

# Interactive Modelling and Tracking for Mixed and Augmented Reality

R. Freeman  
University College London  
Gower Street  
London, WC1E 6BT  
+44 (0)20 7679 0352  
r.freeman@cs.ucl.ac.uk

A. Steed  
University College London  
Gower Street  
London, WC1E 6BT  
+44 (0)20 7679 4435  
a.steed@cs.ucl.ac.uk

## ABSTRACT

Some tasks vital to many mixed and augmented reality systems are either too time consuming or complex to be carried out whilst the system is active. 3D scene modelling and labelling are two such tasks commonly performed by skilled operators in an off-line initialisation phase. Because this phase sometimes needs specialist software and/or expertise it can be a considerable limiting factor for new mixed reality system developers.

If a mixed reality system is to operate in real-time, where artificial graphics are woven into real world live images, the way in which these off-line processes are tackled is critical. In this paper we propose a flexible new approach that reduces the time spent during the off-line initialisation phase by adopting an on-line interactive primitive modelling technique.

Our solution combines two existing and freely available packages, the Augmented Reality Toolkit Plus (ARToolKitPlus) and the Mixed Reality Toolkit (MRT), to enable rapid interactive modelling over live video using a freely moving camera. As a demonstration we show how these can be used to rapidly seed an object appearance-based tracking algorithm.

## Categories and Subject Descriptors

I.4.8 [Image Processing and Computer Vision]: Scene Analysis.

## General Terms

Algorithms, Measurement, Design, Experimentation.

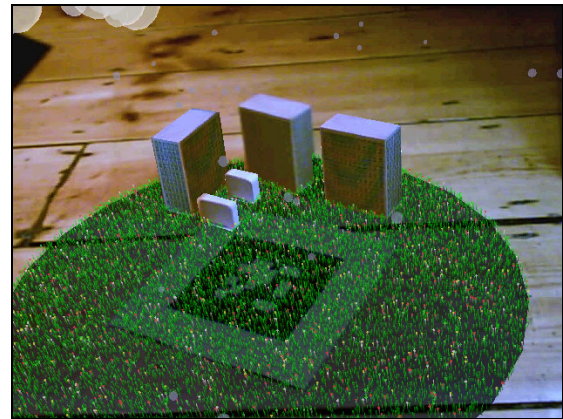
## Keywords

Mixed Reality, Augmented Reality, Image Based Modelling, Model Based Tracking.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

VRST'06, November 1–3, 2006, Limassol, Cyprus.

Copyright 2006 ACM 1-59593-321-2/06/0011...\$5.00.



**Figure 1. Real and virtual objects coexisting in live images from a mobile camera. Real world objects interactively modelled and cloned in real-time. Virtual grass augmented to scene and occluded by objects.**

## 1. INTRODUCTION

Whether presented on a Head Mounted Display (HMD) or tablet see-through PC, it is now possible to augment computer generated graphics over live images of the real world to create novel Mixed Realities (MR). The potential of MR is for increased usability and relevance to the operator, enabling both computers and users to operate together within a human context.

Knowing ‘*what*’ is present in the real world and ‘*where*’ it is, is vital for the operation of many MR systems [1]. Building up this real world knowledge, whilst also minimising user interaction and processing time, is a common hurdle for many MR systems. The significant time required for this task means that it is frequently performed prior to normal system operation in an off-line initialisation phase, with the assumption that the elements of the real world that are modelled will not change between initial capture and MR system deployment. We give some more detail to these specific MR modelling issues in the next section.

As an alternative approach, in section 3, we propose an on-line modelling solution that combines freely available marker-based tracking [4] with interactive primitive-based modelling [5] to enable interactive modelling in real-time using a freely moving camera. Section 3 also describes an object appearance-based tracking solution that can be initialised and calibrated as an on-line task using our proposed modelling solution.

In a demonstration of our solution we simultaneously used two modes of tracking to separately track scene objects and a camera. This meant that we needed to further consider how dynamic switching between and mutual support for these two modes might be best provided at run-time, and is discussed in more depth with other performance results and observations in section 4.

## 2. BACKGROUND

Knowing that a chair, table or particular building is nearby, for example, allows an MR system to appropriately target its visualised information to an operator's surroundings. Knowing also where things are and how they are orientated relative to one-another enables accurate placement and augmentation of the visualisations over real world images.

### 2.1 Modelling

A common hurdle for many MR systems is the building up of this real world knowledge whilst minimising any user burden and processing time. Although there are many existing techniques and commercial packages [6][7] that can be used to reconstruct and track an imaged scene, they often operate in an essentially 'stand-alone' fashion or do not provide a well defined scene model. The commercial applications are predominately desktop applications that are not appropriately designed to be integrated into bespoke MR systems. The dense point clouds produced by some of the more autonomous modelling techniques, whilst requiring very limited initial user intervention, will usually need to be analysed to find out to which object or component points should be attributed [15].

Depending on the specific application of an MR system, it is not always necessary to have a complete or overly detailed geometric scene model [1]. Only those aspects of the scene that are required by the MR systems mixing processes need to be considered. Not only can the more interactive modelling approaches provide an opportunity to define the 'what' that is being modelled but also consider only those parts of the scene that are relevant.

Recent work [3] has shown how interactive primitive-based modelling techniques, originally pioneered by Debevec [2], can be rapidly applied over live video images to interactively reconstruct, register and/or semantically define observed geometry whilst simultaneously extrinsically calibrating a camera. These simple interactive modelling techniques not only gave the operator the opportunity to identify and reconstruct each object, component, volume or location in an imaged scene, but also perform these tasks in an almost instantaneous manner over the live video images. However, this technique [3] was limited to using a camera held in a fixed position, requiring re-registration if the camera were moved. Integrating these model reconstruction techniques with camera position and orientation tracking techniques would be one way of overcoming this limitation.

### 2.2 Tracking

Using high contrast black and white markers placed into a scene the ARToolkitPlus [4] can quickly and automatically detect and process images of the scene to find the relative orientation and position of these markers. Although marker-based tracking can provide a very efficient and fast tracking solution, it is not always well suited to the needs of virtual object placement and augmentation. Because virtual content is augmented relative to the

located markers, if a marker is very much smaller in the imaged scene than the virtual content, or the virtual content is placed at a distance from the marker's centre, the registrations can become visibly unstable and jittery. To overcome some of these shortcomings other methods for both scene and individual object tracking can be used.

Simon, Fitzgibbon and Zisserman [16] proposed using flat surfaces already present in a scene. To increase the robustness they take advantage of the model structure by using the RANSAC algorithm [12]. Unfortunately these techniques suffer problems when trying to go from one surface to another as yet unseen surface patch. Jiang, You and Neumann [17] use both fiducial markers and natural features to overcome this problem.

Instead of planar objects, methods using 3D models of the scene where proposed by Gordon and Lowe [18], Lepetit, Vacchetti, Thalmann and Fua [9] and Okuma, Kurata and Sakaue [19]. Such techniques can provide accurate and locally stable tracking solutions for relevant scene objects. By generating representations of the features which are invariant to rotation and scale, whilst also providing methods for fast feature correlation [9] demonstrated a fast wide-baseline initialisation technique. These approaches have the added advantages of being able to track the internal rearrangement of the various scene objects.

### 2.3 Integrated Modelling and Tracking

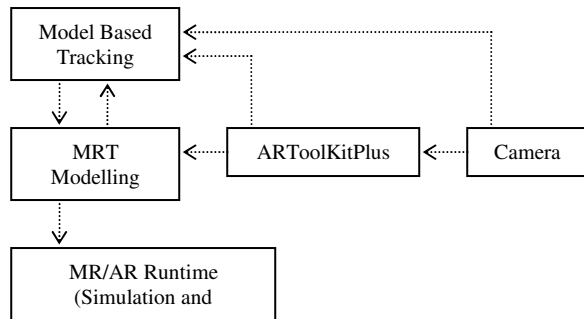
Using structure from motion techniques it is possible to both reconstruct and track a 3D model simultaneously. Because these techniques will produce an unsorted point-cloud or mesh model they lack the ability to describe the scene in terms of separable well defined objects suitable for use in many MR systems. The benefits gained by using these autonomous approaches are therefore somewhat mitigated by still having to collate and label the various reconstructed scene parts.

Using a hybrid of hardware tracking technologies along with hand controlled modelling capabilities; Piekarski [10] produced a mobile immersive MR modelling and tracking system known as TinMith. A user wears a see-through HMD device to see virtual content augmented to the real world. With vision tracked gloves for manipulation the user is able to interact with virtual objects beyond arms length, using a technique termed '*construction at a distance*'. He uses a simple shape-based modelling approach combined with 3D carving and '*bread crumb*' techniques to create new 3D objects. However, because the hybrid of hardware tracking technologies he uses are prone to inaccuracies and drift, the augmented content was often poorly aligned to the images of the real world. Using a completely vision-based tracking approach that compares the live captured images to what we expect to see is one way the results might be improved.

## 3. SOLUTION

To trade-off effort in an off-line phase by adding new actions in the on-line phase our proposed solution extends recent primitive modelling work [3] by combining tracking techniques to overcome its manual re-registration limitations.

When using live video with a moving camera the model and camera registrations must be continuously maintained as the camera moves through the scene. Using the ARToolKitPlus [4] marker-based tracking solution to continuously calibrate a ground



**Figure 2. Overall solution component diagram. Captured images initially processed using the ARToolkitPlus to find markers. Marker tracking information subsequently used to seed primitive modelling and model-based tracking.**

plane our proposed solution integrates the Mixed Reality Toolkit (MRT) [5] primitive-based modelling techniques to enable a user to interactively reconstruct and label geometry relative to a marker placed into a scene.

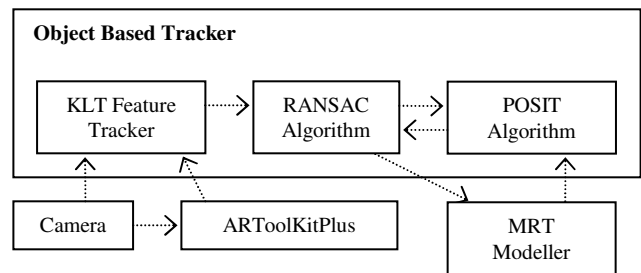
Prior to the MRT being applied over a captured image the marker must first be detected and the position and orientation deduced using the ARToolkitPlus (see Figure 2). Using this information the MRT can be applied to constrain new virtual modelling shapes to remain in a relative position and/or orientation to the marker. Using an interactive process, individual vertices of the virtual modelling shapes can be selected and manipulated to new positions in a 2D image by the user. The MRT can then be applied to estimate the new positions, orientations and dimensions of the virtual shapes relative to the marker.

As previously described, marker-based tracking offers very fast and efficient algorithms, but does have some inherent limitations. Using a single marker the position and orientation calculations are accurate for the area covered by the marker itself. However, the further away from the centre of the marker the worse the result becomes, exaggerating and amplifying any registration jitter.

### 3.1 Model Based Tracking

To overcome some of these shortcomings other more robust methods for both scene and individual object-based tracking need to be used. Given that we have captured a rudimentary model of objects in the scene an obvious solution was to use appearance-based object tracking. Using these techniques the movement of the camera might be tracked relative to the objects directly.

Initialised and calibrated in real-time, using both the primitive modelling tool (MRT) and marker tracking system (ARToolkitPlus), an object-based tracking solution similar to that of Lepetit, Vacchetti, Thalmann and Fua [9] can also be combined. Figure 2 shows how instead of simply calibrating the relative extrinsic camera parameters and inferred ground plane the marker tracking results can also be used to assist the object-based tracking processes, creating a hybrid of tracking techniques. Although stable and robust object tracking can be very accurately achieved for off-line sequences, performing this in real-time using hundreds of features is a difficult task. Any methods which give guidance to constrain the possible solution spaces for feature searching are therefore a good way of reducing the tracking processing workload. With this in mind, in our solution, the



**Figure 3. Internal component diagram for the object appearance-based tracker. Captured images are first processed to find marker positions before the KLT feature tracker is used to find updated feature positions. Using randomly selected minimal subsets of all located features, multiple pose estimations are made (POSIT) until a solution is found that best satisfies the overall data set (RANSAC).**

ARToolkitPlus was used to provide an initial suggestion for the feature location in an image.

Using a set of features identified when surface textures are captured from the live images, the objects pose can be estimated using a combination of the POSIT [13] and RANSAC [12] algorithms (see Figure 3). As with [9] the RANSAC algorithm is used to minimise the effect of erroneous and noisy feature correlations. Unlike [9] where the Harris corner detector [14] is used to identify good features to track, we opted for the freely available and open source Kanade-Lucas-Tomasi (KLT) feature tracker [8][11] to both identify and track features. The standard KLT feature tracker takes the current and previous image frames along with the last known 2D positions of a set of features as its tracking inputs. It was therefore necessary to make minor changes to allow for efficient use of a 3rd rendered virtual model ‘key-frame’ reprojection and also incorporate adjustments to the initial feature search positions provided by the ARToolkitPlus [4] when the markers are visible.

## 4. RESULTS

The above solution was implemented and run on a 3GHz PC with a 256MB 3D graphics card and a modern cheap web camera.

### 4.1 Pausing to Model

Because the camera was often being held and moved around by an unsteady hand, it would sometimes make it much harder for the operator to select and manipulate the correct vertices of the model. Therefore although it was possible to perform the modelling and tracking processes concurrently, to manipulate the virtual shapes with the MRT it was sometimes easier to ‘pause’ the image capture process and then continue once the manipulations can be completed.

### 4.2 Object Based Tracking

The assistance that the marker-based tracking provided (whilst visible) to the object-based tracking, seeding an initial feature search space from a new image, made it possible for the system to be used fairly robustly, with little regard to sudden or fast movements of the camera with minimum processor effort.

An important issue for many object-based tracking solutions is the need for initialisation. This task will often involve the capture and

processing of off-line key-frames to produce accurately textured models of the objects, along with a set of easily trackable features. Although it is possible, and reasonably efficient [9], to use some commercial applications for this task, the integrated approach used in the proposed solution can speed-up the initialisation and makes it a more flexible and reconfigurable process for the operator. Thus, by initially using the marker tracking it is possible to reconstruct geometry whilst using a movable camera to find various viewpoints to create a set of key-frames.

### 4.3 Intermittent Marker Visibility

Because the marker was not always visible in a captured image it was necessary to consider how object tracking could be used in its absence. The marker tracking results were normally used to define the camera's position and orientation in the real world. All the other positions and rotations of the remaining scene model are then calculated as a relative scene graph to this calibration. Therefore, during object tracking these relative positions are only updated with respect to the most recently captured and calibrated frame. If that frame does not contain a detectable marker (e.g. the marker was intentionally removed) then the marker's position information is eliminated from the solution and the positions of the objects are made relative to the camera instead. If a marker should then reappear at some later stage it is incorporated into the tracking solution in its new found location, the hybrid tracking can then continue as normal.

## 5. CONCLUSIONS

In this paper we have described how many existing MR systems spend a significant time modelling and describing an imaged scene as an off-line initialisation task. We have also demonstrated a possible solution that uses existing open source marker-based tracking to support primitive-based modelling techniques. The outcome is a lightweight and fast approach to scene modelling that might be well suited for use on modern handheld and mobile computer devices.

Using the KLT feature tracker [11] we were also able to demonstrate an individual object-based tracking solution. By both seeding the object-based tracking with marker tracking results and performing independent object tracking we demonstrated how an object tracker can be rapidly initialised and configured in real-time using our on-line modelling approach.

## 6. ACKNOWLEDGMENTS

This work was funded by the UK EPSRC Interdisciplinary Research Collaboration Equator (Grant GR/N15986/01).

## 7. REFERENCES

- [1] P. Milgram and A. F. Kishino. Taxonomy of Mixed Reality Visual Displays, *IEICE Transactions on Information and Systems*, E77-D (12), pages 1321-1329, 1994.
- [2] P. Debevec, C. Taylor and J. Malik. Modelling and Rendering Architecture from Photographs, *Proceedings of SIGGRAPH 96*, pages 11-20, 1996.
- [3] R. Freeman, A. Steed, B. Zhou. Rapid Scene Modelling, Registration and Specification for Mixed Reality Systems, *Proceedings of ACM Virtual Reality Software and Technology*, pages 147-150, 2005.
- [4] ARTToolkitPlus: [http://studierstube.icg.tu-graz.ac.at/handheld\\_ar/artoolkitplus.php](http://studierstube.icg.tu-graz.ac.at/handheld_ar/artoolkitplus.php), accessed on the 29<sup>th</sup> of April 2006.
- [5] Mixed Reality Toolkit: <http://www.cs.ucl.ac.uk/staff/rfreeman/>, accessed on the 29<sup>th</sup> of April 2006.
- [6] RealViz ImageModeler & MatchMover: <http://www.realviz.com/>, accessed on the 29<sup>th</sup> of April 2006.
- [7] The Pixel Farm PFBarn, PFTrack & PFMatch: <http://www.thepixelfarm.co.uk/>, accessed on the 29<sup>th</sup> of April 2006.
- [8] J. Shi and C. Tomasi. Good features to track. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'94)*, pages 593-600, 1994.
- [9] Vincent Lepetit, Luca Vacchetti, Daniel Thalmann and Pascal Fua. Fully Automated and Stable Registration for Augmented Reality Applications. *ISMAR*, pages 93-192, 2003.
- [10] Wayne Piekarski. Interactive 3d modelling in outdoor augmented reality worlds, *PhD Thesis University of South Australia*, 2004.
- [11] Kanade-Lucas-Tomasi Feature Tracker: <http://www.ces.clemson.edu/~stb/klt/>, accessed on the 29<sup>th</sup> of April 2006.
- [12] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Comm. of the ACM*, Volume 24, pages 381-395, 1981.
- [13] D.F. DeMenthon and L.S. Davis. Model-Based Object Pose in 25 Lines of Code. *International Journal of Computer Vision*, Volume 15, pages 123-141, 1995.
- [14] C. G. Harris and M. J. Stevens. A combined corner and edge detector. In *Proceedings of Alvey Vision Conference*, 1988.
- [15] H. Hoppe, T. DeRose, T. Duchamp, J. McDonald and W. Stuetzle. Surface reconstruction from unorganized points. In *Proceedings of ACM SIGGRAPH*, pages 71-78, 1992.
- [16] Simon, Fitzgibbon and Zisserman. Markerless Tracking using Planar Structures in the Scene. In *Proceedings of ISAR*, 2000.
- [17] Bolan Jiang, Suya You and Ulrich Neumann. Camera Tracking for Augmented Reality Media. *IEEE International Conference on Multimedia and Expo (III)*, pages 1637-1640, 2000.
- [18] Iryna Skrypnik and David G. Lowe. Scene Modelling, Recognition and Tracking with Invariant Image Features. In *Proceedings of ISMAR*, pages 110-119, 2004.
- [19] Okuma, Kurata and Sakaue. Fiducial-less 3-D Object Tracking in AR Systems Based on the Integration of Top-down and Bottom-up Approaches and Automatic Database Addition. In *Proceedings of ISMAR*, page 260, 2003.