Real-Time 3D Human Capture System for Mixed-Reality Art and Entertainment

Ta Huynh Duy Nguyen, Tran Cong Thien Qui, Ke Xu, Adrian David Cheok, *Member*, *IEEE*,
Sze Lee Teo, ZhiYing Zhou, Asitha Mallawaarachchi, Shang Ping Lee, Wei Liu, Hui Siang Teo,
Le Nam Thang, Yu Li, and Hirokazu Kato

Abstract—A real-time system for capturing humans in 3D and placing them into a mixed reality environment is presented in this paper. The subject is captured by nine cameras surrounding her. Looking through a head-mounted-display with a camera in front pointing at a marker, the user can see the 3D image of this subject overlaid onto a mixed reality scene. The 3D images of the subject viewed from this viewpoint are constructed using a robust and fast shape-from-silhouette algorithm. The paper also presents several techniques to produce good quality and speed up the whole system. The frame rate of our system is around 25 fps using only standard Intel processor-based personal computers. Besides a remote live 3D conferencing and collaborating system, we also describe an application of the system in art and entertainment, named Magic Land, which is a mixed reality environment where captured avatars of human and 3D computer generated virtual animations can form an interactive story and play with each other. This system demonstrates many technologies in human computer interaction: mixed reality, tangible interaction, and 3D communication. The result of the user study not only emphasizes the benefits, but also addresses some issues of these technologies.

Index Terms—3D viewpoint, mixed reality, tangible interaction, art, entertainment.

◆

1 INTRODUCTION

1.1 Background and Motivation

In the past few years, researchers have heralded mixed reality as an exciting and useful technology for the future of computer human interaction and it has generated interest in a number of areas, including computer entertainment, art, architecture, and communication. Mixed reality refers to the real-time insertion of computer-generated graphical content into a real scene (see [8], [9] for reviews). More recently, mixed reality systems have been defined rather broadly, with many applications demanding tele-collaboration, spatial immersion, and multisensory experiences.

Inserting real collaborators into a computer generated scene involves specialized recording and novel view generation techniques. There have been a number of systems focusing on the individual aspects of these two broad categories, but there is a gap in realizing a robust real-time capturing and rendering system which, at the same time, provides a platform for mixed reality-based tele-collaboration and provides multisensory, multi-user interaction with the digital world. The motivation for our work stems from here. 3D Live technology is developed to capture and generate realistic novel 3D views of humans at interactive frame rates in real time to facilitate multi-user, spatially immersed collaboration in a mixed reality environment.

In addition, we also refine and integrate a variety of 3D Live's fast processing and rendering algorithms into Magic Land, a tangible interaction system with fast recording and rendering of 3D humans avatars in a mixed reality scene, which brings to users a new kind of human interaction and self reflection experience. Although the Magic Land system itself only supports the recording and playback feature (because of the ability to self-reflect and interact with one's own 3D avatar), the system can be quite simply extended for live capture and live viewing.

Up to now, the idea of capturing human beings for virtual reality has been studied and discussed in quite a few research articles. In [20], Markus et al. presented "blue-c," a system combining simultaneous acquisition of video streams with 3D projection technology in a CAVE-like environment, creating the impression of total immersion. Multiple live video streams acquired from many cameras are used to compute a 3D video representation of a user in real time. The resulting video inlays are integrated into a virtual environment. In spite of the impression of the total immersion provided, blue-c does not allow tangible ways to manipulate 3D videos captured. There are few interactions described between these 3D human avatars and other virtual objects. Moreover, blue-c is currently single user per portal [20] and, thus, does not allow social interactions in the same physical space. Our Magic Land, in contrast, supports multi-user experiences. Using a cup, one player can tangibly manipulate her own avatar to interact with other virtual objects or even with the avatars of other players. Furthermore, in this mixed reality system, these interactions occur as if they are in the real-world physical environment.

- *T.H.D. Nguyen, T.C.T. Qui, K. Xu, A.D. Cheok, S.L. Teo, Z. Zhou, A. Mallawaarachchi, S.P. Lee, W. Liu, H.S. Teo, L.N. Thang, and Y. Li are with the Interactive and Entertainment Research Center, Level 02-02, Research Techno Plaza, Nanyang Technological University, 50 Nanyang Drive, Singapore 637553.*
  E-mail: {duynguyen, quitran, xuke, adriancheok, szelee, steven, asitham, shangping, liuwei, jason, thangle, jessie}@mixedrealitylab.org.
- *H. Kato is with the Department of Information Machines and Interfaces, Faculty of Information Science, Hiroshima City University, 3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima 731-3194, Japan.*
  E-mail: kato@sys.im.hiroshima-cu.ac.jp.

Manuscript received 7 Dec. 2004; revised 15 Mar. 2005; accepted 25 Mar. 2005; published online 8 Sept. 2005.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org, and reference IEEECS Log Number TVCGSI-0204-1204.

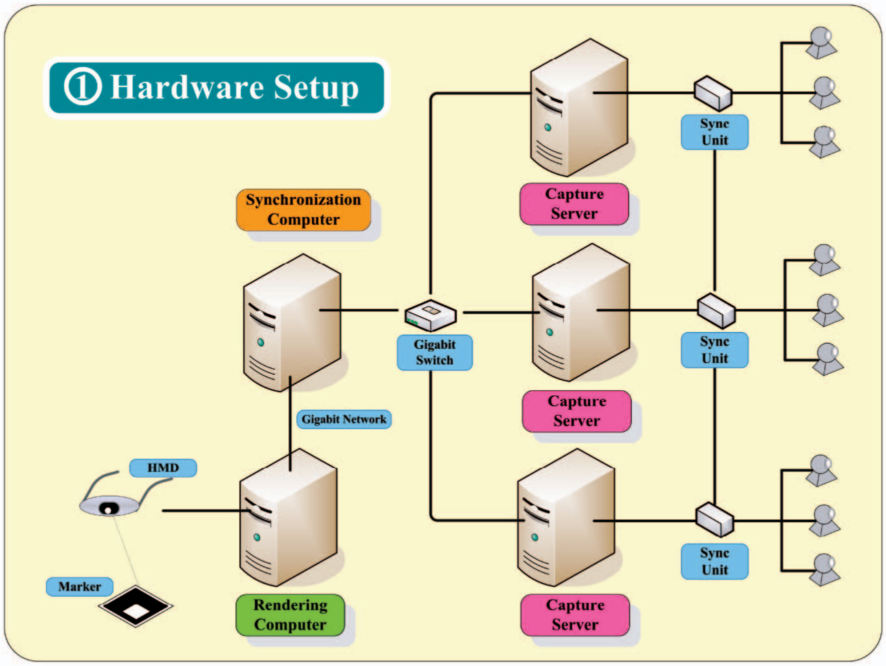


Fig. 1. Hardware architecture.

Another capture system was also presented in [13]. In this paper, the authors demonstrate a complete system architecture allowing the real-time acquisition and full-body reconstruction of one or several actors, which can then be integrated in a virtual environment. Images captured from four cameras are processed to obtain a volumetric model of the moving actors, which can be used to interact with other objects in the virtual world. However, the resulting 3D models are generated without texture, leading to some limitations in applying their system. Moreover, their interaction model is quite simple, based only on active regions of the human avatars. We feel it is not as tangible and exciting as in Magic Land, where players can use their own hands to manipulate the 3D full color avatars.

1.2 Contributions

The major technical achievements and contributions to the research field in realizing the Mixed Reality (MR) Magic Land project can be summarized as follows:

- We propose a complete and robust real-time and live human 3D recording system, from capturing images, processing background subtraction, to rendering for novel view points. Originating from our older and previous system [25], we develop this novel system by integrating new techniques to improve speed and quality. For background subtraction, we improve the quality by filtering misclassified pixels and increase the network speed by optimizing the data size. For 3D rendering, we contribute new methods to compute visibility and blend color. These contributions have significantly improved the quality and performance of our system and are very useful for mixed reality researchers.
- Our real application, MR Magic Land, is the cross-section where art and technology meet. It not only combines the latest advances in human-computer interaction and human-human communication—mixed reality, tangible interaction, and 3D Live technology—but also introduces to artists of any discipline intuitive approaches for dealing with mixed reality content. It brings together the processes of art creation, acting, and reception in one environment and creates new forms of human interaction and self reflection. Moreover, future development of this system will open a new trend of mixed reality games, where players actively play a role in the game story.
- We also introduce our laboratory developed MXRToolkit software package used in developing this research. This MXR Software Development Kit consists of a library of routines to help with all aspects of building mixed reality applications, together with a number of applications to aid with common tasks such as camera calibration, camera tracking, etc. This software is open-source and released under the GNU General Public License. More detailed technical documentations can be found at http://sourceforge.net/projects/mxrtoolkit/.

In the next section, we will proceed by first providing a description of our hardware and system setup. Then, we will discuss the software architecture. After that, our application, MR Magic Land, is presented. Finally, we will analyze a user study conducted and present the results of our findings.

2 HARDWARE AND SYSTEM DESCRIPTION

2.1 Hardware

Fig. 1 represents the overall system structure. Eight Dragonfly FireWire cameras from Point Grey Research [5], operating at 30fps, 640 x 480 resolution, are equally spaced around the subject and one camera views him/her from

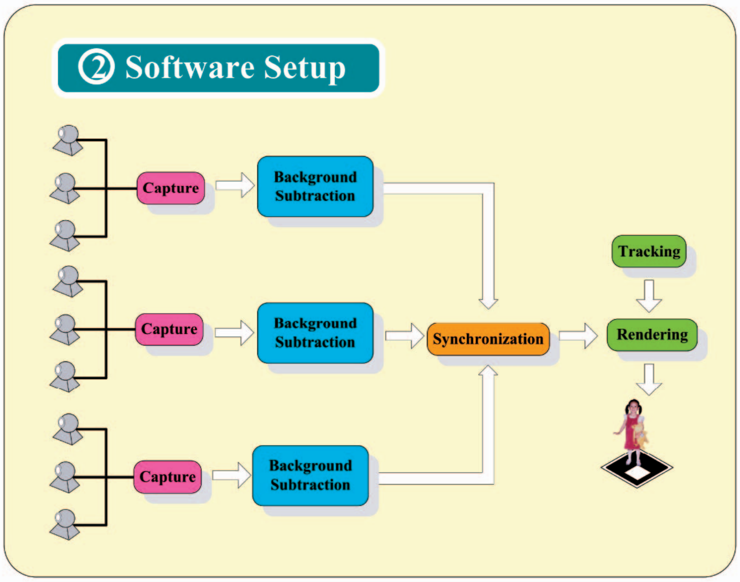


Fig. 2. Software architecture.

above. Three Sync Units from Point Grey Research are used to synchronize the image acquisition of these cameras across multiple FireWire buses [5]. Three Capture Server machines, each one a Dell Precision Workstation 650 with Dual 2.8GHz Xeon CPUs and 2GB of memory, receive the three 640 x 480 video-streams in Bayer format at 30Hz from three cameras each and preprocess the video streams.

The Synchronization machine is connected with three Capture Sever machines through a Gigabit network. This machine receives nine processed images from three Capture Server machines, synchronizes them, and also sends them via gigabit Ethernet links to the Rendering machine, which is another Dell Precision Workstation 650.

The user views the scene through a video-see-through head mounted display (HMD) connected directly to the Rendering machine. A Unibrain firewire camera, capturing 30 images per second at a resolution of 640 x 480, is attached to the front of this HMD. The Rendering machine obtains images from this Unibrain camera, tracks the marker pattern on these images, calculates the position of the virtual viewpoint, generates a novel view of the captured subject from this viewpoint, and then superimposes this generated view onto the images obtained from the Unibrain camera and displays it on the HMD. Details of each step will be discussed later, in Section 3.

2.2 System Setup

First of all, in order to generate the novel view of the subject from any angle/position of the virtual viewpoint, the zoom level, angle, and position of each Dragonfly camera must be adjusted so that it can capture the whole subject even as he/she moves around. Moreover, to guarantee that the constructed visual hull is close enough to the object's shape, the zoom level and the position of each camera should be adjusted so that the camera looks at the subject at a far enough distance. The camera on top to view the subject from above also serves this purpose.

The system is very sensitive to the cameras' intrinsic and extrinsic parameters because the visual hull construction algorithm is based on the relative distances among cameras as well as the distances between the subject and the cameras. Consequently, after being adjusted, the position, zoom level, and angle of each camera have to be fixed so that the camera's parameters are not changed anymore. The next step is to calibrate all the cameras to get the necessary parameters. Both the Unibrain camera attached to the HMD and the Dragonfly cameras which capture the subject have to be calibrated. Their intrinsic parameters can be estimated using standard routines available with the ARToolkit [1] or MXRToolkit [3].

For the Dragonfly cameras, we must not only estimate the intrinsic parameters, but also the extrinsic parameters to get the spatial transformation between each of the cameras. Calibration data is gathered by presenting a large checkerboard to all of the cameras. For our calibration strategy to be successful, it is necessary to capture many views of the target in a sufficiently large number of different positions. Standard routines from Intel's OpenCV library [4] are used to detect all the corners on the checkerboard in order to calculate both a set of intrinsic parameters for each camera and a set of extrinsic parameters relative to the checkerboard's coordinate system. Where two cameras detect the checkerboard in the same frame, the relative transformation between the two cameras can be calculated. By chaining these estimated transforms together across frames, the transform from any camera to any other camera can be derived [25], [24].

3 SOFTWARE COMPONENTS

3.1 Overview

All of the basic modules and the processing stages of the system are represented in Fig. 2. The Capturing and Image Processing modules are placed at each Capture Server machine. After the Capturing module obtains raw images from the cameras, the Image Processing module will extract parts of the foreground objects from the background scene to obtain the sillhouettes, compensate for the radial distortion component of the camera mode, and apply a simple compression technique.

The Synchronization module, on the Synchronization machine, is responsible for getting the processed images from all the cameras and checking their timestamps to synchronize them. If those images are not synchronized, based on the timestamps, the Synchronization module will request the slowest camera to continuously capture and send back images until all these images from all nine cameras appear to be captured at nearly the same time.

The Tracking module will obtain the images from the Unibrain camera mounted on the HMD, track the marker pattern, and calculate the Euclidian transformation matrix relating the marker coordinates to the camera coordinates. Details about this well-known marker-based tracking technique can be found in [24], [1], or [3].

After receiving the images from the Synchronization module and the transformation matrix from the Tracking module, the Rendering module will generate a novel view of the subject based on these inputs. The novel image is generated such that the virtual camera views the subject from exactly the same angle and position as the head-mounted camera views the marker. This simulated view of the remote collaborator is then superimposed on the original image and displayed to the user. In subsequent parts of this paper, we will discuss the techniques we use in each module in more detail.

3.2 Image Processing Module

The Image Processing module processes the raw captured image in three steps: background subtraction (which extracts parts of the foreground objects from the image to obtain the silhouettes), radial distortion compensation, and image size reduction. The second step is done by applying the intrinsic parameters of the camera to estimate the correct position of each pixel. The remainder of this section will concentrate on the background subtraction and image size reduction steps.

3.2.1 Background Subtraction

The result of visual hull construction in the Rendering module largely depends on the output of the background subtraction step. This preprocessing step is one of the most crucial steps to determine the quality of the final 3D model. As it not only has to produce the correct foreground object, the chosen background subtraction algorithm must be very fast to fulfill the real-time requirement of this system. Another important requirement to guarantee the good shape of the visual hull is that the background subtraction algorithm must be able to eliminate the shadow caused by the objects.

There are many works on background subtraction which produce rather good results, such as [14], [28], [22]. However, there normally exists a significant trade-off between processing time and the quality of the result. The simple statistical method we used in our previous work on 3D Live [24] is very fast, but does not produce good enough quality. To fulfill our needs, we use a modified method based on the scheme of Horpraset et al. [14], which has good capabilities of distinguishing the highlighted and shadow pixels. However, this algorithm has been modified in our research to reduce the computational intensiveness and optimize for the real-time constraints of this system.

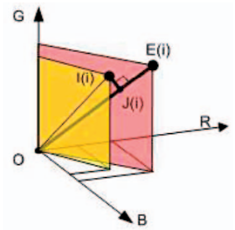


Fig. 3. Color model.

The main idea of this method is to learn the statistics of properties of each background pixel over N precaptured background frames and obtain the statistical values modelling for the background. The pixel properties to be calculated here are chromaticity and brightness, which is obtained from a new model of the pixel color. Based on this, the algorithm can then classify each pixel into "foreground," "background," "highlighted background," or "shadow/shading background" after getting its new brightness and chromaticity color values. In our application, we only need to distinguish the "foreground" type from the rest.

The new color model, which separates the brightness from the chromaticity component, is summarized in Fig. 3. In this RGB color space, the point $I(i)$ represents the color value of the $i\text{th}$ pixel and $E(i)$ represents the expected color value of this pixel, for which coordinates $(\mu_R(i), \mu_G(i), \mu_B(i))$ are the mean values of the R, G, B components of this pixel obtained from the learning stage. $J(i)$ is the projection of $I(i)$ on the line $OE(i)$.

The brightness distortion (α_i) and color distortion (CD_i) of this pixel are defined and calculated as:

$$\begin{aligned}\alpha_i &= \frac{J(i)}{E(i)}\\ &= argmin_{\alpha_i}\left[\left(\frac{I_R(i)-\alpha_i\mu_R(i)}{\sigma_R(i)}\right)^2+\left(\frac{I_G(i)-\alpha_i\mu_G(i)}{\sigma_G(i)}\right)^2\right.\\ &\qquad\left.+\left(\frac{I_B(i)-\alpha_i\mu_B(i)}{\sigma_B(i)}\right)^2\right],\end{aligned} \tag{1}$$

$$CD_i = \sqrt{\left(\frac{I_R(i)-\alpha_i\mu_R(i)}{\sigma_R(i)}\right)^2\left(\frac{I_G(i)-\alpha_i\mu_G(i)}{\sigma_G(i)}\right)^2\left(\frac{I_B(i)-\alpha_i\mu_B(i)}{\sigma_B(i)}\right)^2}. \tag{2}$$

In the above formula, $\sigma_R(i), \sigma_G(i), \sigma_B(i)$ are standard deviations of the $i\text{th}$ pixel's red, green, blue values computed in the learning stage. In our version, we assume that the standard deviations are the same for all pixels to make the CD_i formula simpler:

$$CD_i = (I_R(i)-\alpha_i\mu_R(i))(I_G(i)-\alpha_i\mu_G(i))(I_B(i)-\alpha_i\mu_B(i)). \tag{3}$$

Another assumption is that the distributions of α_i and CD_i are the same for all pixel i. With this assumption, we do not need to normalize α_i and CD_i as was being done in the previous work of [14].

Fig. 4. Results of background subtraction: before and after filtering.

These modifications reduce the complexity of the formula and quite drastically increase the calculation speed from 33ms/frame to 13ms/frame, but produce more small misclassified pixels than the original algorithm. However, these small errors can be easily filtered in the next step.

3.2.2 Filtering

The filtering step is necessary to remove the small misclassified regions. There are many filtering methods to process the images after background subtraction. However, regarding the real-time constraint, we use the simple morphological operators to open and close to filter out small misclassified regions.

3.2.3 Data Size for Real-Time Network Constraints

One very important factor is the amount of data to transfer over the network. In order to reach the fastest network speed, the size of the data has to be as small as possible. In our system, we try to optimize the data size by using the following two main methods:

- Reducing the image size by only storing the smallest rectangular region containing the foreground objects. An algorithm is implemented to find out the contour of the foreground and, based on this result, to calculate the smallest bounding box. This finding of the contour algorithm also acts as another filtering method, which filters all small misclassified foreground regions in which contour lengths are less than a predefined threshold. The size of this smallest rectangular region bounding the foreground objects depends on how closely the camera looks at the object and how large the object is. As mentioned in Section 2.2, all cameras must be adjusted so that they view the object from a far enough distance to guarantee the quality of the visual hull. Consequently, for each camera, the average size of this bounding box of the foreground is normally less than 1/8 the size of the whole image, which is a significant reduction in the data size.
- Using Bayer format [6] with background information encoded to store the images. Instead of using 3 bytes to encode three color components, Red, Green, Blue, for each pixel, we encode the whole image in Bayer format, which costs only 1 byte for each pixel. Moreover, for each pixel, the background information is encoded in the least significant bit of the byte at the position of this pixel in the Bayer image, value 1 for a background pixel and 0 for a foreground pixel. Obviously, this method of storing images leads to some color information lost. However, because the lost information is minimal, the color quality of the output images is still good. Consequently, the lost information is trivial compared with the benefit of greatly reducing data size, which is at least 3 times smaller than the RGB format with background information encoded.

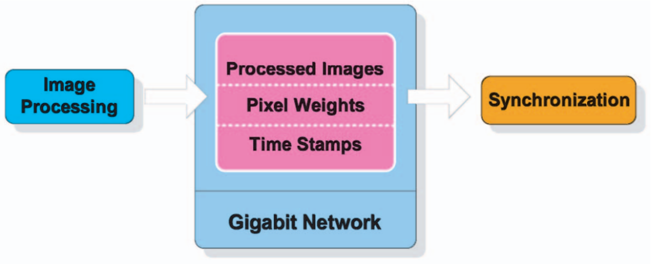


Fig. 5. Data transferred from Image Processing to Synchronization.

3.2.4 Results

The sample results of the image processing step are shown in Fig. 4. We can see that there are small errors after we subtract the background by our optimized algorithm. In the figure, the small green pixels inside the body are the foreground pixels misclassified as background ones and the small black pixels outside the body are the background pixels misclassified as foreground ones. However, these errors are completely removed after the filtering step. The speed of this step is only around 15ms/frame. Compared with the nonsimplified algorithm, which is 37ms/frame including the filtering step, the optimized algorithm is fast enough for this real-time application.

3.3 Synchronization

The main function of this module is receiving and synchronizing images which have been processed by the Image Processing module. The purpose of synchronization is to ensure that all images are captured at the same time.

Fig. 5 describes the data transferred from Image Processing to Synchronization. It includes three parts. The first part is the image which is processed by the Image Processing Module. Instead of sending the whole image, we only transmit the smallest rectangle area of the original image that contains the silhouette. This significantly reduces the amount of data to be transmitted. The second part is the pixel-weights for this image. These weights will be used for blending color in the rendering steps. We will present more about this weight in the Rendering section of this paper. The last part to be transmitted is the Time Stamp, which is the time when this image is captured. Using this timing information, the Synchronization module will synchronize images captured from all nine cameras.

Once one set of images from nine cameras is received, the time stamp of each image will be compared. If the difference in time between the fastest camera and the slowest camera is greater than 30 ms, the Synchronization module will require the Image Processing module to provide a new image from the slowest camera. This synchronizing process will keep looping until the difference is less than 30ms. The reason to choose 30ms as the threshold is because our system operates at 30 fps.

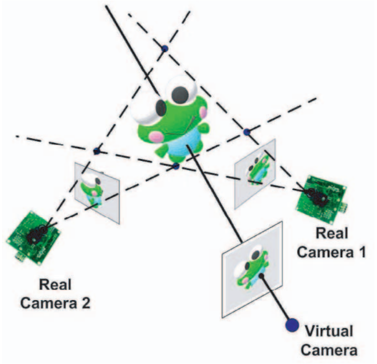


Fig. 6. Novel viewpoint is generated by the visual hull.

3.4 Rendering

Our rendering algorithm used in this system is a new development over our previous algorithm, which is described in [25]. To improve the speed and quality, we introduce new ways to compute visibility and blend color in generating images for novel viewpoints. In this section, the main algorithm will first be briefly described. After that, improvements for speed and quality will be presented.

3.4.1 Overview of the 3D Human Rendering Algorithm

Our algorithm proceeds entirely on a per-pixel basis. We denote the desired image, the "virtual camera image," and its constituent pixels, "virtual pixels." The virtual camera can be determined by taking the product of the (head mounted) camera calibration matrix and the estimated transformation matrix. Given this 4 x 4 camera matrix, the center of each pixel of the virtual image is associated with a ray in space that starts at the camera center and extends outward. Any given distance along this ray corresponds to a point in 3D space. We calculate an image-based depth representation by seeking the closest point along this ray that is inside the visual hull. This 3D point is then projected back into each of the real cameras to obtain samples of the color at that location. These samples are then combined to produce the final virtual pixel color. In summary, the algorithm must perform three operations for each virtual pixel: determining the depth of the virtual pixel as seen by the virtual camera, finding corresponding pixels in nearby real images, and determining pixel color based on all these measurements. We briefly describe each of these operations in turn.

Determining Pixel Depth. The depth of each virtual pixel is determined by an explicit search starting at the virtual camera projection center and proceeding outward along the ray corresponding to the pixel center (see Fig. 6). Each candidate 3D point along this ray is evaluated for potential occupancy. A candidate point is unoccupied if its projection into any of the silhouettes is marked as background. When a point is found for which all of the silhouettes are marked as foreground, the point is considered occupied and the search stops.

Using this method, we can generate the visual hull very efficiently. One problem with the visual hull is that the geometry it reconstructs is not very accurate. When photographed by only a few cameras, the scene's visual hull is much larger than the true scene [30]. One well-known improvement for the visual hull which was discussed in [16], [27], [29], [30], [31], and [32] is to utilize color constraint. Although, using this constraint, we can generate "photo-hull," which is a better approximation than visual-hull, the rendering speed will be decreased significantly and, thus, it is not suitable for real-time applications. Alternatively, we reduce the errors of the visual hull by using more cameras and a larger recording room.

Finding Corresponding Pixels in Real Images. The resulting depth is an estimate of the closest point along the ray that is on the surface of the visual hull. However, since the visual hull may not accurately represent the shape of the object, this 3D point may actually lie outside of the object surface. Hence, care needs to be taken in choosing the cameras from which the pixel colors will be combined. Depth errors will cause incorrect pixels to be chosen from each of the real camera views.

To minimize the visual effect of these errors, it is better to choose the incorrect pixels that are physically closest to the simulated pixel. So, the optimal camera should be the one minimizing the angle between the rays corresponding to the real and virtual pixels. For a fixed depth error, this minimizes the distance between the chosen pixel and the correct pixel. We rank the cameras proximity once per image, based on the angle between the real and virtual camera axes.

We can now compute where the virtual pixel lies in each candidate cameras image. Unfortunately, the real camera does not necessarily see this point in space—another object may lie between the real camera and the point. If the real pixel is occluded in this way, it cannot contribute its color to the virtual pixel. In the previous versions of this research, we increased the system speed by intermediately accepting points that are geometrically certain not to be occluded. However, this geometrical information does not always provide true occlusion. As we can see in Fig. 9, in the left image, we still can see the false shadows of two hands over the body. These false hand shadows are generated because these parts of the body are occluded from the reference cameras by the two hands, but the geometrically-based method cannot detect it. To achieve better results, in this new version, we introduce a new method to compute occlusion.

Determining Virtual Pixel Color. After determining the depth of a virtual pixel and which cameras have an unoccluded view, all that remains is to combine the colors of real pixels to produce a color for the virtual pixel. In the previous research, we took a weighted average of the pixels from the closest N cameras such that the closest camera is given the most weight. This method can avoid producing sharp images that often contain visible borders where adjacent pixels were taken from different cameras. However, there are still some errors along the edge of the silhouette. In the next section, we propose a new method to blend color which can overcome this problem.

3.4.2 New Algorithm Methods for Speed and Quality

Occlusion Problem. As said above, one of the main issues of this algorithm is the occlusion problem. In order to compute visibility, one basic approach is searching in 3D space. To determine if a point A is visible from one camera, we can simply search, point by point, from A toward the center O of this camera. If any point in this ray belongs to the visual hull, A is considered to be invisible from this camera (Fig. 7).

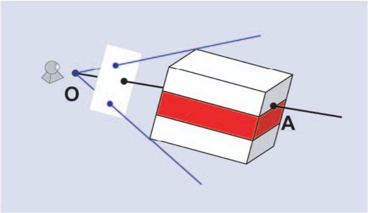


Fig. 7. Example of occlusion. In this figure, A is occluded from camera O.

Instead of brute-force searching in 3D space, [30] proposed a more efficient way which only needs to step along epipolar lines. However, with this method, we still need to search on all captured images. To further increase the speed, we introduce a new method which only requires searching on one captured image.

To compute visibility, Matusik et al. introduced a novel algorithm which can effectively reduce 3D visibility computation to 2D visibility computation [21]. The main idea of this algorithm can be illustrated in Fig. 8. In this figure, camera K is chosen so that the projection Q of P on this camera lies on the edge of the silhouette. This algorithm is based on the fact that the 3D point P has to be visible from the camera K if, on the image plane of one camera K, the 2D point Q is visible from the epipole E (the projection of the center of projection of camera K onto the image plane of camera J). In their paper, they use this algorithm to determine the visibility of each face of the visual hull, but we apply it to compute the visibility of each point of the image-based visual hull. Our algorithm can be summarized as follows:

To determine if point P is visible from camera K, the following three steps will be processed:

1. Find one camera J where the project Q of P lies on the edge of the silhouette.
2. Find the epipole E of camera K on the image plane of camera J.
3. If there is any foreground pixel lying on the line connecting point Q and point E, i.e., Q is occluded from point E, then P will be considered to be occluded from camera K. Otherwise, P will be consider to be visible from camera K.

Using this algorithm, we can avoid 3D searching while still being able to detect occlusion whenever it happens.

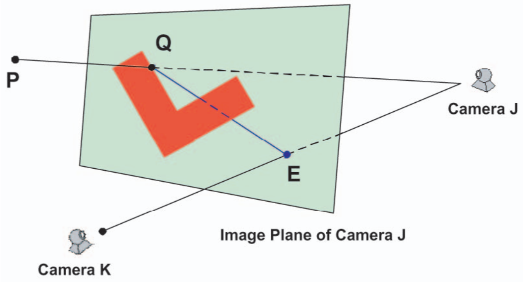


Fig. 8. Visibility computation: Since the projection Q is occluded from the epipole E, 3D point P is considered to be invisible from camera K.

However, this algorithm is overly conservative [21]. It never considers a point visible if this point is occluded, but, sometimes, it considers a point occluded which is in fact visible. As a result, some points in the visual hull will be computed to be occluded from all cameras, which leads to holes in the results. To compensate for this, whenever a point is computed to be invisible from all cameras, we do not accept that, but use the previous version to recompute visibility. The negative effect of this is, for some points, we need to run both methods, but, normally, there are only a few points like that. Thus, it does not affect the overall speed in any significant way.

Fig. 9 shows example rendering results. In the left image, we use geometric information to compute visibility, while, in the right, we use the above-described visibility computing algorithm. As one can see, in the upper image, there are false shadows of two hands over the body, while there are not in the lower image.

Table 1 shows the frame rate we can achieve with our algorithm. All three visibility algorithms, 3D searching, geometrically-based, and our new algorithm, are tested. We also tested with two different resolutions: 320 x 240 and 640 x 480. As we can see, our new method is much faster than 3D searching method. With this new algorithm, we can achieve 23 fps at 320 x 240 and 11 fps at 640 x 480, while, with 3D searching, it is only 7 fps and 3 fps, respectively. Compared with the geometrically-based method, the new method is a little slower, but it provides better results.

New Method for Blending Color. The second improvement is a new method to blend color for the visual hull. Most of current shape-from-silhouette algorithms use the angles

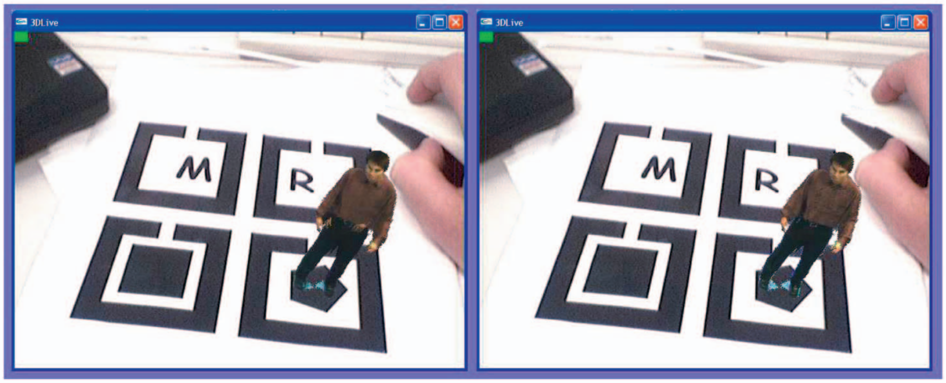


Fig. 9. Rendering results: In the left image, we use geometric information to compute visibility, while, in the right, we use our new visibility computing algorithm. One can see the false hands appear in the upper image.

TABLE 1
Rendering Speed

IMAGE SIZE	3D SEARCHING	NEW ALGORITHM	GEOMETRICAL-BASED
320 x 240	7 fps	23 fps	27 fps
640 x 480	3 fps	11 fps	13 fps

between the desired view and reference views to decide the weights for blending. However, this can cause errors along the edges of foreground images because background subtraction usually generates errors in these areas. For example, in Fig. 10, if we base on the angles of the cameras, point A will get color from camera 2, which is a closer angle to the novel viewpoint. However, the projection of A to camera 2 is at the edge of the silhouette, which usually contains some errors due to the background subtraction.

To address this issue, we utilize a technique from image-mosaicing. In this subject of image-mosaicing, to reduce visible artifacts, that is, to hide the edges of the component images, one can usually use a weighted average with pixels near the center of each image contributing more to the final composite [32]. Similarly to this idea, in our algorithm, to determine the color of the virtual pixel, we take a weighted average with pixels near the center of each silhouette having higher weights. Thus, in Fig. 10, if we use this blending method, A will get color from camera 1, where the projection of A is closer to the center of the silhouette. This new blending method makes the visual hull smoother along the edges of silhouettes.

One problem with this blending method is that it requires more memory and time to store and calculate the weights as each pixel of each reference images got different weights. To increase the speed, instead of computing these pixel weights during rendering, we calculate them during the image processing process. In such a way, we can run this calculation on three different computers, each in charge of images captured from three cameras. This will triple the speed. Thus, for each captured image, the Image Processing module will calculate the weights for each pixel and then pass these weights for the rendering module.

Fig. 11 shows one set of images from nine cameras and their corresponding pixel weights. The brighter one pixel is, the higher weight it gets. Fig. 12 shows two rendering results. The left is rendered with camera weights, while the right is rendered with pixel weights. As we can see, using pixel weights, the result is better and smoother, especially along the edge of silhouettes.

4 Magic Land: An Application of the Live Mixed Reality 3D Capture System for Art and Entertainment

With the abilities of capturing, sending, regenerating the 3D images of live humans and objects in real time, and displaying these objects' 3D images in the augmented reality enviroment, 3D Live technology has many applications in various fields.

The first obvious application is a three-dimensional video-conferencing and collaboration system, which is much better than the traditional 2D video-conferencing system in terms of communication benefits. This is because the 3D images displayed in a real environment can fully represent nonverbal communication such as gestures, which the traditional 2D system cannot. Moreover, using the 3D system, users not only can arrange markers representing several collaborators about them to create a virtual spatial conferencing space, but also can potentially conference from any location and, thus, the remote collaborators become part of any real-world surroundings, potentially increasing the sense of social presence.

Another application of the 3D Live system in education and entertainment is an augmented book in which a different fiducial marker is presented on each page and associated with each is virtual content consisting of both 3D graphics and a narrator who was captured in our system. Other applications of this system in training, entertainment, computer games, etc., can be seen in [25].

The remainder of this section will fully describe a novel application of 3D Live in art and entertainment. This system, named Magic Land, is the cross-section where art and technology meet. From a technology viewpoint, it is a combination and demonstration of the latest advances in human-computer interaction and human-human communication: mixed reality, tangible interaction, and 3D Live

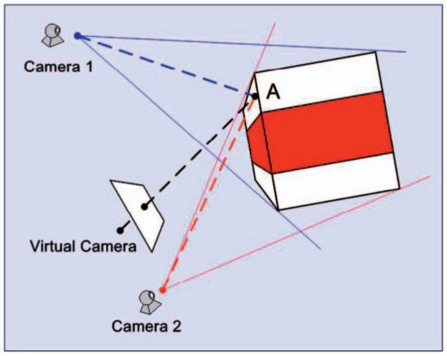


Fig. 10. Example of blending color.

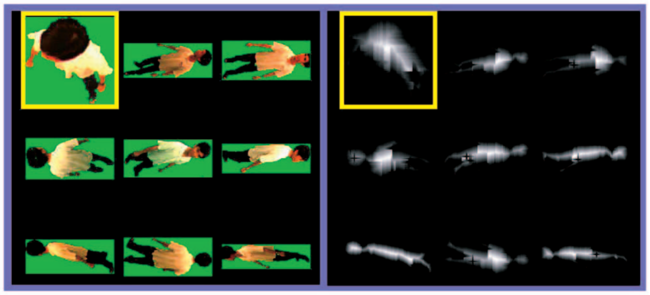


Fig. 11. Original images and their corresponding pixel weights.

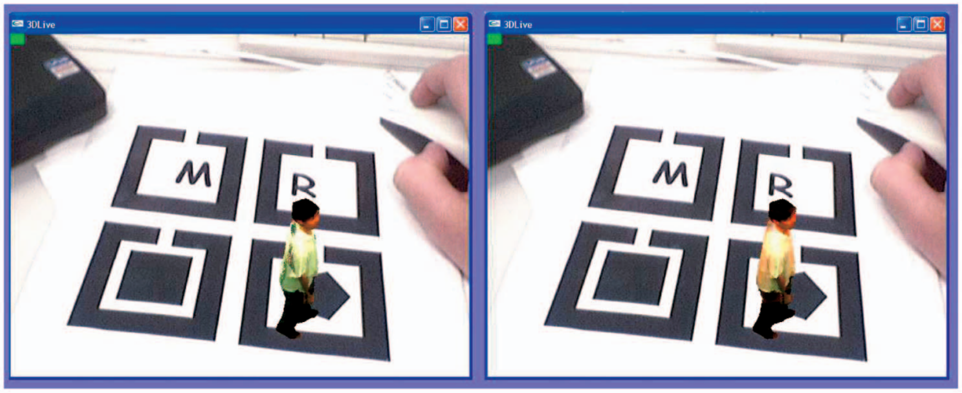


Fig. 12. Rendering results: The right is with the pixel weights algorithm, while the left is not. The right image shows a much better result, especially near the edges of the figure.

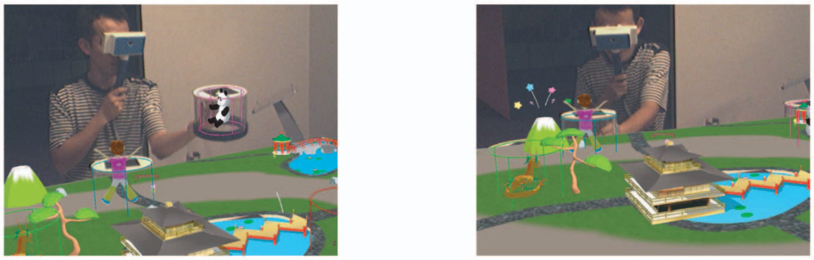


Fig. 13. **Tangible interaction on the Main Table**: (Left) Tangibly picking up the virtual object from the table. (Right) The trigger of the volcano by placing a cup with a virtual boy physically near to the volcano.

technology. From an artistic viewpoint, it aims to introduce tangible approaches of dealing with mixed reality content to artists of any discipline. These approaches, which allow artists to manipulate the mixed reality content intuitively and easily by using cups, was also presented in [15] for a city planning application.

Another main purpose of the Magic Land system is to bring to all users a new special kind of human self reflection and human-human interaction. In this system, users can tangibly pick up themselves or their collaborators and watch them in 3D form encountering other virtual objects. In order to allow users to manipulate their own 3D recorded images in a mixed reality environment, this version of Magic Land does not fully exploit the "live" capturing feature of 3D Live, but, instead, utilizes the fast processing and rendering algorithms for fast 3D Live record and playback features. However, another version of Magic Land, which can be easily built for live capture and live viewing, is discussed further in Section 4.4. The artistic intention and motivation of the project will also be discussed further in Section 4.3.

4.1 System Concept and Hardware Components

Magic Land is a mixed reality environment where 3D Live captured avatars of human and 3D computer generated virtual animations play and interact with each other.

The system includes two main areas: recording room and interactive room. The recording room is where users can have themselves captured into live 3D models which will interact in the mixed reality scene. This room, which has nine Dragonfly cameras mounted inside, is a part of the 3D capture system described above. After the user gets captured inside the system, she can go to the interactive room to play with her own figure.

The interactive room consists of three main components: a Menu Table, a Main Interactive Table, and five playing cups. On top of these tables and cups are different marker patterns. A four cameras system (ceiling tracking system) is put high above the Main Interactive Table to track the relative position of its markers with the markers of the cups currently put on it. The users view the virtual scenes and/or virtual characters which will be overlaid on these tables and cups via the video-see-through HMDs with the Unibrain cameras mounted in front and looking at the markers. The Main Interactive Table is first overlaid with a digitally created setting, an Asian garden in our case, whereas the cups serve as the containers for the virtual characters and also as tools for users to manipulate them tangibly. There is also a large screen on the wall reflecting the mixed reality view of the first user when he/she uses the HMD. If nobody uses this HMD for 15 seconds, the large screen will change to the virtual reality mode, showing the whole magic land viewed from a very far distant viewpoint.

An example of the tangible interaction on the Main Interactive Table is shown in Fig. 13. Here, we can see a user using a cup to tangibly move a virtual panda object (left image) and using another cup to trigger the volcano by putting the character physically near the volcano (right image).

The Menu Table is where users can select the virtual characters they want to play with. There are two mechanical push buttons on the table corresponding to two types of characters: the human captured 3D Live models on the right and VRML models on the left. Users can press the button to change the objects shown on the Menu Table and move the empty cup close to this object to pick it up. To empty a cup

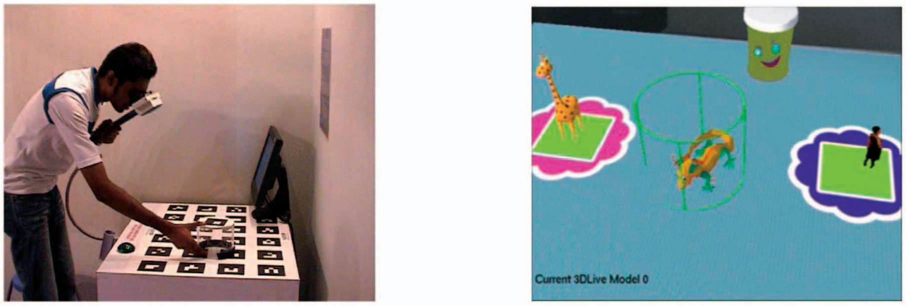


Fig. 14. **Menu Table**: (Left) A user using a cup to pick up a virtual object. (Right) Augmented view seen by users.


(a)
(b)

Fig. 15. **Main Table**: (a) The witch turns the 3D Live human which (b) comes close to it into a stone.

(trash), users can move this cup close to the virtual bin placed at the middle of the Menu Table. In Fig. 14, in the left image, we can see a user using a cup to pick up a virtual object; at the edge of the table closest to the user are two mechanical buttons. In the right image, we can see the augmented view seen by this user. The user had previously selected a dragon which is inside the cup.

After picking up a character, users can bring the cup to the Main Interactive Table to play with it. Consequently, there will be many 3D models moving and interacting in a virtual scene on the table, which forms a beautiful virtual world of those small characters. If two characters are close together, they would interact with each other in the predefined way. For example, if the dragon comes near to the 3D Live captured real human, it will blow fire on the human. This gives an exciting feeling of the tangible merging of real humans with the virtual world. As an example of the interaction, in Fig. 15, we can see the interaction where the witch which is tangibly moved with the cup turns the 3D Live human character which comes physically close to it into a stone.

4.2 Software Components

As shown in Fig. 16, the software system of Magic Land consists of five main parts: 3D Live Recording, 3D Live Rendering, Main Rendering, Ceiling Camera Tracking, and Game Server. Besides these parts, there is a Sound module that produces audio effects, including background music and interactive sounds for the whole system.

In this system, users can record their live model for playback. The 3D Live Recording and 3D Live Rendering parts are a recording capturing system described in the previous section. After going inside the recording room and pressing a button, the user will be captured for 20 seconds. The captured images are then processed and sent to all 3D Live Rendering modules. However, unlike the live version, which sends the processed images of nine cameras immediately for each frame, the recorded version sends all the processed images of all the frames captured in 20 seconds at a time. Another difference is that, instead of using TCP/IP to send the 3D Live data to each User 3D Live Rendering and Menu 3D Live Rendering module of the 3D Live Rendering part, we use multicast to send the data to all of them. This helps to utilize the bandwidth of the network as well as to ensure that all the receivers finish receiving data at the same time.

The Main Rendering part includes a Menu Rendering module and five User Rendering modules. These modules track the users' viewpoints and render the corresponding images to the users. First, they obtain images from the Unibrain cameras mounted on the users' HMDs, track the marker patterns, and calculate the transformation matrix relating the coordinates of these markers with the coordinate of the camera. After that, based on the transformation matrix, each module will render the image and output the result to the corresponding HMD. The Menu Rendering module especially also handles the users' inputs when they press the buttons on the Menu Table or when they use the cups to select and remove virtual characters.

The Ceiling Camera Tracking module receives images from four cameras placed above the Main Interactive Table. It tracks the markers of the table and cups and calculates the transformation matrices of the cups relative to the table from the top view. After that, it sends these matrices to the Game Server.

Last, but not least, the Game Server is the heart of the system, which links all the modules together. It receives and forwards information from the Ceiling Camera Tracking, Menu Rendering, and User Rendering modules. This Game Server coordinates and synchronizes what every user has in their cup in terms of type of the character and its animation, position, and orientation. First of all, it receives the camera tracking data from the Ceiling Camera Tracking module and determines the interaction between the characters inside the cups, based on the distances between cups. After

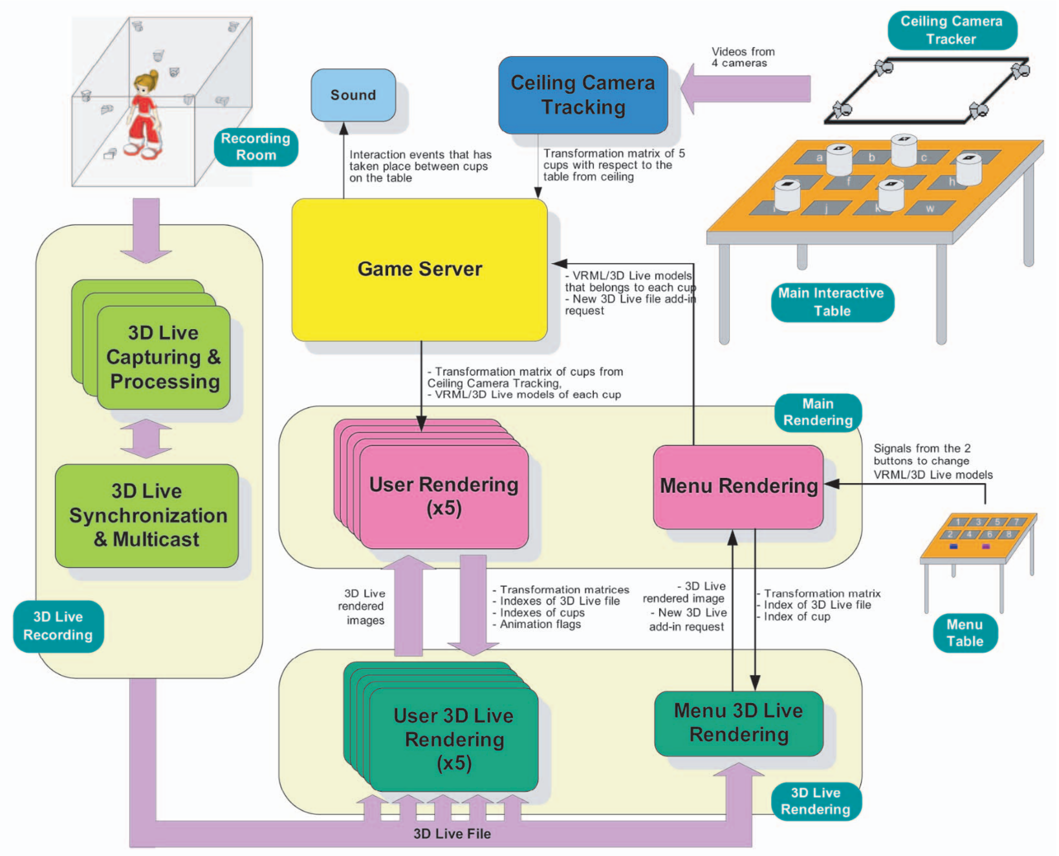


Fig. 16. System setup of Magic Land.

that, it forwards this interaction information to the User Rendering and Sound modules so that these modules can render the respective animations and produce the corresponding interactive sound. The ceiling camera tracking data is also forwarded to the User Rendering modules for use in case that the users' camera lost the tracking of their cups' marker. When the users select a new character, the Game Server also receives the new pair of cup-character indexes from the Menu Rendering and forwards to all the User Rendering modules to update the change.

4.3 Artistic Intention

Magic Land demonstrates novel ways for users in real space to interact with virtual objects and virtual collaborators. Using the tangible interaction and the 3D Live human capture system, our system allows users to manipulate the captured 3D humans in a novel manner, such as picking them up and placing them on a desktop, and being able to "drop" a person into a virtual world using the users' own hands. This offers a new form of human interaction where one's hands can be used to interact with other players captured in 3D Live models.

The artistic aspect of this installation introduces to artists easy, tangible, and intuitive approaches for dealing with mixed reality content. The main challenge of the project is to create a new medium located somewhere between theater, movie, and installation. The outcome of the project is an infrastructure that gives artists new opportunities to transport audiovisual information and encourage artists of any discipline to deal with those new approaches.

We can perceive Magic Land as an experimental laboratory that can be filled by a wide range of artistic content, which is only limited by the imagination of the creators. To watch the scene from above with the possibility of tangible manipulation of elements creates a new form of art creation and art reception that generates an intimate situation between the artist and audience.

The project itself brings together the processes of creation, acting, and reception in one environment. These processes are optimized to the visitors experience in order to better understand the media and lead to a special kind of self reflection. The recording area plays the role of the interface between human being and computer. It is also a special experience for the users to watch themselves acting in 3D on the interactive table from the external point of view, like a "Bird in the sky." Fig. 17 showns two bird's eye views of this system.

4.4 Future Work

Currently, we are developing a new version of Magic Land by exploiting the "real time" capability of 3D Live technolog, in which outside players can see, on the Main Interactive Table, the 3D images of the one who is being captured inside the room in real time. Instead of sending all the processed images of all the frames captured in 20 seconds at a time, this version uses RTP [26] and IP multicast to stream the processed images to all User 3D Live Rendering modules immediately for each frame. To guarantee continuous rendering, User 3D Live Rendering modules will buffer these images for a number of received frames before generate the 3D images inside one of the special cup on the Main Interactive Table. Moreover, inside the recording room, the captured player wears an HMD to view the virtual environment in front of her at the viewpoint corresponding to the position of the cup on the

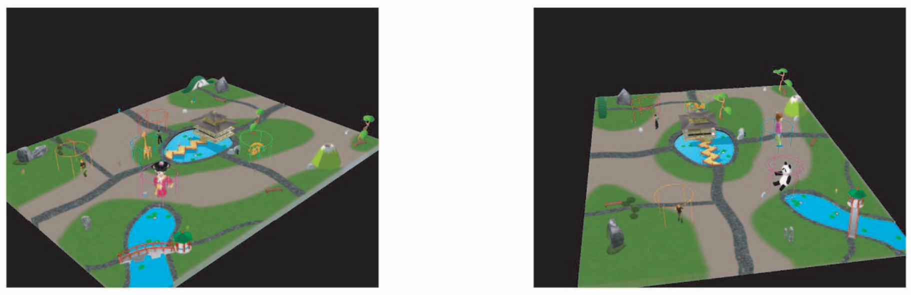


Fig. 17. **Main Table**: The bird's eye views of the Magic Land. One can see live captured humans together with VRML objects.

table. The HMD is connected to a computer outside by a small cable going through the ceiling of the recording room. The cable is painted the same color as the room and its width is small enough to be eliminated by the filter step of the 3D Live background subtraction and image processing modules.

In this context, the captured player can actively interact with other virtual objects in the virtual reality environment when seeing them on the HMD and the outside players will have fun seeing her reaction in the mixed reality environment. In our further future work, we want to explore the problem of whether the cup which represents the 3D Live object can automatically move when the captured player moves inside the room. Such a system will give the captured person more freedom exploring the whole virtual world herself. Technologies in Touchy Internet [17] can be applied to automatically move the special cup around the table. Touchy Internet uses special sensors and a wireless system to track the movement of a pet at home in the backyard and control a doll's movement placed at the office corresponding to the pet's movement.

The future version of Magic Land will open a new trend for mixed reality games in which players can actively play the role of a main character in the game story, be submerged totally in the virtual environment, and explore the virtual world themselves, while, at the same time in the mixed reality environment, other players can view and construct the virtual scene and new virtual characters to challenge the main character. Consequently, the game story is not fixed, but will depend on the players' creativity and imagination and follow their reactions when they travel around the virtual word.

4.5 Magic Land's Relationship to Mixed Reality Games

Nowadays, computer games have become a dominating form of entertainment due to their higher level of attractiveness to game players. There are some superior advantages which make computer games more popular than traditional games. First, they attract people by creating the illusion of being immersed in an imaginative virtual world with computer graphics and sound [7]. Second, the goals of computer games are typically more interactive than those of traditional games, which gives players a stronger desire to win the game. Third, usually designed with the optimal level of information complexity, computer games can easily provoke players' curiosity. Consequently, computer games intrinsically motivate players by bringing them more fantasy, challenge, and curiosity, which are the three main elements contributing to the fun in games [18]. Moreover, compared with many traditional games, computer games are also easier to play at any individual's preferred location and time.

However, the development of computer games has often decreased players' physical activities and social interactions. Addressing this problem, the growing trends of current games, especially mixed reality games, are trying to fill in this gap by bringing more physical movements and social interactions into games while still utilizing the benefit of computing and graphical systems.

A typical VR game, CAVE Quake [2], increases the players' sense of 3D space by surrounding them with a life-sized 3D virtual world, instead of constraining them within a limited 2D screen. However, CAVE Quake players still lack of physical movement, tangible interactions, and social communications.

AR2 Hockey [23], an air-hockey AR game in which users use a real mallet to play with a virtual puck on a real table, enhances physical interactions and social communication, but does not utilize the graphical power of computer systems.

AquaGauntlet [33] is another AR game in which several players gather in a small place with some physical egg-shaped objects to shoot computer-generated creatures superimposed onto the real scene as if they came from these egg-shape objects. This game enhances physical interactions and social communication and also utilizes the graphical power of computer system. However, players of AquaGauntlet, as well as AR2 Hockey, still have limited movement and little interaction with the physical space (as they must stand in a fairly constant location).

Another embodied computing-based mixed reality game which also enhances physical interactions and social communication is Touch-Space [12]. This game is carried out in the physical world with a room-size space where two players will collaboratively finish some tasks and then rescue a princess in a castle controlled by a witch. This game provides different levels of interaction in different environments: physical, augmented reality, and virtual reality. However, all these interactions are limited to a room-size space and only two users.

Pirates! [10] and Human Pacman [11] are two typical outdoor mixed reality games aimed at enhancing physical activities and social interactions to as great an extent as possible. Pirates! uses handheld computers and proximity-sensing technology to make real-world properties, such as locations or objects, important elements of game mechanics. Meanwhile, in Human Pacman, the player who acts as "Pacman," wearing a wearable computer and an HMD, goes around the physical game space to collect cookies, where another player acting as "Ghost" will find and touch to kill the Pacman. There are two other players acting as Pacman's and Ghost's helpers, sitting inside offices, using a computer's graphical information to search for their enemy's locations in order to help their partners. These games are very successful in term of enhancing physical

TABLE 2
Comparison of Magic Land with Other Mixed Reality Games

Games	**A**dvantages	**D**isadvantages
CAVE Quake	Significantly increase players' sense of 3D space by fully immersing them into a 3D virtual world. Provide beautiful graphics and interesting game story.	Very limited physical movement. No tangible interaction and social communication.
AR2 Hockey	Provide 3D mixed reality experience and tangible interaction with virtual object. Enhance social communication.	Limited physical movement and tangible interaction. No attractive 3D graphics of virtual world.
AquaGauntlet	Provide 3D mixed reality experience, tangible interaction and nice 3D graphics of virtual characters. Enhance social communication.	Limited physical movement and tangible interaction.
Touch Space	Tangible interaction with virtual object, enhance social communication, nice graphical virtual characters in mixed reality world. Different levels of interaction in different environments: physical environment, augmented reality, and virtual reality.	Limited physical movement and number of players.
Pirates!	Provide physical movement and social interaction to great extent.	Limited tangible interaction and graphical virtual characters. No 3D mixed reality experience.
Human Pacman	Provide physical movement and social communication to large extent. Enhance 3D mixed reality experience and tangible interaction.	Physical movement is slightly limited due to wearable computer and HMD.
Magic Land	Provide varied tangible interaction with virtual objects, beautiful 3D mixed reality virtual scene and characters, and social interactions among players. *Players can be captured and become new characters encountering with other virtual characters in mixed reality world.*	Not fully provide physical movement like outdoor games such as Pirates! and Human Pacman.

interactions and social communications, however, they have not fully utilized the graphical power of a computing system to create an appealing imaginative virtual world. Pirates! is played on a PDA screen, which does not allow a 3D mixed reality experience. Human Pacman requires quite heavy and bulky wearable computers and equipment.

As summarized in Table 2, compared with the above typical AR/VR games, as an indoor mixed reality and tangible interaction game, Magic Land exploits physical tangible interaction, social interaction, and also utilizes 3D graphics rendering to create an attractive imaginative virtual world. Moreover, the act of putting 3D images of real human beings into that inventive world and making them new characters of the game story is unique in game context. Most importantly, Magic Land is a kind of "free play" game [19] in which players are free to use their imagination and creativity to design the game story and rules. Thus, as mentioned before, the game story and rules is not fixed, but depend on the players' imagination and decisions.

5 USER STUDY OF THE MAGIC LAND SYSTEM

5.1 Aim of This User Study

We conducted this user study to obtain feedback from the users regarding their perception of our Magic Land system, for example, their feeling on interacting with virtual objects, being captured in 3D in a special recording room, etc. This survey also helps to assess how much this system promotes social interaction and remote 3D collaboration. The improvements that may continue to be made in future work are also expected to benefit from this user study.

5.2 Design and Procedures

Thirty subjects (13 females and 17 males) were invited to participate in this study. The age group of the subjects ranged from 15 to 54 years old, with the average age of 25.4 years old. All of them reported clear vision and normal hearing abilities.

During the user study, each subject went into the 3D Live recording room first and followed the system voice instructions to record herself. After the recording was finished, the system asked her to leave the recording room and go to the Menu Table and wait for her 3D data to be transferred over. Once her captured 3D Live data was sent to the Menu Table, the subject could then use the green button on the Menu Table to find herself among the various recorded human characters. Once she had found herself, she could then use one of the empty cups to pick herself up and put herself onto the main interaction table. She could then go and pick some more captured human 3D Live characters and virtual 3D VRML characters and add them to the interaction table and try interactions among them. Subjects were also encouraged to play this system with their friends at the same time (social collaboration).

After the subjects tried all the functions of this system, they were asked to fill in a questionnaire paper with 13 questions, as shown in Table 3 which can be found on the Computer Society Digital Library at http://computer.org/tvcg/archives.htm.

5.3 Results of This User Study

Questions 1 and 2 are used to assess the overall feelings of the subjects about the Magic Land system. The two main features here are merging the user into the virtual world and interacting with other virtual objects. From the feedback, we found that 25 subjects out of 30 felt *Very Exciting* about the concept of merging themselves into the virtual world and 20 subjects felt *Very Exciting* about the concept of interacting with virtual object. From these results, we can see that this technology is indeed very attractive to the general public.

Question 3 concerns how much this technology can help in promoting social interaction. The feedback was quite positive. In total, 20 subjects felt that this system can help to promote social interaction and six of them felt it is *Very helpful*. Questions 4, 5, and 6 concern about the 3D Live recording room. Eighteen subjects felt that the 3D Live recording process is *Comfortable and good* and nine felt it was *Moderate*. Only three subjects felt that the recording processing was *Uncomfortable* or made them *Nervous and feel uneasy*. Out of all of the testing subjects, 73.4 percent of them felt this system can be used for remote 3D collaboration in the future and 63.4 percent of them believed such a system would be collaborative. It shows that this 3D Live capturing process can be accepted by most of the population. The feedback on Question 8 shows that nearly two-thirds of the testing subjects think this system is useful in telepresence compared to current 2D video teleconferences.

Another important part of this Magic Land system is that we are using physical cups to pick up and move the virtual objects or 3D Live characters. From the answers to Questions 9, 10, and 11, we can see that most testing subjects like the method of using physical cups compared to using mice and keyboards as in traditional computer games. Comparing to mice and keyboards, 17 subjects felt that the cups were easier for picking up virtual objects and 18 subjects felt it was easy for them to move the objects around using the cups. Also, 18 subjects felt that using cups is helpful in promoting social interaction.

As a multimedia system, we also evaluate how entertaining this system is through Question 7. In the results, 10 subjects enjoyed the game a lot and 11 said *It is a nice game. Good for playing occasionally*. This result is quite encouraging for applicatios of this technology in further digital entertainment development. To check how friendly the user interface is, we used Question 12 to see how the users felt about the way of deleting a virtual object inside the cup. It shows that 70 percent of the subjects like our idea of using the virtual trash can. And, from Question 13, we can see that more than 90 percent of the subjects would like to try this kind of system again in the future.

5.4 Conclusions of the User Study

Overall, from the user study, we can conclude that our Magic Land system is shown to have produced a tangible, natural, and novel interaction interface to the users. The diagrams of the results for all 13 questions can be seen in Fig. 18 which can be found on the Computer Society Digital Library at http://computer.org/tc/archives.htm.

Most testing subjects claimed that this system is very attractive and they were excited to see themselves being captured in 3D and then being put into the interaction table together with the other 3D objects. Although a few people complained that the capturing process makes them feel uncomfortable or nervous, most of them felt comfortable or natural with the system. So, we can say that this system is acceptable to the general public and maybe minor modifications can be made to make it more user friendly.

From these results, we can see that most of the test subjects felt that mixed reality technology helps to promote social interaction among participants. More than half of the participants think this technology will be useful for a remote 3D collaboration system in the future. But, a few of them still think there is little collaboration in this system or no collaboration at all. The reason for this could be that all the 3D Live characters we use now are captured separately, without any relationship among them. But, when the technology is used in remote 3D collaboration in the future and the captured characters must be related, users should feel differently.

Using physical cups instead of traditional mice and keyboards also proved to be a more natural way of controlling virtual objects from the results of the user study. Most participants felt it was easy to use and helpful in promoting social interaction. Also, we can see that most users think it is a good idea to use the virtual trash can to delete the objects. This result shows that mixed reality technology provides a natural user interface.

Additionally, further improvements in this system may be made by increasing the gaming complexity and hardware refinement. There were 30 percent of the testing subjects who still felt that this system is not so entertaining. We can improve that by adding more meaningful interactions, 3D sound effects, better computer graphics, etc.

6 CONCLUSION

This paper has described a complete system for novel real-time capturing and rendering of 3D images of live subjects in a mixed reality environment. We believe that this is a significant step toward the goal of perfect "tele-presence" for remote collaborations in the near future. The hardware and software issues, together with new and novel improving algorithms and methods to speed up the system and obtain good quality, have been discussed in detail. We also presented Magic Land, a novel application of 3D Live in art and entertainment. Results of the survey on Magic Land's users reveal some important issues and emphasize the effectiveness of 3D Live, mixed reality, and tangible interaction on human-computer interaction. Our future work is to continue to improve the quality and speed of the whole 3D Live system, especially the image processing, 3D Live rendering, and networking parts, and to develop Magic Land as a new trend of mixed reality game.

REFERENCES

[1] ARToolKit, http://www.hitl.washington.edu/artoolkit/, 2005.
[2] CAVE Quake II, http://brighton.ncsa.uiuc.edu/prajlich/caveQuake/, 2005.
[3] MXRToolkit, http://sourceforge.net/projects/mxrtoolkit/, 2005.

[4] OpenCV, http://sourceforge.net/projects/opencvlibrary/, 2005.

[5] Point Grey Research Inc., http://www.ptgrey.com, 2005.

[6] RGB "Bayer" Color and MicroLenses, http://www.siliconimaging.com/RGB Bayer.htm, 2005.

[7] A. Amory, K. Naicker, J. Vincent, and C. Adams, "The Use of Computer Games as an Educational Tool: Identification of Appropriate Game Types and Elements," pp. 311-321, *British J. Educational Technology,* 1999.

[8] R. Azuma, "A Survey of Augmented Reality," *Presence,* 1997.

[9] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre, "Recent Advances in Augmented Reality," *IEEE Computer Graphics and Applications,* Nov./Dec. 2001.

[10] S. Bjork, J. Falk, R. Hansson, and P. Ljungstrand, "Pirates!—Using the Physical World as a Game Board," *Proc. Interact 2001, IFIP TC. 13 Conf. Human-Computer Interaction,* 2001.

[11] A.D. Cheok, K.H. Goh, W. Liu, F. Farzbiz, S.W. Fong, S.Z. Teo, Y. Li, and X. Yang, "Human Pacman: A Mobile Wide-Area Entertainment System Based on Physical, Social, and Ubiquitous Computing," *Personal and Ubiquitous Computing,* vol. 8, no. 2, pp. 71-81, May 2004.

[12] A.D. Cheok, X. Yang, Z. Zhou, M. Billinghurst, and H. Kato, "Touch-Space: Mixed Reality Game Space Based on Ubiquitous, Tangible, and Social Computing," *J. Personal and Ubiquitous Computing,* pp. 430-442, 2002.

[13] J.M. Hasenfratz, M. Lapierre, and F. Sillion, "A Real-Time System for Full Body Interaction," *Virtual Environments,* pp. 147-156, 2004.

[14] T. Horprasert et al., "A Statistical Approach for Robust Background Subtraction and Shadow Detection," *Proc. IEEE Int'l Conf. Computer Vision (ICCV '99) Frame Rate Workshop,* 1999.

[15] H. Kato, K. Tachibana, M. Tanabe, T. Nakajima, and Y. Fukuda, "Magiccup: A Tangible Interface for Virtual Objects Manipulation in Table-Top Augmented Reality," *Proc. Augmented Reality Toolkit Workshop (ART03),* pp. 85-86, 2003.

[16] K.N. Kutulakos and S.M. Seitz, "A Theory of Shape by Space Carving," *Int'l J. Computer Vision,* vol. 38, no. 3, 2000.

[17] S.P. Lee, F. Farbiz, and A.D. Cheok, "Touchy Internet: A Cybernetic System for Human-Pet Interaction through the Internet," *SIGGRAPH 2003, Sketches and Application,* 2003.

[18] T.W. Malone, "Toward a Theory of Intrinsically Motivating Instruction," vol. 5, pp. 333-369, 1981.

[19] R.L. Mandryk and K.M. Inkpen, "Supporting Free Play in Ubiquitous Computer Games," *Proc. Workshop Designing Ubiquitous Computer Games (UbiComp),* 2001.

[20] G. Markus, W. Stephan, N. Martin, L. Edouard, S. Christian, K. Andreas, K.M. Esther, S. Tomas, V.G. Luc, L. Silke, S. Kai, V.M. Andrew, and S. Oliver, "blue-c: A Spatially Immersive Display and 3D Video Portal for Telepresence," *Proc. ACM SIGGRAPH,* 2003.

[21] W. Matusik, C. Buehler, and L. McMillan, "Polyhedral Visual Hulls for Real-Time Rendering," *Proc. 12th Eurographics Workshop Rendering Techniques,* 2001.

[22] R. Mester, T. Aach, and L. Dmbgen, "Illumination-Invariant Change Detection Using Statistical Colinearity Criterion," *Proc. DAGM Symp. Pattern Recognition,* pp. 170-177, 2001.

[23] T. Oshima, K. Satoh, H. Yamamoto, and H. Tamura, "Ar2 Hockey System: A Collaboration Mixed Reality System," vol. 3, no. 2, pp. 55-60, *Trans. VRSJ,* 1998.

[24] S.J.D. Prince, A.D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato, "Live 3-Dimensional Content for Augmented Reality," *IEEE Trans. Multimedia,* submitted.

[25] S.J.D. Prince, A.D. Cheok, F. Farbiz, T. Williamson, N. Johnson, M. Billinghurst, and H. Kato, "3D Live: Real Time Captured Content for Mixed Reality," *Proc. Int'l Symp. Mixed and Augmented Reality (ISMAR),* 2002.

[26] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "Rtp: A Transport Protocol for Real-Time Applications," Internet Eng. Task Force, Audio-Video Transport Working Group, 1996.

[27] S. Seitz and C. Dyer, "Photorealistic Scene Reconstruction by Voxel Coloring," *Proc. IEEE Conf. Computer Vision and Pattern Recognition,* June 1997.

[28] M. Seki, H. Fujiwara, and K. Sumi, "A Robust Background Subtraction Method for Changing Background," *Proc. Fifth IEEE Int'l Workshop Applications of Computer Vision,* 2000.

[29] G. Slabaugh, W.B. Culbertson, T. Malzbender, and R. Schafer, "A Survey of Volumetric Scene Reconstruction Methods from Photographs," *Proc. Int'l Workshop Volume Graphics,* 2001.

[30] G. Slabaugh, R. Schafer, and M. Hans, "Image-Based Photo Hulls," *Proc. First Int'l Symp. 3D Data Processing, Visualization, and Transmission,* 2002.

[31] G.G. Slabaugh, W.B. Culbertson, T. Malzbender, M.R. Stevens, and R.W. Schafer, "Methods for Volumetric Reconstruction of Visual Scenes," *Int'l J. Computer Vision,* 2004.

[32] R. Szeliski, "Video Mosaics for Virtual Environments," *IEEE Computer Graphics and Applications,* Mar. 1996.

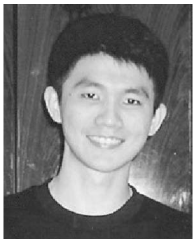
[33] H. Tamura, H. Yamamoto, and A. Katayama, "Mixed Reality: Future Dreams Seen at the Border between Real and Virtual Worlds," *IEEE Computer Graphics and Applications,* vol. 21, no. 6, pp. 64-70, 2001.

Ta Huynh Duy Nguyen is an MEng degree candidate (Research Scholar) under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. He received the BEng degree in information technology from the Department of Information Technology, HCM City University of Technology, Vietnam, in 2003. His research interests include augmented reality, human-computer interaction, 3D computer vision, and multimedia systems.

Tran Cong Thien Qui is an MEng degree candidate (Research Scholar) under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. His reearch interests include augmented reality, multimedia systems, artificial intelligence, and computer vision. He received the University Gold Medal in recognition of having the most outstanding academic performance among the Bachelor's degree graduates from the Department of Information Technology, HCM City University of Technology, Vietnam, in 2003.

Ke Xu is an MEng degree candidate (Research Scholar) under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. He received the BEng (Hons.) in electrical and computer engineering from the National University of Singapore in 2001. His research interests include augmented and virtual reality, human-computer interaction, and computer vision, especially exploring robust natural feature tracking techniques for wearable computers.

Adrian David Cheok is director of the Mixed Reality Lab National University of Singapore. He has worked in real-time systems, soft computing, and embedded computing at both Mitsubishi Electric Research Labs (Osaka, Japan) and the National University of Singapore (NUS). At NUS, he has been working on research covering mixed reality, human-computer interaction, wearable computers and smart spaces, fuzzy systems, embedded systems, power electronics, and multimodal recognition. He has successfully obtained funding for four externally funded projects in the area of wearable computers and mixed reality from the Defense Science Technology Agency (DTSA) Singapore. He is currently an assistant professor at the National University of Singapore, where he leads a team of more than 20 researchers and students. He has been a keynote and invited speaker at numerous international and local conferences and events. He has been invited to exhibit for two years in the Ars Electronica Museum of the Future, launching at the Ars Electronica Festival 2003. He is a member of the IEEE, was the IEEE Singapore Section Chairman in 2003, and is presently ACM SIGCHI Chapter President. He was awarded the Hitachi Fellowship 2003, the A-STAR Young Scientist of the Year Award 2003, and the SCS Singapore Young Professional of the Year Award 2004.



Sze Lee Teo received the BEng (Hons.) degree in electrical and computer engineering from the National University of Singapore in 2003. He is currently an MEng degree candidate (Research Scholar) under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. His main areas of research include software architectures and techniques for virtual reality and human computer interface design.

ZhiYing Zhou graduated from the Mixed Reality Lab, National University of Singapore led by Dr. Adrian Cheok. He is currently a research fellow of the lab. His research interests include augmented and virtual reality, human-computer interaction, and computer vision. He is a founding member of the ACM SIGCHI Singapore Chapter. His research work has been featured by the *New Scientist*, *Washington Post*, *Strait Times*, CNBC, BBC, etc. He was awarded the "Tan Kah Kee Young Inventor's Award 2004" and the second prize at the International Competition on Idea to Product. He was nominated to the "World Technology Award 2004" founded by CNN, Times, NASDAW, and Microsoft and was named a "Lifetime Fellow" due to his outstanding work in mixed reality research.

Asitha Mallawaarachchi is currently pursuing the BEng (electrical) degree in the Department of Electrical and Computer Engineering at the National University of Singapore. He is working on augmented reality applications for his undergraduate thesis. In 2003, he was awarded a summer undergraduate research fellowship by the California Institute of Technology.

Shang Ping Lee received the BEng (Hons.) degree in mechanical and production engineering from Nanyang Technological University and the MSc in mechatronics from the National University of Singapore (NUS). He is currently a research engineer under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering, NUS. His research projects are in the area of human power harvesting and personal area network.

Wei Liu received the BEng degree in electrical engineering from Xi'an Jiaotong University in 1994 and the MEng in electrical engineering from Xi'an Jiaotong University in 1997. She is a research engineer under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. Her research interests include augmented and virtual reality, human-computer interaction, and computer vision.

Hui Siang Teo received the BEng (Hons.) in electrical and computer engineering from the National University of Singapore in 2003. He is currently an MEng degree candidate (Research Scholar) under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. His research work involves development of augmented reality and virtual reality applications for both entertainment and use.

Le Nam Thang is an MEng degree candidate (Research Scholar) under the supervision of Dr. Adrian David Cheok in the Department of Electrical and Computer Engineering at the National University of Singapore. He received the BEng degree in 2003 from the Department of Information Technology at the Ho Chi Minh City University of Technology, Vietnam. His research interests include multimedia, computer vision, and human-computer interaction.

Yu Li is a technical support offices (TSO) at the Mixed Reality Lab at the National University of Singapore, where she is in charge of all of the art things in the group, including creating 3D models, designing all of the lab's logos, posters, and covers of conference proceedings, and constructing and maintaining the lab's official Web site. She has more than two years of experience in these areas and has obtained several official certifications in 3D modeling and graphics processing from Adobe and Autodesk. She is a master of 3D max, Adobe Photoshop, Flash, Dreamweaver, etc.

Hirokazu Kato received the BE, ME, and DrEng degrees from Osaka University, Japan, in 1986, 1988, and 1996, respectively. He was with the Department of Control Engineering at Osaka University from 1989 to 1999. In 1998, he joined the Human Interface Technology Laboratory (HIT Lab) at the University of Washington as a visiting scholar and worked for the wearable computing project. In 1999, he joined the Department of Information Machines and Interfaces at Hiroshima City University, Japan. Since 2003, he has been with the Graduate School of Engineering Science at Osaka University, where he is currently as associate professor. He has been studying pattern recognition, image processing, and computer vision. He has also been interested in human-computer interaction and computer mediated communication. Recently, he has been studying augmented reality technology and developed the ARTookKit, which is a world-famous software library to build augmented reality applications.