Image Manipulation through Gestures

José Miguel Salles Dias
Miguel.Dias@iscte.pt

Pedro Nande
Pedro.Nande@iscte.pt

Pedro Santos
Pedro.Santos@iscte.pt

Nuno Barata
Nuno.Barata@iscte.pt

André Correia
Andre.Correia@iscte.pt

ADETTI/ISCTE, Associação para o Desenvolvimento das Telecomunicações e Técnicas
de Informática, Edifício ISCTE, 1600-082 Lisboa, Portugal, www.adetti.iscte.pt

Abstract

In this work we present a novel free-hand gesture user interface based on detecting the trajectory of fiducial markers attached to the user's fingers and pulse, able to interact a sequence of images of a digital video piece. The model adopted for the video representation, is based in its decomposition in a sequence of frames, or filmstrip. Totally sensor-less and cable-less interfaces, provide the means for a user to intuitively interact through gestures with the filmstrip, within the framework of an Augmented Virtuality usage scenario. By simply gesturing, users are able to select at random, drag, release, delete or zoom image frames, browse the filmstrip at a controlled user-defined rate and issue start, end, stop and play commands to better control the digital video sequence. A fixed video camera monitors the user interaction through gesturing of the mentioned fiducial markers. This scheme enables the system to simplify the more complex problem of marker-less free-hand gesture tracking. Once the markers are detected and recognized in real- time by the computer vision layer, the system obtains the 3D pose (position and orientation) of the marker centres in relation to a virtual camera reference frame, whose mathematical model matches the real video camera. We are specifically interested in obtaining the pose of the left and right hand pulses, left and right thumb, and left and right hand index. By projecting the positions of these poses in the 2D visualization window, simple topological analysis based in the study of the kinematics evolution of distances and angles, can be implemented, enabling gesture recognition and the activation of system functions and, subsequently, of specific gesture-based user interaction for a given active functionality. This interaction will affect the shape, scale factor, position and visualisation of scene objects, that is, of filmstrip frames. For the computer vision layer, our system adopts AR Toolkit, a C/Open GL-based open source library that uses accurate vision based tracking methods to determine the virtual camera pose information through the detection in real-time of fiducial markers. The graphical output is implemented with C++/Open GL. Our proposed system is general, in the sense that it can interact with any filmstrip, obtained "a priori" from a digital video source.

Keywords

Augmented Virtuality, Tangible Interfaces, AR Toolkit, Hand and Finger Gesture, Digital Video Editing, Image Browsing, Filmstrip.

1. INTRODUCTION[1]

With our work, we are aiming to provide innovative Mixed Reality technology, applicable for user interaction with images of a digital video sequence. The model adopted for the video representation, is based in its decomposition in a sequence of frames, or filmstrip. Totally sensor-less and cable-less interfaces, used in previous works as Tangible Interfaces [Dias 03a], but adopted here as simple fiducial markers for the underlying vision

tracking system, provide the means for a user to intuitively interact through gestures with the filmstrip, within the framework of an Augmented Virtuality usage scenario. The augmentation of the virtual environment introduced by our system, were image frames from the filmstrip are mapped into a 3D model of a projection screen, enables the user to concentrate in the task at hand and does not add a cognitive seam in the visualisation process. In video or image editing and visualization scenarios, creating a technological environment where a user could experience intuitive gesture-based interaction with a filmstrip, could be something very rewarding and revolutionary for these users.

---

[1] The test images and filmstrip here presented are part of the Official Trailer of the Minority Report Motion Picture © by DreamWorks available in their corporate site.

This work was in part motivated by Spielberg's "Minority Report" gestural and visual-based HCI language. The paper focuses in describing the basic techniques and in showing the potential of the concept of image manipulation by means of gesturing, and should be seen as an illustration of a technological platform, whose validity evaluation in a given usage context, still needs to be accessed by a proper usability evaluation methodology, such as in [Dias 03b], currently on-going and to be published in the near future. Although the test application described here is related to image editing, the purpose of the paper is to deploy a technique that could be used also in other application scenarios, such as large image database browsing. The system uses hand gesture detection as an HCI[2] modality, by adopting a trajectory approach and a marker-based technique. This marker-based scheme can be seen as a transitional technology for this type of gestural HCI modality systems, since it simplifies the more complex and general problem of marker-less free-hand gesture trajectory tracking.

In synthesis, in this paper, we address the appropriateness and technical viability of marker-based free-hand gesture user interfaces in Augmented Virtuality, for the enhancement of visualisation and interaction with digital images in video sequences.

The paper is organised as follows: in section 2, we provide a background and state-of-the-art in the issues of Augmented Reality, Mixed Reality and Tangible Interfaces. In section 3, we present our modular system functions and detail the hardware and software of the developed prototype platforms. Section 4 covers the issues of the User Interfaces adopted in our system. Finally, in section 5, conclusions and future directions of research are given.

2. BACKGROUND

Augmented Reality (or AR) systems and technologies were introduced in 1992 by Caudel and Mizell [Caudel92], in the context of a pilot project, where they were used to simplify an industrial manufacturing process in a Boeing airplane factory. In general these systems provide the means for "intuitive information presentation, which enhances the perceiver's situational awareness and cognitive perception of the real world" [Behringer 99]. This enhancement is achieved by placing virtual objects or information cues into the real world, which is made possible by performing "virtual camera calibration", that is, by computing the virtual camera parameters that match the position and orientation of the observer of the real scene. With this technique, "virtual" objects can then be registered in relation to "real" objects which means that these objects can be seen in the same position and orientation of other "real" objects of the scene, as perceived by the user. This is usually done using optical or video see-through head mounted displays and tracking devices, linked to either standalone computers with 3D graphics capabilities, or mobile wearable computers. Video see-through AR is where virtual images are over-

laid on live video of the real world. A possible alternative is optical see-through AR, where computer graphics are overlaid directly on a view of the real world.

According to Azuma [Azuma 97], AR "1) combines real and virtual environments; 2) is interactive in real-time; 3) is registered in 3D. We can conclude that AR is a challenging multidisciplinary field and a particularly promising new user interface paradigm, integrating computer graphics, computer vision, positional and orientation tracking technologies, wearable computer architectures, psychology, artificial intelligence, ergonomics and others. In view of this integration framework and associated complexity, Milgram and Herman [Milgram 99], have proposed a new taxonomy that supplements the definition of Augmented Reality, by introducing the concepts of Augmented Virtuality (AV) and Mixed Reality (MR). They argue that Real Environments (RE) and Virtual Environments (VE) are, in fact, two poles of the Reality-Virtuality Continuum, being RE the left pole and VE, the right pole. REs include sample data acquired and digitised from real scenarios, such as image, video, ultrasound, X-ray, laser range and light intensity data and others, whereas VEs, refer to computer models of 3D scenarios that can be rendered in real-time. AV includes 3D scenarios that are supplemented by the inclusion of sample data acquired and digitised from real scenarios, through for example, texture mapping techniques. Mixed Reality is thus the overall framework that includes the continuum transitions from RE, to AR, passing through AV and towards VE.

On the user interaction side of the equation within immersive Mixed Reality environments, standard input devices such as keyboards or mice are useless, since they distract the user from the task at hand, thus creating a severe cognitive seam within the interaction process. On the other hand, traditional input devices used in virtual environments, such as the data glove or a 3D mouse with 6 degrees of freedom (6 DOF), introduce undesirable complexity in the tracking technology, or add "strange" gadgets to the user's workspace, with whom she's not daily accustomed. To face this problem, Kato [Kato 01a] have proposed Tangible Interfaces for Augmented Reality, as a new approach for the design of interactions in that context. According to the author, Tangible Interfaces, "are those in which 1) each virtual object is registered to a (tangible) physical object; and 2) the user interacts with virtual objects by manipulating the corresponding tangible object". Tangible Interfaces are described in the literature as being intuitive, since physical object manipulations are mapped to virtual object operations. If the mapping is one-to-one, we classify the interface as a space-multiplexed one. A time-multiplexed interface is the one where a physical object may correspond to different functions, at different points in time.

The related issue of hand and body gestures recognition have been tackled by using two different approaches: an image reference frame-based history approach [Dias 02a], where the bitmaps or pixmaps of the gesture path in the image reference frame are maintained and managed

---

[2] HCI – Human Computer Interaction

to derive gestures, or a world reference frame-based trajectory approach [Kopp 03], where the vision system tracks the evolution of the 6 degrees of freedom pose of the hand (its kinematics law) in the world reference frame, to compute the gestures.

Since 1996, AR and its applications, have also been researched by a group of FhG-IGD[3], in areas like repair and maintenance of complex facilities, interior and exterior architecture, as well as in medical scenarios, such as needle biopsy and minimal invasive treatment, combining Computed Tomography (CT), Magnetic Resonance (MR) and Ultrasound (U) imaging, with the real scene, enabling news means of interaction of the physician with the patient [Reiners 98]. More recently, ARVIKA [Friedrich02], a large initiative funded by the German Ministry of Education and Research, is developing Augmented Reality technology and applications in the areas of product development, production and service in the automotive and aerospace industries, for power and processing plants and for machine tools and production machinery. This initiative includes some of the worlds largest European and global players in the automotive and aerospace industries. ARVIKA efforts are focused in mobile use of AR in industrial applications, involving the fusion of real and virtual information, the delivery of such information as a function of context and multi-modal input/output, including speech, Tangible Interfaces and hands-free operations. The general problem of satisfactory user tracking and object registration is being investigated, using both fiducial markers on the environment and marker-less tracking.
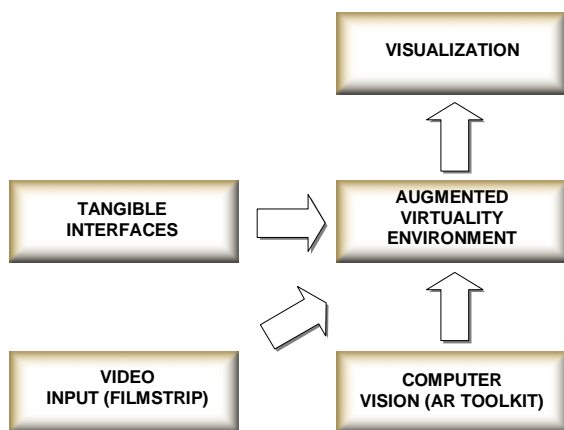

Figure 1 System Architecture.

3. SYSTEM ARCHITECTURE

To deploy our new user interaction paradigm for image frames of digital video, a number of modules comprise our proposed system, referred to as FilmstripWorks (**Figure 1**).

The **Computer Vision** Module includes a lower layer of Video Capture, which feeds into **AR Toolkit** (the Computer Vision and Image Processing layer [Kato 01b]),

---

[3] Fraunhofer-Institute for Computer Graphics, www.igd.fhg.de

which is able to recognize and track, by computer vision means and in real-time, the 3D pose (position and orientation) of fiducial markers centres, relatively to the virtual camera reference frame, whose mathematical model matches the real video camera capturing live images. In order to provide the required user interaction, a dedicated system module was developed, which comprises real-time **Tangible Interface** Gesture Recognition and Virtual Object Transformations (picking, translation and scaling of image frames). Our system requires the avail-

Figure 2 FilmstripWorks interaction window (top), with multiple image frames that were picked, dragged and zoomed. Detail showing frame numbers and the corresponding time stamps of the filmstrip (bottom).

ability at program start-up of a **Filmstrip**, produced by a third party system. The system then maps the filmstrip band in the lower layer of the visualization window (see **Figure 2**). The upper zone of this window, corresponds to the visualization and interaction area, where the user can edit (pick, drag, drop, delete, zoom, play the video sequence starting from a frame, browse the filmstrip) various image frames (our scene objects), selected from the filmstrip, using our technique of marker–based free-hand gesture detection. The lower zone or the window, depicts the frame number and the time stamp of each frame of the filmstrip.

4. SYSTEM CONFIGURATION

The typical hardware and software platforms required by our system, are as follows:

- Hardware - CPU: HP Pentium IV 2.5 GHZ; 1 G byte RAM; Graphics card: NVIDIA GFORCE4 128 Mbytes; Webcam (input video with 640x480 pixel);

- Software - MS Visual C++ 6.0 enterprise edition; Ar-Toolkit 2.5.2; Video Input - Direct X 8.1; Graphics - Open GL and Open VRML;

- Indoor Tracking method - Computer vision-based ap-proach (ArToolkit) with sparsely placed fiducial markers.

Figure 3: *Magic Rings* and *Magic Bracelets* enable the manipulation, through hand and finger gestures, of a filmstrip.

5. FREE-HAND AND FINGER GESTURING

Users generally require real-time user interaction in Mixed Reality (and more specifically, in Augmented Virtuality, as in our case). Having this in mind, we have developed Tangible Interfaces as intuitive as possible and we have come up a series of tools, similar to the ones found in [Billinghurst 01] [Kawashima 01] or [Dias 02b], which are suitable for our user interface requirements. One of these tools is referred as Magic Ring (see Figure 3) and the other, the Magic Bracelet, which have a specific marker attached. As a visual aid, when one of these markers is recognized by the system, an identification text will be displayed on top of it. Instead of registering virtual objects to real ones as in the mentioned works, in the present work we add novel uses for these interfaces, since they are the only physical means of user interaction that will trigger the functionalities of the system. This is achieved by detecting the trajectories of the centres of the fiducial markers attached to the Tangible Interfaces (as provided by ArToolkit), thus deriving the free-hand and finger gestures of the user.

Both the activation of system functions and the tracking of the markers, are performed in the 2D visualization window domain, through simple topological analysis between the projection, in the same window, of the visible markers centre vectors. This analysis consists in evaluation of simple mathematical relations between the mentioned centre vectors. The current implementation assumes a viewport of 1024x768 pixel that occupies the full screen, with a coordinate system that has its origin at the left upper corner, were the *yy´* axis points downwards. The input video source window of 640x480 pixel is mapped into this viewport.

Given vectors **a** and **b**, corresponding to say, the projected centres in the visualisation window of two fiducial markers, we can easily compute, respectively, their Cartesian distance, *dist*(**a**, **b**), the minor angle between them, (**a** ^ **b**), and their mid point, **P** ≡ (P_x, P_y), through:

$$dist\,(\mathbf{a},\mathbf{b}) = \sqrt{\left(\mathbf{a_x} - \mathbf{b_x}\right)^2 + \left(\mathbf{a_y} - \mathbf{b_y}\right)^2}$$

$$(\mathbf{a}^\wedge\mathbf{b}) = \arctan\left(\frac{\mathbf{a_x} - \mathbf{b_x}}{\mathbf{a_y} - \mathbf{b_y}}\right)$$

$$P_x = \min\left(\mathbf{a_x},\mathbf{b_x}\right) + \left|\frac{\left(\mathbf{a_x} - \mathbf{b_x}\right)}{2}\right|$$

$$P_y = \min\left(\mathbf{a_y},\mathbf{b_y}\right) + \left|\frac{\left(\mathbf{a_y} - \mathbf{b_y}\right)}{2}\right|$$

5.1 Activation of System Functions[4]

The activation of a given system function is highly dependent in the degree of proximity between two specified markers associated to the Tangible Interfaces (Magic Ring and Magic Bracelet). This degree of proximity is approximated by the evaluation of *dist*(**a**, **b**), where **a** and **b** are the projected visible marker centres in the visualisation window, such that the markers are not mutually occluded within a certain tolerance (which is dependent on the window dimensions). Each projected marker centre will in fact behave as if it was a cursor controlled by a specific finger gesture. Since the user has a total of six markers (A, B, C, D, F and |), he will control a total of six cursors in the visualisation window. In **Table 1** we list the available system functions and the correspondent gestures that trigger them.

---

[4] The + operator is, in this context, overloaded to a physical operation of "proximity" between the marker centres positions.

System Functions and Combination of Markers	Activation Gesture
1: Pick a frame or Drop a frame: F + D	
2: Resize a frame (zoom): (B + F) + (A + D)	
Stop resizing a frame: A + B	
3: Delete a frame from the main window: A + D	
4: Start or Stop Browsing through the filmstrip: D + C	
5: Start playing the filmstrip: A + |	
6: Stop playing the filmstrip: B + F	
7: Jump to the Start of the filmstrip: F + C	
8: Jump to the End of the filmstrip: B + |	

Table 1 Different gestures that trigger system functions.

5.2 Tracking of Gestures

Once a specific system function is activated, such as enabling image browsing in the filmstrip, we need to track the kinematics of the projected markers centres in the visualization window, so that we can map these kinematics laws into parameters of transformation matrixes, which are then applied to objects we intend to manipulate (image frames of the filmstrip).

5.3 Picking and Translation

Picking (or selecting) and translating (or dragging) objects (image frames) are dependent on one or two visible markers. These gestures have an analogy with the action of pickling and grabbing an object with the right index and thumb fingers (**Table 1.1**). The picking action itself is performed in Open GL, with the middle point of two markers centres, which must be inside the picked image frame. The translation option is executed by tracking the trajectory of the middle point between markers centres, or, in the case one of them is not visible, of the marker centre of the one that is visible.

5.4 Image Zooming

The gesture defined for this feature was largely influenced by the practice in cinema and photography of framing a scene, by the definition of the corners of the frame with both hands (**Table 1.2**). By approximating or separating these corners, a previously selected filmstrip frame is reduced or amplified. The scale factor applied to the filmstrip frame, is a function of the variable distance between two characteristic points. The efficiency of this function depends on the number of visible markers. It is activated with the necessary presence of four markers in the screen, but the tracking of the kinematics change of the zooming parameters, does not depend on the visibility of all four. The two characteristic points, **Point 1** and **Point 2,** are evaluated in the following way:

1. With four visible markers: **Point 1** - The middle point between the left index and the left thumb fingers projected marker centres; **Point 2** - The middle point between the right index and the right thumb projected marker centres;

2. With three visible markers: **Point 1** - The middle point between the index and thumb fingers projected visible marker centres, of one hand; **Point 2** - The projected visible marker centre of the other hand;

3. With two visible markers: **Point 1** and **Point 2** - The projected visible marker centre of each hand.

5.5 Image Browsing

With this functionality, the user is able to browse through the filmstrip, at a variable rate of frames per second. The gesture that activates this feature, just approximates the right thumb to the left hand pulse (**Table 1.4**). In the visualization window area, centred on the left pulse marker centre, a "browsing" graphical interaction pops-up, as depicted in **Figure 5**. It's a kind of a circular section with three holes that match the visible marker centres of the left index and thumb fingers and of the left pulse. This graphical interaction widget, is able to rotate relative to its centre, in the clockwise or counter-

clockwise directions. With this interaction a user is able to browse the filmstrip frame by frame or, to play the filmstrip forwards or backwards, at an interactively defined rate, much in the same manner as with a Jog-Shuttle wheel of a non-linear video editor. To support this specification, we need to identify two vectors, from the visible marker centres, **a** and **b**, in visualization window space, compute the difference vector, **b** – **a**, evaluate the angle between this vector and the *yy´* axis, and study the kinematics change of this angle through time. This technique will enables us to compute both the browsing direction as well as the browsing rate. After being activated, the instantaneous change of the parameters of this system function, re-quires the visibility of two specific markers, but its kinematics change in time, does not require the visibility of none of them, as long as there is no need to change the browsing rate and direction or issue another system function. The algorithm goes like this:

- Let **a** e **b** be the projected marker centres in study.

- Let *yy´* be the ordinate axis.

- Let $angle = \arctan\left(\dfrac{\mathbf{b}_x - \mathbf{a}_x}{\mathbf{b}_y - \mathbf{a}_y}\right)$

- Store the sign of *angle*

- Let $angle = \lVert angle \rVert$

Once *angle* is computed, it is compared against a series of predefined limit angles. From this comparison, different browsing rates will be computed and applied to the filmstrip (a larger angle will trigger a larger browsing rate, a smaller angle can just operate a frame-by-frame browsing). The browsing direction is given by the sign of *angle* (evaluated before the modulus operation): a positive and negative signs correspond, respectively, to forward (clockwise) and backward (counter-clockwise) browsing.
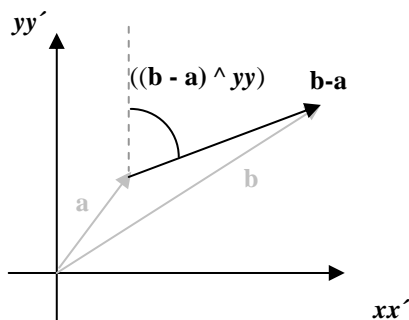
Figure 4 Determining the browsing direction and the browsing rate.

6. CONCLUSIONS AND FUTURE DIRECTIONS

This system presented in this paper describes a general and basic Augmented Virtuality platform that can be ap-

plied in a technical viable way, to interactive manipulation of digital images obtained from a video sequence, previously decomposed in a sequence of frames, or filmstrip. The modular system functions were presented and
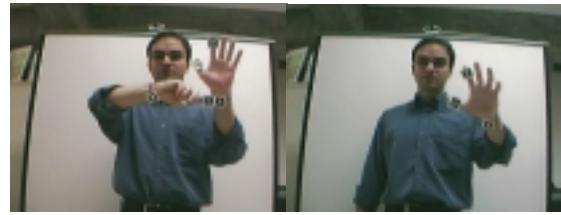
(a)   (b)
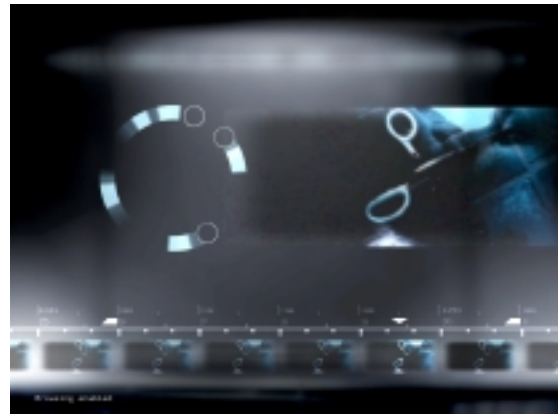
(c)

Figure 5: Browsing the filmstrip: (a) activation gesture; (b) browsing gesture; (c) visualization result**.**

several issues relevant for this type of Mixed Reality applications were raised, namely, the system architecture, the gesturing of Tangible Interfaces and the system functions. We have chosen the use case of video editing as a "proof of concept" to show that, given the recent advances in vision tracking algorithms, namely, the deployment of the AR Toolkit library and the availability of faster video cameras, CPU's and 3D graphic boards, marker-based free-hand gesture user interfaces, are technically viable for image manipulation. This does not contradict the fact that current desktop-based systems for video editing with 2D mouse interaction performs perfectly adequately and, in fact, we don't claim that our system will eventually replace the former. By conducting structured usability evaluation with a panel of users, on this type of usage scenario, which is part of our future plans, we are hoping of getting valuable feedback on the issue of the applicability of new gesture-based HCI modality for this type of task, and from there, we can start studying the more general applicability of this technique to other fields.

As a generalisation of our work, we have demonstrated the feasibility of using marker-based tracking, in the framework of AR Toolkit, to simplify the task of hand gesture recognition using a trajectory approach. With our technique we project the 3D coordinates of the visible marker centres to 2D coordinates of a given visualization

window, enabling the use of two Magic Rings and one Magic Bracelet in each hand, to control six new cursors projected on a given visualization window of an AV environment. By simple topological analysis between visible markers centre vectors distances, mid points and angles, we enable gesture recognition which can trigger functions of the system and, subsequently, of specific gesture-based user interaction to operate a given active functionality. This interaction affects the selection, deselection, shape, scale factor, position and visualization of scene objects (filmstrip frames in our case), enabling also filmstrip browsing at a user-defined rate. The gestures that activate and operate system functions, were developed as a compromise between the ease of usage and the limitations of the marker-based approach. Since the system functions and operations are triggered by the relative proximity and orientation of the visible marker centres, special care was given to the fact that the number of markers (six in total), could provide a sufficient number of possible proximity/orientation combinations for the definitions of HCI gestures, that could be conceptually close to the task at hand. When creating these gestures the authors have not yet considered the possible adoption of one the available hand gesture languages used by deaf persons, but this could be a promising line of research for the future. There are currently available research tools for automatic sign language recognition and synthesis (e.g, for the American Sign Language and British Sign language), which include formal methods (such as using the XML markup language MURML) to describe coordination of verbal and gestural constituents in deictic utterances [Wachsmuth 03]. MURML could be applied in our technique, as a standard way to codify the recognized gestures, which would enable a convenient integration with other modalities, such as speech recognition. With the support of MURML, our system could also be used as a platform to recognize, describe and translate different sign languages.

As another result, our gesturing-based control of cursors (just like mouse cursors), opens the possibility to provide user interaction with custom made "augmented" menus, including menu manipulation, menu items browsing and selection [Dias 03a].

Several other future directions can be envisaged for this kind of system, towards the enhancement of the new paradigm of Tangible Augmented Virtuality, for Interactive Visualisation and Editing of Digital Imaging. Concerning the system architecture, enhancements could be made in sense of immersion and the cooperative support of the system, to increase the perception of Augmented Virtuality, that is, of seamless fusion of virtual environments with real-live video and image input, both in standalone and collaborative working scenarios. In fact, as a future evolution, we could support, multiple collaborators (such as in [Dias 02c]), where collaborative image interaction working sessions could take place amongst local and remote users, immersed on different collaborative environment modalities over the Internet 2 backbone. As concluded by preliminary technical tests to the system, we need to adjust the camera tracking methods,

given various indoor lighting conditions. Hybrid indoor virtual camera tracking could also be developed, using a combination of vision-based approaches (both marker-based, as in AR Toolkit, and marker-less), ultra-sound sensors positioned along the workspace area, and accelerometers located on a pair of gloves or on a wearable unit dressed by the user. This approach could be a possible solution for the accurate registration problem, enabling, for example, the user to walk freely within a certain working area, while experiencing augmented virtual worlds.

Related use cases of rich image manipulation, as with gestural-based HCI modality in medical diagnosis applications, such as X Ray image analysis, or large image database browsing, could be well applied by our method, in the direction of the potential generalisation of our system to other fields.

7. ACKNOWLEDGEMENTS

8. REFERENCES

[Azuma 97] Azuma, R. T., "A survey of augmented reality", in Presence: Teleoperators and Virtual Environments, 6, 355-385, 1997

[Behringer 99] Behringer, R., et al, editors, Augmented Reality, Placing Artificial Objects in Real Scenes, Pro-ceedings of 1st International workshop on Augmented Reality - 1998, A. K. Peters ed., pp xi-xx, 1999.

[Billinghurst 01] Billinghurst, M., Kato, H., Poupyrev, I., "The MagicBook: A Transitional AR Interface", in Augmented Reality: the Interface is Everywhere, SIGGRAPH 2001 Course Notes 27, 2001

[Caudel 92] Caudel, T., Mizell, D., "Augmented Reality: An Application of Heads-up Display Technology to Manual Manufacturing Processes", Proc. Hawaaii International Conference on Systems Sciences, Maui, Hawaii, IEEE press, pp 659-669, January 1992.

[Dias 02a] Dias, J.M.S., Santos, P., et al, "MIXDesign, Tangible Mixed Reality for Architectural Design", SIACG 2002, Universidade do Minho, Portugal, June 2002

[Dias 02b] Dias, J.M.S., Santos, P., et al, "Tangible Interac-tion for Conceptual Architectural Design", "ART02, First IEEE International Augmented Reality Toolkit Workshop", ISMAR 2002, IGD, Darmstad, Germany, September 2002

[Dias 02c] Dias, Cooperative Work Support And Automatic Design Verification In Architecture And Building Construction, eSM@RT 2002 , University of Salford, United Kingdom, 19 - 21 November 2002.

[Dias 03a] Dias, J.M.S. Santos, P., Bastos, R., "Gesturing with Tangible Interfaces for Mixed Reality", Gesture Workshoop, Genova, April 2003

[Dias 03b] Dias, J.M.S. Jorge, J., Carvalho, J., Santos, P., Luzio, J., "Usability Evaluation of Tangible User Interfaces for Augmented Reality, submitted to "ART03, Second IEEE International Augmented Reality Toolkit Workshop", ISMAR 2003, Japan, October 2003

[Friedrich 02] Friedrich W. "ARVIKA – Augmented Reality for Development, Production and Service", Prooceed-ings of the ISMAR 2002, IEEE and ACM Interna-tional Symposium on Mixed and Augmented Re-ality, Darmstad, pp 3-4, 30th Sept-1 Oct, 2002

[Kato 01a] Kato, H., Billinghurst, M., Poupyrev, I., "Tangible Augmented Reality", in Augmented Reality: the Interface is Everywhere, SIGGRAPH 2001 Course Notes 27, 2001

[Kato 01b] Kato, H., "Developing Applications with AR-Toolkit", SIGGRAPH 2001 Course Notes 27, 2001.

[Kawashima 01] T. Kawashima, K. Imamoto, H. Kato, K. Tachi-bana, M. Billinghurst, "Magic Paddle: A Tangible Augmented Reality Interface for Object, Manipu-lation", Proc. on ISMR2001, pp.194-195, 2001

[Kopp 03] Koop, S. et al., "Imitation games with artificial agent: From mimicking to understanding shape-related iconic gestures, University of Bielefeld, Germany, Gesture Workshoop 2003, April 2003, Genova, Italy

[Milgram 99] Milgram, P., Herman, C. J., "A Taxonomy of Real and Virtual World Display Integration", in Mixed Reality, Merging Real and Virtual Environments, Ohmshda & Springer-Verlag, pp 5-30, 1999

[Reiners 01] Reiners D., Stricker, D., Klinker, G. "Augmented Reality: Applications", Computer Graphik Topics, reports on Computer Graphics, 4/98 Vol 10, pp 36-38.

[Wachsmuth 03] Wachsmuth I. et al, "Deixis in Multi-modal Human Computer Interaction: An Interdisciplinary Approach", University of Bielefeld, Germany, Gesture Workshoop 2003, April 2003, Genova, Italy