome or too hard to set up for an ordinary user. Our “AR for the masses” system is cheap to build, easy to set up, and it does not require the users to wear a head-mounted display (HMD). Its cost is low because the whole system consists of only two web cameras (for 3D tracking), a paper box (as the proxy object) and a projector which is getting cheaper. It is also easy to set up: we can move the whole system to a new location easily and finish the calibration within just a few minutes.

CR Categories: I.3.7 [Computer Graphics]: Three-Dimensional Graphics and Realism—Virtual Reality

Keywords: augmented reality, low-cost, projector, camera, calibration.

1 Introduction

With the rapid evolvement of computer graphics, it is not a dream anymore to create a huge virtual environment for people to walk through. Wearing a virtual-reality helmet or glasses with separate left and right views gives the user even more compelling stereo feeling. However, those devices are often cumbersome and could cause psychological discomfort to the users due to the fact that human eyes tend to focus on the nearest flat surface and at the same time they tend to meet at the object of the observation. That is the origin of typical headaches caused by their prolonged use. Moreover, systems of virtual reality isolate the users from the rest of the world and make it hard to have the communication or interaction with the other users or the environment.

To remove the HMD and let user have more intuitive access to the virtual environment, the question becomes how to enhance already built environment rather than replace them by virtual environments. This is the main research topic of Augmented

Reality (AR). However, most AR systems use see-through head-mounted displays which still share the most disadvantages of the traditional helmets or glasses of virtual reality.

In this paper, we introduce a projection-based augmented reality system which works just like a showcase in a museum with scientific or cultural artifacts inside except that the objects in our case are virtual. Also we intend to make the system suitable for ordinary users rather than well-funded institutes. It means that the system should be low-cost and easy to set up.

There are three main components in our system:

- 1) The cameras for head tracking.
- 2) The showcase, which also serves as a proxy object.
- 3) The projector to create the illusion.

The system tracks the position of the user’s head and generates a pre-warped image that is then projected onto the proxy object, i.e., the showcase, to create the illusion that there is a virtual object inside the showcase.

To lower the cost of this system, we use two ordinary web cameras, which are the minimal setup for optic 3D head tracking, as our tracking system. To reduce the user’s discomfort, we replace the traditional HMD by a projector. And we provide a simple scheme for the users to set up the system quickly. Note that an HMD-based system requires the tracking of the head position and orientation. However, a projector-based system requires only the tracking of the head position. That is the main reason why we may build a low-cost head tracking system.

2 Previous Work

Projection-based AR systems such as [Raskar et al. 1998a] provide an alternative to HMD. By projecting images directly onto the surface of a physical object, the users can enjoy the illusion created by the computers while still view the objects in a natural way. You can walk in or out and move your head freely. However, this type of systems can only augment physical objects rather than produce a virtual object. For example, in the Shader Lamp project [Raskar et al. 2001], they built a physical Taj Mahal model that was not colored. In [Oliver et al. 2001], they proposed a virtual showcase system, where they used half-silvered mirrors and a projector to build a showcase and create the illusion that a virtual object exists in the showcase. However, we can still see some drawbacks such as:

- 1) This system is hard to set up and not portable.
- 2) The showcase must be placed upon a big piece of glass, and the environment near it must be cleared to avoid occlusion with the projected images. In other words, there exists some isolation of the virtual object and the surrounding physical environment.

*e-mail: timchen@ibr.cs.nthu.edu.tw

†e-mail: chunfa@cs.nthu.edu.tw

Most AR systems also pay little attention to their costs and the ease of setup and calibration, while our goal is to build a system that is relatively cheap and can be quickly set up and calibrated when it is moved to a different location.

In [Raskar et al. 2003], they have the same general idea of bringing this kind of AR system to everyday life as we do and present a portable projection-based system. Their goals are to generate projected texture which is customized to the shape of the surface so that it is consistent with our usual viewing experience, and to create a large display by an ad-hoc cluster of projectors. Their system does not actually track the user's viewpoint. Their system also needs calibrated projector-camera pairs with tilt sensors.

3 Methods

As mentioned in the introduction, there are three main components in our system. They are two web cameras for head tracking, a proxy object, and a projector for rendering virtual object. We need to calibrate the web cameras to obtain the 3D position of the user's head and to calibrate the projector so that it can produce perceptively correct image when it is viewed from the user's viewpoint. The main problem here is how to find suitable correspondence between 2D points and 3D points. We solve this problem by providing the user a mouse-clicking user interface for clicking markers on the proxy object. It is quite reasonable to treat the proxy object and the markers on it as the reference coordinate system since the projector must illuminate the proxy object which means that there are correspondences between 2D projector pixels and 3D marker positions. The cameras should be able to see the proxy object as well if the user's head is not far away from the proxy object, which means that there are correspondences between 2D camera pixels and 3D marker positions too. The biggest advantage of using this scheme is to ensure that the whole system uses the same reference coordinate system, which makes the rendering process much easier. However, to give the user more freedom of head movement, we usually do not force the camera to see the proxy object. In such a case, we use extra markers (to be discussed in 3.1) to calibrate the camera.

3.1 The Setup Process

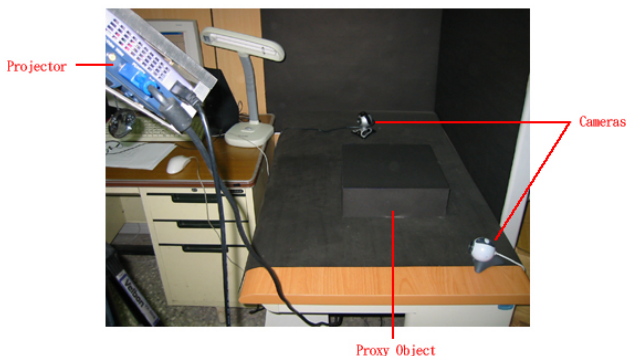


Figure 1: Our experimental system consists of two web cameras, a paper box on the table as the proxy object and a projector at the left.

First, we casually put the web cameras, the proxy object, and the projector in the room. Note that there is no strict constraint as

long as the cameras can see the user's head, and the projector can illuminate the proxy object (Figure 1).

Second, we use the user interface we designed to click the markers on the proxy object to calibrate the projector (Figure 2 Left). Note that if a camera cannot see the markers on the proxy objects then we will need extra markers to help accomplish the calibration of the cameras. That may happen when the user intends to stand far away from the proxy object or simply because the field of view of the web camera is too narrow to see both the user's head and the markers. In our case, we use an extra checkerboard as our marker to calibrate the cameras because the restricted field of view of the web cameras (Figure 2 Right).

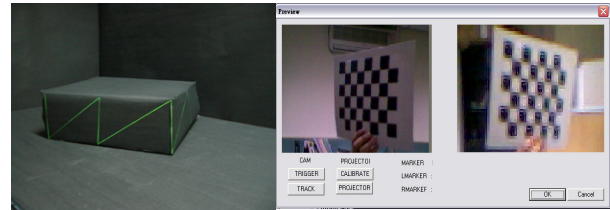


Figure 2: (Left) The calibration of the projector is achieved by moving the mouse cursor that is projected on the proxy object to click on the markers. (Right) Shown are the user interface and the checkerboard used for the camera calibration.

3.2 Camera Calibration and Head Tracking

A caveat of camera calibration is that the field of view of most available web cameras is about 20 degree, which is only half of typical single-lens-reflex (SLR) camera lenses. In this case, we can ignore the distortion and nonlinear characteristics of the cameras and simply use the DLT (Directly Linear Transformation) algorithm [Hartley and Zisserman 2001] which produces satisfying calibration results for our system.

DLT is summarized as follow: Given point correspondences between 3D points X_i and 2D image points x_i we are required to find a camera matrix P , a 3-by-4 matrix, such that $x_i = PX_i$.

Since this is an equation involving homogeneous vectors, x_i and PX_i may have the same direction but differ in magnitude by a non-zero scale factor. Thus $x_i \times PX_i = 0$. For every 2D-3D correspondence we have the relationship of

$$\begin{pmatrix} 0^T & -\omega_i X_i^T & y_i X_i^T \\ \omega_i X_i^T & 0^T & -x_i X_i^T \\ -y_i X_i^T & x_i X_i^T & 0^T \end{pmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \end{pmatrix} = 0$$

Where $(P^i)^T$ is the i -th row of P . Since the three equations above are linear dependent, we can use only two of them. From a set of n points, we have a $2n \times 12$ matrix A by stacking up the equations. And the projection matrix P is computed from solving the equation $Ap = 0$, where p is the vector containing the entries of the matrix P . For over-determined solutions, having more than 6 points, we should add constraints like $\|P\| = 1$.

For head tracking, an additional problem is how to find the user's head on the frames grabbed from both cameras. There are many

existing methods utilizing the features on human face like eyes, nose or mouth. However, since this part is beyond the scope of this paper, we just assume there is a marker on the user's head for simplicity. In our implementation, we use a ping-pong ball painted red as the marker and find its position on the image based on the ratios of RGB component values.

After knowing the 2D head (marker) position in both camera views, we then use the traditional triangulation method to calculate the 3D head (marker) position.

3.3 Projector Calibration

The internal and external parameters of a projector can be expressed in the same way as those of a camera. Given more than 5 correspondences between 2D projector pixels and corresponding 3D points, we can calibrate the projector using DLT again. As mentioned in [Raskar and Beardsley 2001], there is no difference between the projector model and the camera model except that a traditional assumption about simplified cameras models, that the principal point is close to the image, is not valid for projectors. Most projectors use an off-axis projection since they are usually set on a table or hang on a ceiling and the image is projected from upper or lower half of the lens. Hence, the principle point is vertically shifted.

Though there is an automatic projector calibration algorithm proposed in [Raskar and Beardsley 2001], it is not suitable in our system because we project the image on a proxy object rather than a plane. Thus we provide a user interface for the users to calibrate the projector by manually moving the mouse cursor and clicking on markers on the proxy object. It is not as tiresome as it sounds since providing merely 6 correspondences between 2D and 3D points, that is clicking the mouse 6 times, gives us enough information for the projector calibration. However, we recommend the users to give more than 6 correspondences, because in this way, the users do not have to click the marker very precisely and can finish the calibration process even faster. In our case, we ask the users to give 12 correspondences.

3.4 Rendering

To create the image of virtual objects that is perceptively correct when viewed from the user's viewpoint after it is projected on the proxy object, we utilize the two-pass rendering algorithm proposed in [Raskar et al. 1998b].

In the first pass, we render the virtual object from the user's viewpoint and save the output into a pixel buffer. In the second pass, we use the projective texture technique [Segal et al. 1992] to project the image in the pixel buffer to the 3D model of the proxy object, by setting the eye to the (real) projector position and the (projective texture) projector to the user's viewpoint for the final rendering. The final rendering may look distorted on its own, but will produce perceptively correct result when it is projected to the proxy object and observed from the user's viewpoint.

The cost of the first pass is the effort of rendering the virtual object to be displayed in the showcase. The cost of the second pass is the effort of rendering the proxy object with projective texture on it, which depends mainly on the complexity of the 3D

model of the proxy object. In our case, the cost of the second pass can be neglected since the proxy object is only a cube.

4 Implementation and Results

We implement our system using a BenQ PB2220 projector, two cheap Logitech Quickcam web cameras, and an ordinary paper box as our proxy object. We choose black color for the proxy object because it reduces the influence of the environment lighting. We rendered the image on a Pentium4 2.0G PC with a GeForce4 Ti4200 graphics card. By using the DirectShow API, we are able to grab two frames simultaneously from two cameras. Since the cost of rendering part is so low, the bottleneck of this program becomes the time we use DirectShow to acquire frames from cameras.

Setting up the whole system takes less than 5 minutes, including the time to place the projector to illuminate the proxy object and the time to properly place the cameras to cover the potential space of user's head position.

We show the results in the following. In Figure 3, we can that the image projected from the projector is extremely distorted (Figure 3 Top); however, it appears perceptively correct when it is observed from the user's viewpoint (Figure 3 Bottom). After moving the head a little bit to the upper right, we get the pictures in Figure 4.

Note that by rendering the wire frame of the proxy object, we can exam the precision of the system. As seen in Figures 3 and 4, the wire frame matches the proxy object well. It shows that even with our simple and quick setup scheme, we can still produce illusion of virtual objects with satisfying precision.

5 Conclusion and Future Work

We have presented a low-cost and easy to set up AR system using only off-the-shelf components. We believe that the AR technology is going to be one part of everyday life in the future, and this system is an attempt to achieve such a goal.

Our system can be improved in many aspects. First, it is desirable to broaden the field of view of cameras and allow the users to move more freely. It can be done easily by replacing the web cameras with better CCD video cameras which are no longer expensive nowadays. Second, it is possible to add stereo display to our system. Since the image in our system is projected onto the surface of the proxy object, the users tend to focus on the surface rather than the inside of the box. However, the users may have to wear stereo glasses in such a case. Third, it is possible to model and render the environment surrounding the proxy object. With the environment rendered, we can even make the proxy object less intrusive and give user a more realistic illusion of the co-existence of the virtual objects and their nearby environment.

6 Acknowledgement

We would like to thank the anonymous reviewers for their valuable suggestions and comments. This work is supported by R.O.C. DOE Grant 89-E-FA04-1-4 (Program for Promoting Academic Excellence of Universities).

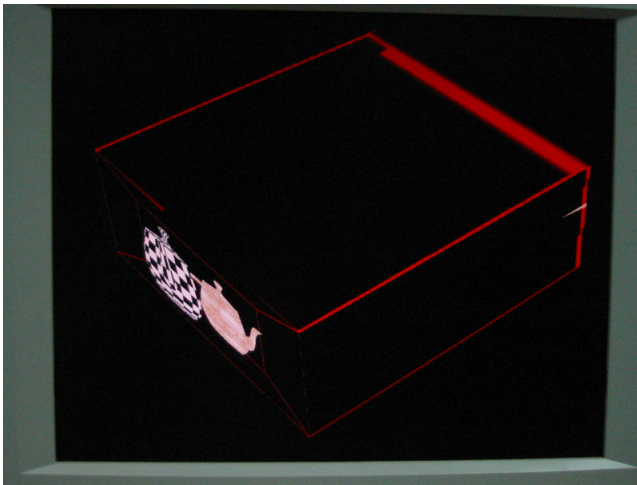
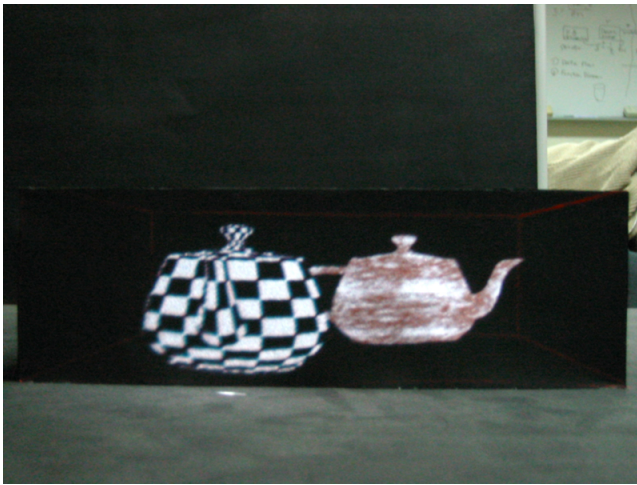


Figure 3: (Top) The image observed from the user's viewpoint. (Bottom) The pre-warped image projected by the projector.

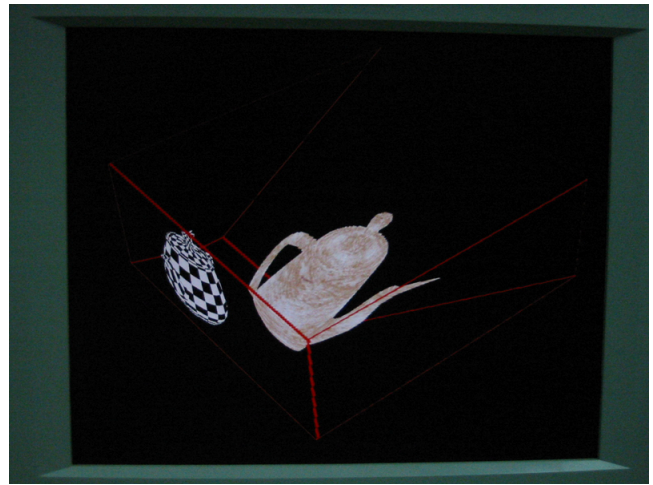


Figure 4: (Top) The observed and (Bottom) the pre-warped image when the user's viewpoint moves to the upper right from the one in Figure 3.

References

- BIMBER, O. FROHLICH, B. SCHMALSTIEG, D. AND ENCARNACAO, L. M. 2001. The Virtual Showcase. In *Proceedings of IEEE Computer Graphics and Applications*, November/December, 2001.
- HARTLEY, R., AND ZISSERMAN, A. 2001. *Multiple View Geometry in Computer Vision*. Cambridge University Press.
- RASKAR, R. WELCH, G. AND FUCHS, H. 1998a. Spatially Augmented Reality. In *Proceedings of First International Workshop on Augmented Reality*, San Francisco, November 1, 1998.
- RASKAR, R. CUTTS, MATT. WELCH, GREG. AND STÜERZLINGER, W. 1998b Efficient Image Generation for Multiprojector and Multisurface Displays. In *Proceedings of 9th Eurographics Rendering Workshop*, 1998.
- RASKAR, R. WELCH, G. LOW, K. L. AND BANDYOPADHVAY, D. 2001. Shader Lamps: Animating Real Objects With Image-Based Illumination. In *Proceedings of Eurographics Workshop on Rendering*, London, England, June 25-27, 2001.
- RASKAR, R. AND BEARDSLEY, P. 2001. A Self-Correcting Projector. In *Proceedings of IEEE Computer Vision and Pattern Recognition (CVPR) 2001*, Hawaii, Dec 2001.
- RASKAR, R. BAAR, J. V. BEARDSLEY, P. WILLWACHER, T. RAO, S. AND FORLINES, C. 2003. iLamps: Geometrically Aware and Self-Configuring Projectors. *ACM Transactions on Graphics*, 22, 3, 809-818.
- SEGAL, M. KOROBKIN, C. WIDENFELT, R. V. FORAN, J. AND HAEBERLI, P. 1992. Fast shadows and lighting effects using texture mapping. In *Computer Graphics (Proceedings of ACM SIGGRAPH 92)*, pages 249-252.