



Research White Paper

WHP 145

January 2007

Mixed Reality Techniques for TV and their Application for On-Set and Pre- Visualization in Film Production

G. A. Thomas

BRITISH BROADCASTING CORPORATION

Mixed Reality Techniques for TV and their Application for On-Set and Pre-Visualization in Film Production

G. A. Thomas

Abstract

An overview of the technologies for the real-time mixing of real and virtual images for live TV production is presented. On-set visualization in film production has much in common with this application area; it aims to generate a view of the scene, including virtual elements, in real-time, to help the production crew in tasks such as shot framing. A short review is presented of how some techniques developed for TV been applied to film. A related area is pre-visualization, where computer graphics are used to simulate a view of a scene before any shooting takes place. Pre-visualization is already applied extensively in film production, but in TV production the traditional tools of paper-based storyboards and cardboard models of a set are still in common use. An example of a tool that provides an intuitive way of planning and visualizing a production is briefly described, which is applicable to both TV and film.

This paper was presented on 22nd October 2006 at the International Workshop on Mixed Reality Technology for Filmmaking, University of California at Santa Barbara, USA (in association with ISMAR 2006).

Additional key words: Virtual studio, augmented reality, TV production

White Papers are distributed freely on request.
Authorisation of the Head of Research is required for
publication.

© BBC 2007. All rights reserved. Except as provided below, no part of this document may be reproduced in any material form (including photocopying or storing it in any medium by electronic means) without the prior written permission of BBC Future Media & Technology except in accordance with the provisions of the (UK) Copyright, Designs and Patents Act 1988.

The BBC grants permission to individuals and organisations to make copies of the entire document (including this copyright notice) for their own internal use. No copies of this document may be published, distributed or made available to third parties whether by paper, electronic or other means without the BBC's prior written permission. Where necessary, third parties should be directed to the relevant page on BBC's website at <http://www.bbc.co.uk/rd/pubs/whp> for a copy of this document.

Mixed Reality Techniques for TV and their Application for On-Set and Pre-Visualization in Film Production

Graham Thomas

BBC Research

ABSTRACT

An overview of the technologies for the real-time mixing of real and virtual images for live TV production is presented. On-set visualization in film production has much in common with this application area; it aims to generate a view of the scene, including virtual elements, in real-time, to help the production crew in tasks such as shot framing. A short review is presented of how some techniques developed for TV been applied to film. A related area is pre-visualization, where computer graphics are used to simulate a view of a scene before any shooting takes place. Pre-visualization is already applied extensively in film production, but in TV production the traditional tools of paper-based storyboards and cardboard models of a set are still in common use. An example of a tool that provides an intuitive way of planning and visualizing a production is briefly described, which is applicable to both TV and film.

Additional Keywords: Virtual studio, augmented reality, TV production

1 INTRODUCTION

Graphics have been used in live TV programme production for many years, with chroma-key being used in applications such as placing a virtual weather map behind a presenter. However, it was not until the 1990s that graphics rendering hardware became powerful enough to render high-quality graphics at video frame rate. This allowed camera movement, as 3D virtual objects or backgrounds could then be re-rendered for each TV field (50 or 60 times each second), to match the current camera view. The freedom to move the camera allowed virtual objects or backgrounds to be used in a much wider range of programmes, and the term “virtual studio” was used to distinguish these new systems from simple chroma-keying techniques. Early virtual studio systems relied on high-end graphics supercomputers such as the SGI Onyx; an example of this was the ELSET system [1], demonstrated in 1994 at IBC in Amsterdam.

Since the early days of virtual studios, the technology has developed significantly. Camera tracking systems need no longer rely on mechanical sensors on the camera mounting, and graphics can be rendered on a conventional PC system. The way in which the technology is used has also developed: the initial enthusiasm for replacing the entirety of a real set with a virtual background has to some extent given way to the addition of virtual objects into a real environment, so that only those elements of the set that cannot be easily created for real are synthesized. Examples include virtual video walls and graphics for news, and overlaid

graphics for sports analysis. The technology has also enabled the production of innovative programmes where the virtual elements are the key part, such as the childrens’ game show BAMZOOKi [2].



Figure 1. The BBC programme BAMZOOKi, in which virtual creatures designed by children race in front of their eyes

In parallel with the developments in real-time augmented reality for TV, special effects using mixed real and virtual content for film have also developed apace. However, these effects are generated in post-production, in contrast to TV production where there is often a requirement for live use. This is generally a reflection of the vast differences in production budgets and timescales between film and TV, although with the reducing cost of 3D post-production, some TV programmes are starting to make significant use of mixed reality techniques in post-production.

One area in which both TV and film production can benefit from real-time mixing of real and virtual content is whenever there is significant interaction between the virtual elements and the presenters, actors or production crew. Framing a shot where a virtual element is a significant point-of-interest, or making eye-contact with a virtual actor, are very difficult unless there is some means of visualizing the virtual elements in real-time during filming. On-set visualization for film production is a technique that has been developed to address this problem, and the technology developed for virtual TV studios can play a key part in this.

Section two reviews some of the key technologies that are currently in use for real-time mixed-reality TV production. Section 3 presents some examples of how this technology is being applied to on-set visualization in film production. Section 4 discusses the requirements for pre-visualization for TV production, and shows an example of how these are currently being addressed.

Kingswood Warren, Tadworth, Surrey, KT20 6NP, UK
email: graham.thomas@rd.bbc.co.uk

2 TECHNOLOGIES FOR MIXED-REALITY TV

2.1 Camera Tracking

To insert a virtual object into a real image from a TV camera, the position of the object must appear correct as the camera moves. This requires the camera position, orientation and field-of-view to be measured at the same rate as the video signal (50 or 60Hz), and with a stability sufficient to ensure negligible visible drift between the real and virtual elements.

Specifying the measurement accuracy required for a convincing result is difficult, as it depends on numerous factors including the field-of-view of the lens, the resolution of the TV system (standard or high-definition), and the composition of the scene (whether virtual and real objects appear close together). For example, if the virtual object was to drift by no more than one TV line in a standard-definition image of around 500 lines vertical resolution, at a lens angle of 5° (a fairly tight zoom), an angular accuracy of around 0.01° is needed. If the virtual object in this scene was at a distance of around 6m, a camera movement of about 1mm would correspond to one picture line, indicating that the spatial position needs to be measured to an accuracy of about 1mm. Measurement inaccuracies may take various forms, such as random noise (which may have both low frequency and high frequency components, and may vary in amplitude depending on the speed of camera motion) and drift as a function of camera position. The influence of these inaccuracies will be different; for example an error that varies smoothly with camera position will generally be much less of a problem than a rapid noise-like variation that is present even when the camera is still.

Most systems in use for augmented reality TV production use sensors on the camera lens to measure the settings of zoom and focus, and the lens is pre-calibrated so that its focal length can be derived from these settings. As the lens zooms, the effective position of the principal point (the point at which an ideal pin-hole camera would be placed to get the same view) moves along the axis of the lens (sometimes by as much as 1m or more), and this should also be taken account of in the calibration process. The point in the image that is at the focus of expansion during a zoom is not generally exactly in the centre of the image, and this offset also needs to be measured and accounted for by the rendering software. Lens distortion will also vary with zoom, and can become significant (particularly at wide angles), even for high-quality zoom lenses. The distortion can also be pre-calibrated, although most real-time rendering systems do not currently allow images to be rendered with distortion.

In the early days of virtual studios, the only way of measuring camera movement was by fitting sensors to the camera mounting (e.g. to measure pan and tilt, plus motion along a track), or by using motion-control mounts that were already equipped with sensors. A range of different tracking systems that do not rely on mechanical mounts are now available, allowing the use of conventional camera mounts, including cranes and hand-held. Most systems are based on optical tracking technology, and can be classified as “outside in” or “inside out” systems.

An “outside-in” tracking system uses cameras mounted around the operating area, and a set of passive or active markers on the camera. These systems can be expensive to scale up to large

studios, as a large number of cameras may be needed. An example of such a system is Walkfinder [3].

An “inside out” tracking system uses images from either the TV camera itself, or an additional camera fixed to the camera being tracked. The camera views markers, patterns or other easily identifiable features in the studio, whose 3D coordinates are known. By analyzing the image, the pose of the camera can be computed. Such systems were first developed for head-mounted augmented reality using active beacons [4], but have been developed to be suitable for large TV studios. An example of such a system is the BBC-developed *free-d* system [5], which uses circular retro-reflective barcoded markers on the studio ceiling (Figure 2). A small upward-looking monochrome camera surrounded by a ring of LEDs is mounted on each studio camera to be tracked (Figure 3).



Figure 2. Circular barcoded markers mounted on the ceiling of a TV studio



Figure 3. *free-d* tracking camera mounted on studio TV camera.

Further discussion of other commercially-available tracking systems may be found in Wojdala [6]. These systems all rely on some form of sensors or special markers in the environment in

which the system is being used. A current topic of active research is the development of tracking systems that can use naturally-occurring features in the scene. One example of such work is the MATRIS project [7]. Results from this project have already been applied commercially in sports applications, where lines on a football or rugby pitch can be automatically identified and tracked.

There are applications where it is useful to be able to attach virtual graphics to small physical objects rather than fixing them in the world reference frame. This allows a presenter to pick up a virtual object, and can provide an interesting way to interact with such objects [8]. An example of the use of an image-based marker tracking system for TV production is shown in Figure 4. Similar systems have been used for hand-held augmented reality using a PC and web-cam [9], although for broadcast use the systems need to satisfy the high frame-rate and robustness requirements for TV. A potential problem with the use of such systems is that by the very nature of the close interaction between presenter and object, there is a risk of the presenter covering the marker, or placing a hand into the space that the virtual object should occupy. Careful rehearsal is usually required.

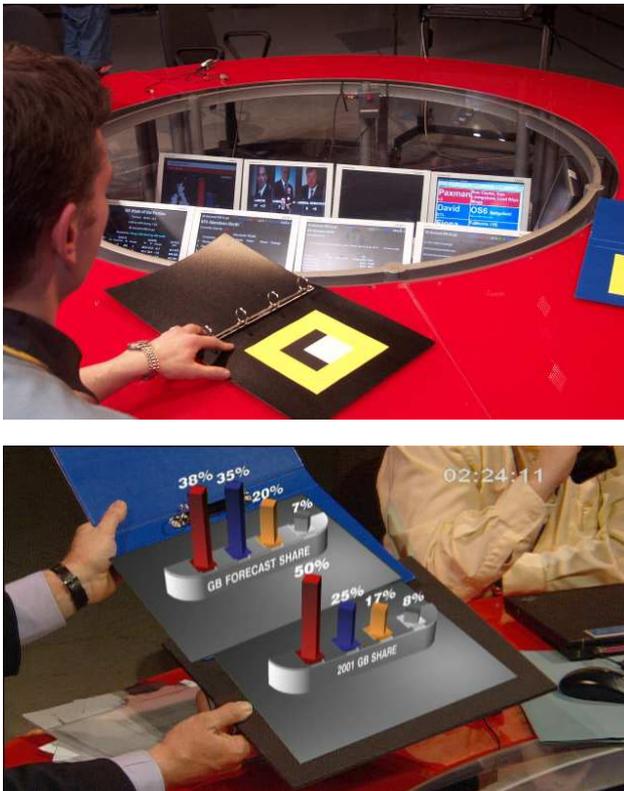


Figure 4. Vision-based tracking of rectangular markers (top) used to produce hand-held 3D graphics (bottom)

2.2 Graphics generation

There are a range of software packages available commercially for the real-time rendering of graphics for mixed reality / virtual studio production. Most require high-end PC systems, equipped with broadcast-standard video capture and output cards. The

features that differentiate such systems from related PC applications such as 3D games include:

- **Broadcast-standard video output:** this is usually achieved by using either a graphics card that supports serial digital video output, or a separate card that provides broadcast-standard output, in conjunction with software that copies the rendered image from the frame-buffer on the graphics card. An alternative is the use of a stand-alone scan converter connected to the video output of a conventional graphics card, although this can make genlock and frame-rate synchronization difficult (see below).
- **Genlock:** the video signal needs to be locked to the video signal from the camera.
- **Consistent and stable high frame rate:** the image must be re-rendered exactly once per field, otherwise the motion of virtual objects will not match the smooth motion of the real camera. This is in contrast to applications such as games, where any frame rate above 20-30 frames per second may be considered good, and the higher the rate, the better.
- **Alpha (or key) output:** an additional video signal is often required to indicate those parts of the graphics that should always appear in the foreground.
- **Anti-aliasing:** whilst this is important in applications such as games, it is more important when the rendered images are being combined with video, as any differences in quality between the real and virtual parts tends to shatter the illusion that both are really present in the same scene.

2.3 Keying

If the virtual objects are always to appear in front of the real scene, all that is required is a key signal (sometimes called a mask or alpha signal) from the graphics engine, which can be fed to a mixer or keyer to indicate which parts of the rendered image should be forced into the foreground. An alternative to this is to render the graphics against a coloured background, and use a chroma-keyer to key the camera video into the coloured areas.

If virtual objects are to appear behind real objects, a key signal for the real objects is required. Where the relevant part of the background is to be completely virtual, this is usually achieved using a coloured background and a chroma-keyer. An alternative to a conventional blue or green background is to use a retro-reflective cyclorama [16] in conjunction with a ring of coloured lights around the camera. This is particularly useful if low light levels are used, when a conventional coloured background may not appear bright enough. An example of the use of such a retro-reflective background was shown in Figure 3.

However, in situations where a virtual object needs to be inserted behind one real object but in front of another, and neither has a distinct colour, chroma-key cannot be used. Although a key signal could be generated in post-production by using "rotoscoping" or other (semi) manual techniques to draw a key for the foreground object, this approach clearly cannot be used for live production.

One approach that has been developed to allow the real-time generation of a key signal for an arbitrary foreground object relies on illuminating the foreground object with pulsating coloured light [10]. Real-time image processing is used to detect the areas

of the image having the appropriate phase and frequency of variation of the chosen colour and generate a key signal for these areas, whilst also generating a “clean” video signal with the flashing component removed. Figure 5 shows an example of this “FlashKey” technique, where a strobe light running at 75Hz illuminates the arm of a presenter. The camera captures images of the scene at 50Hz, but with a shutter time under $1/75^{\text{th}}$ of a second, so that the strobe light is only visible on alternate fields.



Figure 5. Use of the “FlashKey” technique: (a) shows studio set-up, with strobe light on left, illuminating the presenter’s arm, and (b) shows the composite image after processing.

Other alternatives to chroma-key include depth estimation, using either multiple cameras [11] or time-of-flight of infra-red light [12]. In situations where the background is stationary, an approach based on difference-keying may be used [13]. However, at the current state of development, none of these approaches are as robust or effective as chroma-key, and real-time operation is still a challenge.

2.4 Presenter Feedback

For any situation in which a presenter, actor or other participant in a programme need to interact with virtual content, it is important to provide some form of visualization for them. At simplest, this can be provided by positioning a TV monitor out-of-shot but visible to the presenter, showing the composite image. Sometimes, the auto-cue on the camera is used for this, often with the image reversed to simulate the behaviour of a mirror.

In order to allow the presenter to maintain the correct eye-line whilst viewing the virtual content, a view of the virtual object can be projected into the scene, at the position where the virtual object will appear. It is important that this projected image does not appear in the final composite image. If a chroma-key background is being used, this can usually be achieved by projecting a low-brightness image that does not interfere with the keying process, and ideally using back-projection so that the projected image does not fall on the presenter.

However, where the image needs to be projected onto real scene elements (for example, if the presenter needs to see virtual objects that are to be placed on a real table), a low-brightness projection may be visible to the camera, particularly if the rendered virtual objects do not exactly obscure the projected image. A method of rendering such a projected image invisible to the camera is to blank the projector synchronously with the camera integration period (for example, by reducing the shutter time on the camera so that it integrates for $1/100^{\text{th}}$ of a second at a field rate of 50Hz, and applying a similar but opposite-phase shutter on the projector). An early example of the use of this technique [14] used a specially-designed liquid crystal shutter on

the projector, but recent advances in DMD projectors, particularly those designed for time-sequential stereo [15], mean that off-the-shelf projectors can now achieve the same results. Figure 6 shows an example of projection feedback being used to show virtual graphics on a table top. The projector is positioned high above the table, projecting downwards.



Figure 6. Projection feedback in use on the BBC show *BAMZOOKi* (the monitors on the right show the composite image and the original camera image)

Even when a shuttered projection technique is used, it is advantageous to avoid projecting light directly onto a presenter, in order to avoid dazzling them. This can be achieved by generating a key signal for the presenter, using a camera located close to the projector, and using this to blank out the portion of the projected light that would otherwise fall on them. Figure 7 shows such a set-up: the camera is surrounded by a ring of blue LEDs as a retro-reflective cyclorama [16] is being used as a chromakey background. Figure 8 shows the cyclorama with the projected image. One advantage of using a retro-reflective cyclorama in this application is that it appears grey (rather than bright blue) to the presenter, making the projected image easier to see. The studio TV camera is also fitted with a blue LED ring, so that the cyclorama appears bright blue from its point of view. An example of a composite image obtained from this set-up is shown in Figure 9. Note how the presenter is able to see what he is pointing at.

The projection method described above works well for virtual objects that appear roughly coincident with surfaces (such as tables, walls and floors) onto which images could be projected. However, in situations where a virtual object (such as a virtual actor) is standing freely in space, this relatively straightforward approach will not work. For example, if a top-down view of a virtual actor was projected onto the point on the floor at which he

stands, the eye-line of the presenter will be wrong when looking at the head of the virtual character. This problem can be solved by using a view-dependant projection method [17].

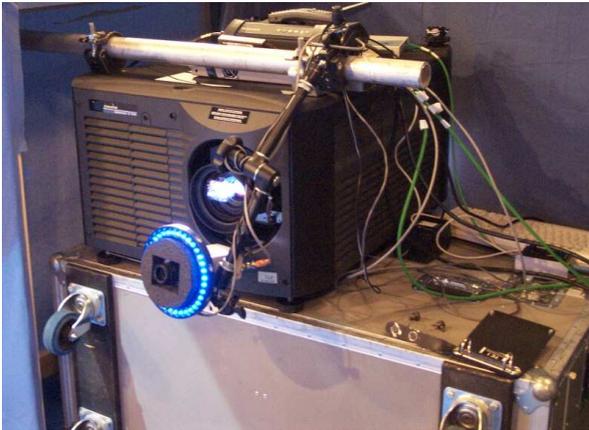


Figure 7. Projector with associated camera to generate key image to avoid projected light landing on presenter



Figure 8. Projected image, showing blanked portion for presenter



Figure 9. Composite image as broadcast, with presenter using projected image to view virtual graphics

3 APPLICATION TO ON-SET VISUALIZATION FOR FILM PRODUCTION

Many of the technologies described in the previous section can be applied during film production, in order to give the production crew a real-time view of roughly how the completed shot will look. Because the requirement is only to give a rough indication of the final shot, for purposes such as shot framing and checking of eye-lines, camera tracking generally does not need to be as accurate as for TV production, and the graphics can be rendered at a lower frame-rate. The quality of the rendered graphics need not be anything like as high as in the final film.

One of the first times that on-set visualization was used was in 2001, when ILM used the *free-d* system [5] during the production of the film *A.I. Artificial Intelligence*. This allowed Steven Spielberg to view the virtual environment and experiment with different camera angles.



Figure 10. *free-d* targets used for on-set visualization



Figure 11. Augmented view from camera viewfinder

At about the same time, *free-d* was also used by BBC Resources in providing on-set visualization for *Harry Potter and the Philosopher's Stone*. The signal from the "video assist" (a small TV camera seeing the view through the film camera

viewfinder) was overlaid with images rendered using the Maya animation package, driven by the live camera tracking data. Pre-generated animations could be triggered and viewed in real-time as the film camera moved.

4 PRE-VISUALIZATION FOR TV AND FILM PRODUCTION

Pre-visualization for film generally makes use of high-end 3D animation packages to create mock-ups of complex scenes in films.

For TV production, pre-visualization using graphics tools is still relatively uncommon, with manually-drawn storyboards and cardboard models of sets still being in common use. To be attractive to the production crew, a graphical tool needs to be easy to use, with an interface that works in familiar terms such as floorplans and pedestal-mounted cameras, and it must offer easy-to-use models, rather than having the bewildering range of functionality that high-end 3D animation packages have to offer.

An example of some early work to produce a tool specifically designed for pre-visualization is that of Higgins [18]. More recently, commercial products have started to appear that cater for this market. One such example is Antics [19], developed with advice from the BBC, and building on experience gained with earlier simulators used for training of TV directors. Some example screenshots from this program are shown in Figure 12.

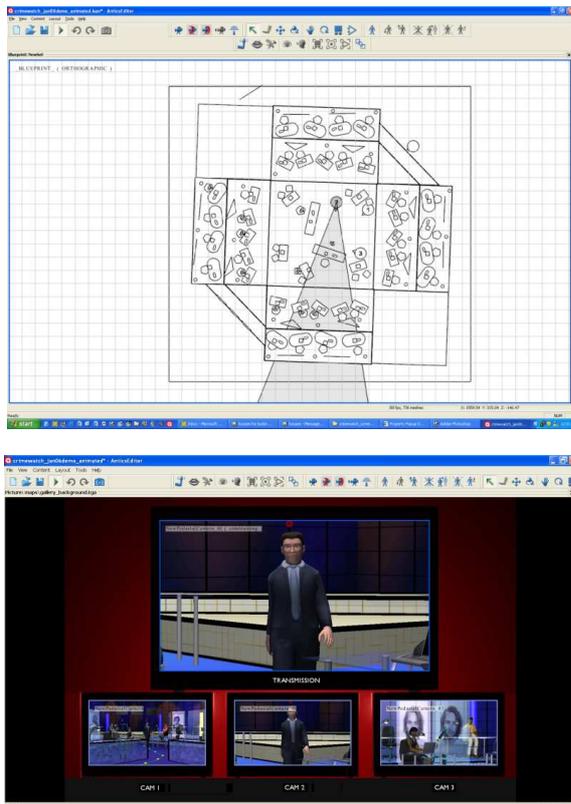


Figure 12. Screen-shots from a program designed specifically for pre-visualization for film and TV, showing a floorplan view (above) and a simulated production gallery (below)

5 CONCLUSION

This paper has reviewed some of the technology currently available for real-time mixed reality techniques for TV production, and showed how some of these have been successfully used to provide on-set visualization for films. The key thing about both application areas is the need for mixed real and virtual scenes to be generated in real time, during shooting.

Pre-visualization does not generally require mixed reality techniques, but is an important area where computer graphics can help with the production of both TV and film. Tools have recently been developed that make pre-visualization both easier, cheaper and quicker.

REFERNCES

- [1] Blondé, L., et al. A Virtual Studio for Live Broadcasting: the Mona Lisa Project. *IEEE Multimedia*, Vol. 3, No. 2, pp. 18-29, Summer 1996.
- [2] The BBC programme BAMZOOKi. www.bbc.co.uk/cbbc/bamzooki
- [3] <http://www.thoma.de/walkfinder1e.html>
- [4] Ward, M. et al. A Demonstrated Optical Tracker With Scalable Work Area for Head-Mounted Display Systems. *Proceedings of 1992 Symposium on Interactive 3D Graphics* (Cambridge, MA, 29 March - 1 April 1992), pp. 43-52.
- [5] Thomas, G.A., Jin, J., Niblett, T., Urquhart, C. A versatile camera position measurement system for virtual reality TV production. *IBC*, Amsterdam, September 1997. IEE Conference Publication No. 447, pp. 284-289.
- [6] Wojdala, A. Virtual Studio in 2000 the State of the Art. *Virtual Studios and Virtual Production Conference*, New York, 17-18 August 2000.
- [7] Chandaria, J. et al. Real-time camera tracking in the MATRIX project. *IBC*, Amsterdam, September 2006, pp. 321-328.
- [8] Lalioti, V., Woolard, A. Mixed Reality productions of the future. *IBC*, Amsterdam, September 2003, pp. 312-320.
- [9] Billinghurst, M., Kato, H., Poupyrev, I. The MagicBook: A Transitional AR Interface. *Computers and Graphics*, November 2001, pp. 745-753.
- [10] FlashKey. <http://www.bbc.co.uk/rd/projects/virtual/flash-keying>
- [11] Thomas, G.A., Koppetz, M., Grau, O. New methods of image capture to support advanced post-production. *IBC*, Amsterdam, September 2003, pp. 211-220
- [12] Iddan, G.J., Yahav, G. 3D Imaging in the studio (and elsewhere...) *Proc. SPIE, Conference Proc. of Videometrics and Optical Methods for 3D Shape Measurement*, pp. 48-55, January 2001.
- [13] Thongkamwitoon, T. et al. An Adaptive Real-Time Background Subtraction and Moving Shadows Detection. 2004 IEEE International Conference on Multimedia and Expo, pp. 1459-1462.
- [14] Fukaya, T. et al. An Effective Interaction Tool for Performance in the Virtual Studio - Invisible Light Projection System. *IBC*, Amsterdam, Sept. 2002, pp. 389-396.
- [15] Christie Mirage S+14K projector. <http://www.christiedigital.com/AMEN/Products/christieMirageS14K.htm>
- [16] Truematte retro-reflective chroma-key system. <http://www.bbc.co.uk/rd/projects/virtual/truematte>
- [17] Grau, O., Pullen, T., Thomas, G.A. A combined studio production system for 3D capturing of live action and immersive actor feedback. *IEEE Transactions on Circuits and Systems for Video Technology*, Volume 14, No. 3, March 2003.
- [18] Higgins, S. The Moviemaker's Workspace. <http://alumni.media.mit.edu/~scott/abstract.html>
- [19] Antics www.antics3d.com