The ORIGAMI Project: advanced tools and techniques for high-end mixing and interaction between real and virtual content

G. Bazzoni (PhT[1]), E. Bianchi (ChT[2]), O. Grau (BBC[3]), A. Knox (FmS[4]),
R. Koch (CAU[5]), F. Lavagetto (DIST[6]), A. Parkinson (FmS), F. Pedersini (PhT),
A. Sarti (PdM[7]), G. Thomas (BBC), S. Tubaro (PdM)

Abstract

ORIGAMI is an EU-funded IST project with the goal to develop advanced tools and new production techniques for high-quality mixing of real and virtual content for film and TV productions. In particular, the project proposes new techniques for 3D camera calibration and tracking and for 3D scene reconstruction and plenoptic modelling. Such solutions will support a variety of different object classes and application scenarios, including single-camera modelling of static environments, and multi-camera modelling of dynamic studio scenes. The developed methods will be integrated into a new software framework, which is being developed within the project.

ORIGAMI also proposes novel techniques that enable real and virtual actors to interact with each other and with the real and virtual elements of the extended set.

1. Introduction

In the past few decades, the use of digital effects has become more and more common in film production, not just for the motion picture industry but also for the production of TV programmes and commercial videos. In fact, the latest advances in the fields of image processing, computer vision and computer graphics have enabled the film industry to merge synthetic content and natural content with a stunning level of realism and progressively decreasing costs.

Originally, digital techniques were primarily used to simulate physical conditions of extreme danger or to obtain results that would have been impossible to achieve otherwise. Thanks to the level of realism that can be achieved today, digital effects are currently being used more and more to cut down production costs by constructing accurate computer models of environments (e.g. large constructions, buildings, neighbourhoods or even entire cities) or by altering pre-existing ones (e.g. landscapes).

Unfortunately, digital effects for filmmaking can only be manufactured and applied in an off-line *postproduction* phase that makes massive use of computational power and requires highly specialised technicians. This fact still discourages many (especially European) film directors from making extensive use of digital effects for fear of losing control over the film. This, indeed, represents a serious limitation as, besides ruling out a wide range of spectacular options, it also forces them to adopt less convincing mechanical solutions and physically build expensive and less realistic sets, with the result of making the film less competitive without a reductions of production costs.

Those film directors who choose to embark onto the use of heavy-duty digital effects, however, have to accept dramatic changes in their way of working. In fact, during the actual filming they cannot easily have a preview of what the result will look like at the end of the postproduction work. The lack of a WYSIWYG (What-You-See-Is-What-You-Get) approach to the filming of live action in a *chroma-key studio* (a controlled environment where actors perform in front of a blue screen for chroma-keying purposes), for example, makes it difficult to plan camera trajectories and lighting. As the film director and the director of photography can do little more than *guessing* the final outcome, they often end up either keeping what they blindly obtain (and settling for sub-optimal results), or having to pay for a number of expensive screen tests (prototypes of digital

[1] Phoenix Tools, SRL, Via Cardano 8, 20124 Milano, Italy
[2] Chinatown SRL, Via Aleardo Aleardi 12, 20154 Milano, Italy
[3] BBC Research and Development, Kingswood Warren, Tadworth, Surrey, KT20 6NP
[4] FrameStore. 9 Noel Street, London W1V 8GH, England
[5] Multimedia Information Processing, Christian-Albrechts-Universitaet zu Kiel, Institut fuer Informatik und Praktische Mathematik. Herman-Rodewaldstr. 3, D-24098 Kiel, Germany
[6] Dipartimento di Informatica, Sistemistica e Telematica, Via all'Opera Pia 13 - 16145 Genova, Italy
[7] Dipartimento di Elettronica e Informazione - Politecnico di Milano. Piazza Leonardo Da Vinci 32, 20133 Milano, Italy

Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission (3DPVT'02)
0-7695-1521-5/02 $17.00 © 2002 IEEE

effects). Similar problems are encountered by the actors in a virtual studio, as all they see is a blue background. Once again, they lack a visual feedback of the key elements of the scene, not to mention the necessary positional and synchronisation cues, which are usually replaced by rough marks on the floor and timed gestures made by the crew.

All such problems have a severe impact on the final production costs, as it is difficult to predict how much postproduction work will be necessary. Another source of uncertainty is in the production costs related to the set (real and virtual), as it is usually difficult for film director, scenographer, director of photography and producers, to come to an agreement over something that cannot be quickly sketched or pre-visualised.

1.1. Project objectives

Currently, the mixing between real, virtual, off-line and on-line content is possible only after endless and painstaking post-production work, based on a trial-and-error approach. In fact, with the technology available today, the film footage relative to the off-line content can either be used *as is*, or it can be altered through the addition of synthetic elements. One thing that cannot easily be done is to arbitrarily alter the viewpoint from which the off-line scenes are acquired. All this forces the camera motion used for filming the live action to strictly follow the camera motion of the off-line footage, which is exactly the opposite of what would be desirable. In fact, the final viewpoint should mainly depend on the on-line content, where the acting takes place. Furthermore, controlling or even just constraining the camera motion during the filming of off-line content is usually less easy (if not impossible) than in a controlled environment such as a studio. Finally, a number of problems related to illumination conditions, shadows, blurring, occlusions, inevitably arise.

ORIGAMI intends to develop advanced technology to perform a seamless mixing of *real* and *virtual* content coming from both *live action* and *off-line* footage. The interaction between real actors and virtual actors (computer animations that are often equipped with a certain intelligence and autonomy) constitutes the "live" content. All the elements of the environment that do not react to the live action are considered part of the off-line content. In fact, such elements are obtained through the "virtualisation" (digital modelling) of off-line footage of a real environment, which may be digitally altered and/or enriched with synthetic elements (*extended set*). The off-line content is usually made of static or quasi-static elements, but it could also include passive "actors", which are dynamic elements that are not expected to react to the live action.

The need to distinguish between *on-line* and *off-line* content comes from the fact that we would like final viewpoint to be decided during the filming of the live action. This means that the off-line content needs to be authored in advance through a process of virtualisation of the reality (3D/plenoptic modelling from off-line footage, set extension, passive actor modelling, etc.). This is particularly relevant when safety is at risk (e.g. hazardous or hostile environments); or in case of external constraints (e.g. timed or non-repeatable environmental conditions); or in the presence of cost limitations.

In order to remove the a-priori constraint on camera motion and provide the film director with the freedom to decide it during the live action, the ORIGAMI project will develop and integrate techniques that will enable a high-end digitisation of the off-line content. The result of this operation will be a virtual environment that can be visualised from an arbitrary viewpoint trajectory, with lighting conditions that can be quite easily altered.

A number of results in this direction are already available in the literature and have been developed in previous EC projects, but never before with the goal of achieving the level of realism expected for the applications addressed by ORIGAMI. As we may expect, image-based digitisation techniques based on the geometric modelling of the viewed surfaces have some intrinsic limitations as far as surface complexity is concerned. Furthermore, they usually suffer from serious problems when transparent or highly reflective surfaces are involved. In order to overcome these limitations, ORIGAMI approaches the modelling problem with an integrated solution of practical usability, based on the synergetic use of a variety of existing and novel computer vision techniques [1][3][4], in order to cover a wide range of possible cooperative and non-cooperative scenes. More specifically, ORIGAMI is working on a synergistic integration of image-based rendering techniques (plenoptic scene modelling) with geometric/radiometric surface reconstruction techniques. Furthermore, the ORIGAMI project is working on novel techniques for the complete or partial geometric 3D modelling of scene objects from existing off-line footage without requiring an explicit segmentation between object and environment. All such solutions are uncalibrated, as they will use video sequences acquired with unconstrained camera motion [1][4][5][6].

As for the live content, the ORIGAMI project is focusing on the dynamic 3D modelling of actors in the studio [1], using a multicamera system. The goal is to obtain rough shape estimation in real time, and more accurate 3D models in an off-line session. As we will see later, a rough real-time shape estimation will be used to provide virtual actors with a feedback on position and volume occupation of real actors.

A key innovative contribution of ORIGAMI is the mixing of realities, which is conducted in such a way to enable real and virtual actors to interact with each other and with real and virtual elements of the extended set.

Real content is acquired in a studio and mixed with virtual or virtualised content in an interactive fashion. As part of ORIGAMI activities, existing "studio" technology is being improved and integrated with novel solutions that will enable an effective interaction between real and virtual scene content in a variety of forms. In particular, new techniques for the estimation of location, shape, texture, lighting and radiometric properties of actors and objects in the studio, based on calibrated multi-camera analysis are being developed [9]. Furthermore, novel solutions are being developed and tested, to provide the actors in the virtual set with a visual feedback of the extended (virtualised) set. All this enables several forms of interaction between real and virtual scene content. Mutual "visual" feedback can be used, for example, for event synchronisation; obstacle avoidance; occlusion awareness; or even eye contact. Information on shape, illumination and reflectivity of actors and elements in the studio enables radiometric interaction between real and virtual scene content. For example, it will be possible to have a real actor in the studio cast a shadow onto the virtual environment, and to improve the quality of photorealism in the rendering process.

The overall ORIGAMI workflow is summarised in Fig. 1, where all the building blocks of both off-line and on-line activities are integrated together. In this scheme, pre-visualisation and virtual theatre are kept separate, because of their different functional roles. Pre-visualisation, in fact, is here intended for set planning and direction planning, while the virtual theatre corresponds to the real-time visual feedback used by the actors, the director for film production, or the final product in the case of live TV programs.

1.2. Consortium

The ORIGAMI consortium was created with the intent of synergistically combining complementary technical-scientific expertise and diverse marketing interests. All the participants have their own areas of expertise in key aspects of the project.

The ORIGAMI consortium contains a mixture of commercial and academic partners, ranging from Phoenix Tools, which is a commercial software developer; to pure research institutions such as the University of Kiel, the Politecnico di Milano and the University of Genova; to the R&D department of a large commercial enterprise (the BBC). End users play a key role in the Consortium. In addition to the BBC, the Consortium includes Framestore, which is a successful post-production and digital effects company for film production; and Chinatown, a well-known end-user in the area of advertisement.

2. Scene Reconstruction Techniques

One key problem that the ORIGAMI Project is dealing with is the high-end image-based modelling of natural scenes, i.e. environments and objects. This off-line modelling process is expected to be conducted in a scalable and effective fashion, in order to be of practical usability not just for set-planning, but also for high-end compositing and finishing. In order for the system to be able to adapt to a wide range of cooperative and uncooperative scenes, two classes of solutions based on unconstrained video sequences are being developed and synergistically integrated:

- **full-3D geometric and radiometric object modelling:** metric reconstruction of all the surfaces that describe the object's geometry, plus modelling of the reflectance map of such surfaces and of the illumination. Among the methods under study are:
  - volumetric techniques based on the evolution of multi-resolution levelsets, driven by several sources of information, including texture mismatch, feature proximity and extremal boundary deformation. The methods are topology-independent, robust and fast. They produce models of scalable complexity, and do not necessitate background segmentation.
  - radiometric analysis techniques based on the robust estimation of the non-lambertian reflectance from the available footage/lightfield. Such techniques will be tested against inadequate sampling of the viewpoint space, and will return an inherently consistent approximation of the surface's reflectivity.
- **plenoptic modelling** of the scene. The methods under development are based on an extensive acquisition of the scene based on an uncalibrated "scanning" of the viewpoint space (e.g. zigzagged camera trajectory). In order to do so, ORIGAMI is working on novel camera tracking techniques aimed at the "weaving of the viewpoint fabric" through the construction of virtual camera motion passing through crosspoints in the zigzagged camera trajectory. This connected grid of oriented viewpoints will be used to interpolate new viewpoints without the need of explicit (metric) estimation of the scene geometry. This method will prove suitable for the image-based rendering of extremely complex, highly reflective, or transparent surfaces, where geometric reconstruction would

inevitably fail. Finally, new solutions are under study to speed up the rendering process, reduce the storage requirements, and enable real-time visualisation.

Combining plenoptic and geometric techniques is a scientific challenge in the ORIGAMI project. Such solutions, in fact, implement two entirely different approaches to 3D scene representation as the former is image-based while the latter is model-based. A signal-driven solution such as the plenoptic one could be considered as unstructured as it does not use an explicit geometric model for view-point interpolation, and it does not require any radiometric estimation. Conversely, a model-driven solution based on geometric/radiometric estimation is the most structured example of 3D representation solution. In fact, not only can it rely on an explicit geometric model, but such a model is globally consistent within the whole viewpoint space.

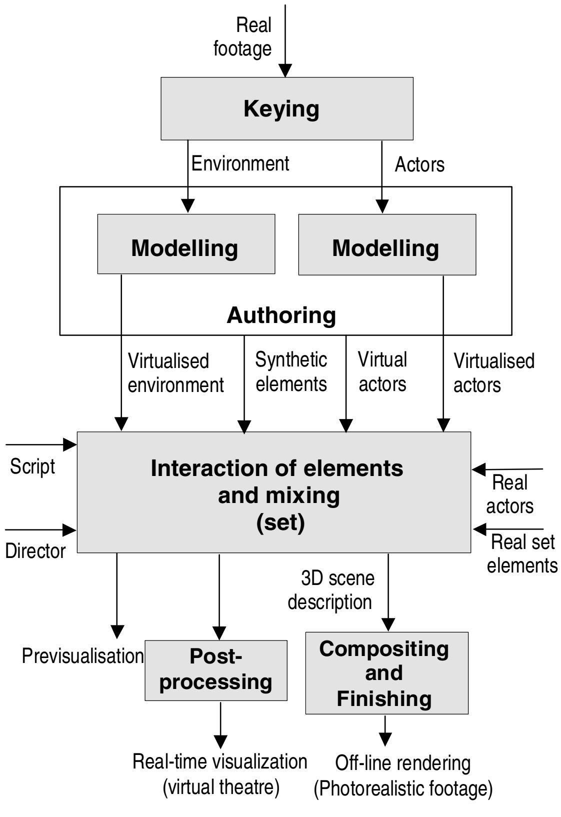

Fig. 1: Overall ORIGAMI workflow.

Among the advanced research goals, the ORIGAMI project is exploring the possibility of progressively adding "structure" to a plenoptic scene representation, with a control over the consistency of the underlying geometry throughout the viewing space. This would allow the two worlds (plenoptic and geometric/radiometric) not just to coexist but also to enable a certain degree of mutual interaction.

3. Interactive mixing of realities

The process of digitisation of the off-line content returns an virtualised environment that is ready to be visualised from an arbitrary viewpoint. This environment can now be populated with real and virtual actors whose mutual interaction is one of the key aspects of the ORIGAMI project.

One way to classify virtual actors could be as follows:

- *passive extras*, which perform pre-recorded actions
- *autonomous agents*, which are equipped with a limited intelligence (e.g. a flock of birds)
- *puppets*, whose motion is directly linked to that of an actor (puppeteer) using a body motion tracking system.
- *avatars*, which can be controlled and instructed at a semantic level (high-level behavioural rules and symbolic emotional controls)

Indeed, virtual actors may have full interaction with the virtual environment, just like real actors are allowed a full visual interaction with the real objects that are present in the ORIGAMI production studio. The ORIGAMI project, however, focuses on the development of novel techniques that enable real and virtual actors to interact with each other and with the real and virtual elements of the extended set, according to the scheme of Fig. 2.

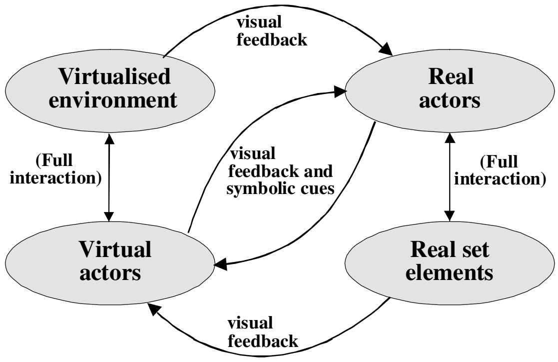

Fig. 2. Interaction map for live action.

Mutual "visual" feedback can be used to enable event synchronisation; obstacle avoidance; eye contact; occlusion awareness; etc. Information on shape, illumination and reflectivity of actors and elements in the studio can also enable radiometric interaction between real and virtual scene elements. For example, it would be possible to have a real actor in the studio project a shadow onto the virtual environment, and to improve the quality of photorealism in the rendering process.

Further forms of interaction will be developed, such as symbolic interaction (script and cues), and emotional interaction.

Feedback from real to virtual scene elements will be possible through the development of novel extended virtual studio techniques based on the simultaneous use of a set of calibrated fixed cameras. Chroma-keying will enable the extraction of the silhouettes of actors and objects in the virtual set. Using this information together with the images acquired in the studio, it will be possible to extract positional information, shape and reflectivity of objects and actors in the scene. A rough 3D model coming from this information can then be fed back to virtual actors in the studio, and may also be used to model some sort of radiometric interaction between real and virtual. For example, real actors could project a shadow onto the elements of the virtual environment.

Multicamera (dynamic) shape modelling could also benefit from techniques using motion tracking and motion compensation for rigid and flexible objects. Such solutions are currently being investigated to enable the computation of surface properties and a significant reduction of the amount of data to be processed.

Finally, in order to harmonise the captured studio scene with the extended (virtualised) set, algorithms for the estimation of scene illumination are under investigation, based on the contextualisation of lightfield calibration techniques to studio applications.

As for symbolic feedback, a variety of commercial solutions may be employed, ranging from systems for body pose estimation, head trackers, gaze trackers, eye-blinking detectors, etc. Additional symbolic feedback (mood description) could even come from a prosodic analyser.

Feedback from virtual to real scene elements will be possible through the development a projection system that allows a visualisation of the virtual scene without interfering with the studio illumination and without losing the possibility to perform chroma-keying within the studio. For this purpose, a special retro-reflective cloth in conjunction with active illumination will be used. In addition to the virtualised environment, the content to be projected will also include positional cues (scene objects and virtual characters), temporal and textual cues (symbolic information on what to say and when to say it), and other types of symbolic high-level cues.

In particular, a significant contribution to this activity will be the development of avatar-based emotional feedback. This will, for example, help real actors that are not trained for acting in a studio and, in particular, improve the action direction.

4. Production system

4.1 Off-line elements of the production system

In productions like movies it is mostly assumed that the material and the scene description can always be modified, refined, and composited in a post-production phase. These productions are strictly script-based, and require high resolution and cutting-edge quality in the final rendering results. Besides enabling the correction and the refinement of the scene models, the post-production session allows an accurate compositing and finishing for photorealistic rendering purposes.

A schematic view of the production system is shown in Fig. 3. The various blocks of this type of application operate in an off-line fashion, including the mixing of real and virtual content. The main functionalites are:

- Modelling of the static environment (3D shape and texture)
- Segmentation of moving objects
- Creation of plenoptic models of static environments
- Modelling of moving actors in the studio (3D shape and texture)

A more detailed description of the approach for modelling actors in the studio can be found in [9].

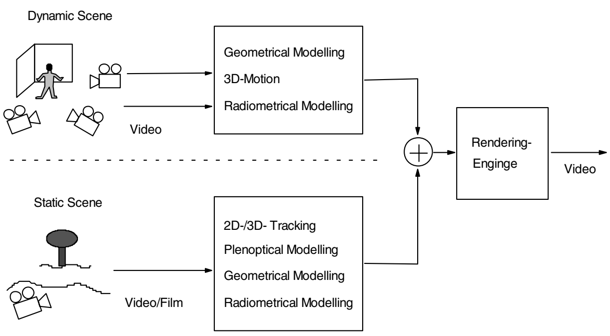

Fig. 3: Functional block diagram for the production system.

4.2 Real-time elements of the production system

The real-time elements to be developed in the ORIGAMI production system are used in a studio environment. They are intended for on-set visualisation of the virtual parts of a scene for both the actors and the director. The arrangement envisaged is shown in the Fig. 4 below. The figure shows a cyclorama covering three sides of the acting area. The cyclorama is made of retro-reflective cloth, onto which views of the virtual parts of the scene are projected. The floor also has a covering of retro-reflective cloth, and virtual elements of the scene are projected onto the floor from a projector mounted high up and towards the back of the set. The projected images show a view of the virtual elements of the scene, with the correct perspective to match the viewpoint of the actor.

The actor is imaged by a number of TV cameras (at least six), each with a ring of blue LEDs around the lens, so that the cyclorama appears blue to the camera. Most cameras are fixed in known positions, while at least one is moveable and its position/orientation will be tracked by a real-time tracking system developed by BBC.
The studio setup will be able to provide some information on the 3D scene geometry in real time. Some of this information will be necessary for a successful interaction between real and virtual:

- the position of the actor's head will be used to control the rendering of the projected images (to match the actor's point of view),
- the approximate volume occupied by the actor, which can be used by the virtual actor’s control system and by the rendering process (for the control of occlusions between real and virtual elements). This will also be used to blank the portion of the front/top-projected images that would otherwise fall onto the actor and thus be visible to the cameras.

A high-quality video image of the actor from a moveable viewpoint will be provided using a TV camera equipped with the BBC's real-time camera tracking system. The image from this camera could be keyed onto a rendered background image as in a conventional studio, providing a broadcast-quality real-time signal if operation as a conventional studio was required. The rendered background image may include occlusion effects computed using the rough real-time 3D actor data.
The data captured in the studio can be combined in post-production with elements of the off-line production. This will allow, for example, a programme that is recorded 'live' to be enhanced by the inclusion of scenes in post-production that make full use of the high-quality 3D model of the actor created from the multiple camera views. Such scenes might include, for example, shots from camera positions that would be physically impossible to achieve in the confines of the studio, or shots incorporating special effects, or shots with highly-complex backgrounds.
A key characteristic of live television productions is the fact that they may not be strictly script-based. This situation, indeed, turns out to be more demanding as far as the management of the interactions between real and virtual are concerned, therefore some constraints on the real-virtual interaction will have to be set. In particular, virtual actors will have to be equipped with enough intelligence to grant them a certain autonomy. Behavioural animation mechanisms will make them sensitive to the environment, and will allow them to take decisions on the set, depending on the behaviour of the other actors (real or virtual). The feedback provided by the enhanced virtual set, coming from the 3D shape estimation mechanisms will thus play a crucial role in controlling all such interaction mechanisms between real and virtual actors.

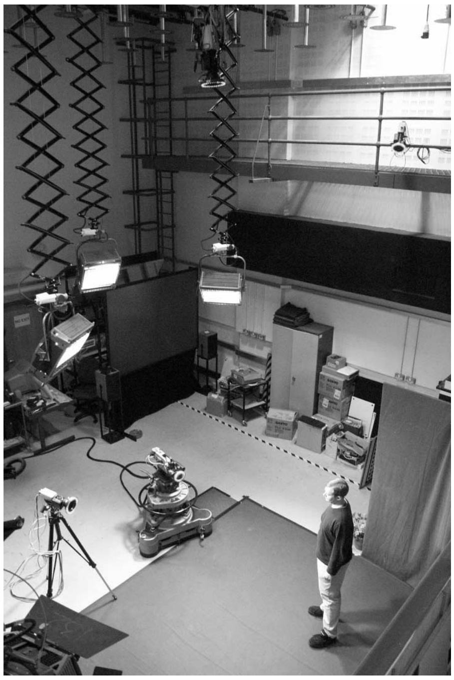

Fig. 4: Studio arrangement.

5. SW Framework

The ORIGAMI project presents many challenges, mainly associated to the expected level of quality of the results (considering the application scenarios), and to the complexity of the interactions that could take place in the ORIGAMI studio. In order to achieve these goals, ORIGAMI intends to integrate all software modules within a common flexible environment, and to provide the user with a friendly and intuitive graphical interface.
The SW architecture will be based on common object sharing, which constitutes a generalisation of the Object Linking and Embedding (OLE) structure. A powerful and flexible Application Programming Interface (API), will be developed in order to facilitate SW integration.
Finally, ORIGAMI will provide the user with a programmable non-modal Graphical User Interface (GUI) based on connection trees, for a WYSIWYG approach to application development.
The proposed structure exhibits a great deal of potential, as it enables a synergistic use of all software models developed within the Project. In fact, it will be not only possible to process data with several techniques in parallel, but it will also be possible to implement processing loops by feeding data back from application to application. This will enable a progressive refinement of the information and of the quality of the final result.

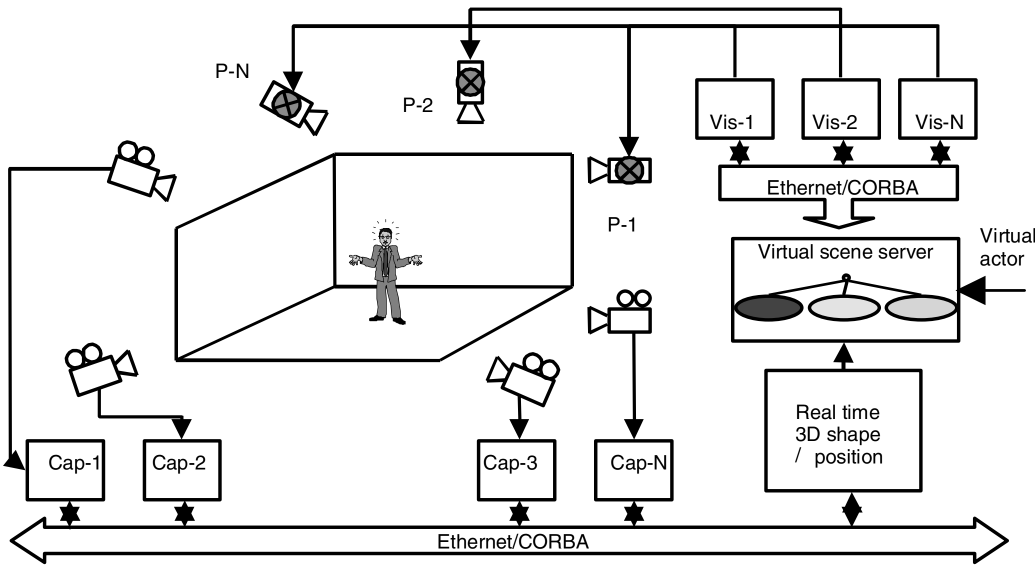

Fig. 5: Components of the real-time production and feedback system. Each camera of the multi-camera setup is connected to a server (Cap-1,..,Cap-N). A rough 3D model is created in real-time in addition to the actor head position. This information is distributed together with the virtual scene to a number of visualization units (Vis-1,..,Vis-N) that drive a data projector for the actor feedback.

Furthermore, this will enable not just a customisation of the software with respect to the application, but the creation of a dynamic user interface. The GUI, in fact, will be context-dependent and will bring forward the required modules while hiding the unneeded ones, depending on the processing phase.

6. Conclusions

The ORIGAMI project exhibits a mix of scientifically challenging goals and extremely pragmatic application scenarios, with measurable performance in terms of cost-effectiveness. In fact, one peculiarity of this project is in the fact that the quality of the results is not a goal but a constraint.

The goals of ORIGAMI led us to the choice of the project name. In fact, origami is the art of paper folding. The word is Japanese, literally meaning "to fold" (oru) "paper" (kami). In fact, it is more than that, as it gives life to a flat world.

References

[1] O. Grau, M. Price, G.A. Thomas, “Use of 3D Techniques for Virtual Production”, in Proc. of SPIE, Conf. Proc. of Videometrics and Optical Methods for 3D Shape Measurement, Vol. 4309, Jan. 2001, San Jose, USA.

[2] F. Pedersini, A. Sarti, S. Tubaro: "Multicamera Systems: Calibration and Applications". *IEEE Sig. Proc. Mag.*. Vol. 16, No. 3, May 1999, pp. 55-65.

[3] F. Pedersini, A. Sarti, S. Tubaro: "Visible Surface Reconstruction with Accurate Localization of Object oundaries", *IEEE Tr. on Circ. and Sys. for Video Technology*, Vol. 10, No. 2, March 2000, pp. 278-291.

[4] R. Koch, M. Pollefeys, L. Van Gool: Realistic surface reconstruction of 3D scenes from uncalibrated image sequences. J. Vis. Comp. Anim. 2000;11:115-127

[5] M. Pollefeys, R. Koch, L. Van Gool. “Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters”, IJCV Special Marr Price Issue. Vol. 32, No. 1, pp. 7-25, 1999.

[6] R. Koch, M. Pollefeys, B. Heigl, L. Van Gool, H. Niemann. “Calibration of Hand-held Camera Sequences for Plenoptic Modeling”, Proc. ICCV'99.

[7] F.Lavagetto, “Time-Delay Neural Networks for Estimating Lip Movements from Speech Analysis: A Useful Tool in Audio/Video Synchronization”, IEEE Trans. on Circuits and Systems for Video Technology, Vol. 7, n. 5, 1997, pp. 786-800.

[8] F.Lavagetto, R. Pockaj, “The Face Animation Engine: towards a high-level interface for the design of MPEG-4 compliant animated faces”, IEEE Tr. Circ. and Sys. for Video Techn., Vol. 9, N.2, March 1999.

[9] G.A. Thomas, O. Grau, “3D Image Sequence Acquisition for TV & Film Production”, Proc. of 1st Int. Sym. on 3D Data Processing Vis. and Transm. (3DPVT 2002), Padova, Italy, Jun 19-21, 2002.