Visual Spatialization of a Meeting Room from 2D Uncalibrated Views

Katia Fintzel[‡ł] & Jean-Luc Dugelay[‡]

[‡]**Institut Eurécom**, Multimedia Communications Department
2229, route des Crêtes, BP 193,
06904 Sophia Antipolis, FRANCE
Tel: +33 (0)4 93 00 26 26; fax: +33 (0)4 93 00 26 27

[ł]**Espri Concept**,
Les Taissounières HB2, BP 277,
06905 Sophia Antipolis, FRANCE
Tel: +33 (0)4 92 38 82 50; fax: +33 (0)4 92 38 82 51

Abstract

In this paper, we introduce an image processing tool, *Video Spatialization*, used for designing a new approach for multipoint teleconferencing systems for very low bit rate links (internet, mobile communications). This type of systems is based on the immersion of all the participants in a common virtual meeting place, like real meetings do, to increase teleconferencing realism. The tool proposed here is used for background control in the virtual scene, especially for consistence with the users position and motion during the meeting session. So this paper contains (1) a review of the trilinearity theory, (2) an efficient algorithm for real views reconstruction, (3) some extensions to synthesize unknown views and (4) the integration of *Video Spatialization* in the context of the TRAIVI project.

1 Introduction

The problem discussed in this paper is the reconstruction of real points of view of a meeting room and the synthesis of virtual ones, from a limited set of 2D uncalibrated views and without resorting to a 3D CAD model of the scene. This is called *Video Spatialization*.

An efficient "mesh-oriented" approach for this kind of synthesis, based on the trilinearity theory and a step of analytical inference are presented in section 2. Such a process aims at offering the possibility for the user to visualize the 3D scene from anywhere and towards any direction. In section 3 we present early visual results, concluding remarks and perspectives for the TRAIVI project, which takes advantage of Video Spatialization techniques in the context of virtual teleconferencing system.

2 Views Synthesis

2.1 An efficient approach for real image reconstruction

We propose an algorithm for real view reconstruction from uncalibrated 2D points of view of a 3D scene based on *trilinear tensors*, first modeled by A. Shashua [1, 2, 3, 4] to understand the geometry of correspondences between three initial images. These relations generalize well known bilinearities, called epipolar constraints [5, 6], and allow us to reconstruct an existing view from two other neighboring views without explicit calibration stage. Following are the steps:

- an analysis step, using more than seven corresponding points in the three original uncalibrated views, to estimate the eighteen parameters of a trilinear form, (for more details about trilinear parameters definition see [7, 1, 2] and [8]).
- a synthesis step, using corresponding points of the external images and the estimated parameters $(\alpha_i)_{i=1..18}$ to reconstruct the cen-

tral view, by the following system:

$$\left\{\begin{array}{l} x'(\alpha_1 x'' + \alpha_2 y'' + \alpha_3) + x'x(\alpha_4 x'' + \alpha_5 y'' + \alpha_6) + \\ x(\alpha_7 x'' + \alpha_8 y'' + \alpha_9) + \alpha_{10} x'' + \alpha_{11} y'' + \alpha_{12} = 0 \\ y'(\alpha_1 x'' + \alpha_2 y'' + \alpha_3) + y'x(\alpha_4 x'' + \alpha_5 y'' + \alpha_6) + \\ x(\alpha_{13} x'' + \alpha_{14} y'' + \alpha_{15}) + \alpha_{16} x'' + \alpha_{17} y'' + \alpha_{18} = 0 \end{array}\right.$$

where (x, y), (x', y') and (x'', y'') point at the pixel coordinates of homologous points respectively in the left, middle and right view.

In the literature, the synthesis step of the method often requires a dense matching preprocessing stage between points of external images in order to compute the luminance of each point of the regenerated central view [1, 2, 9]. This is still an unresolved computationally-expensive operation, so in this paper we suggest that the original images should be represented by a texture warped on associated meshes. In order to resynthetize the central view we only map the texture on a new mesh produced from the mesh nodes of the external images as shown in figure 1. As a result we have a plain synthesized image with no hole (misinformed point) and the time necessary to regenerate an image significantly decreases because it only depends on the number of mesh nodes, according to the scene complexity.

In terms of compression, a complete real view can be represented by only eighteen floating point numbers, called trilinear parameters, but encoding pre-processing and synthesis post-processing are obviously needed. However, possible extensions of the above method by infering on trilinear parameters values have been studied to create virtual points of view from the set of initial images.

2.2 Trilinear Parameters Manipulations for Virtual Views Synthesis

We can analytically corrupt the trilinear parameters as shown in figure 2 to simulate a focal length change or a geometrical 3D displacement of the camera relative to the reconstructed view and by this way we synthesize new unknown views.

All the analytical manipulations of trilinear parameters to create new points of view are detailled in the research reports [8, 10]. A complete explanation would be dull here, but the following array (figure 3) sums up the manipulations and especially shows the inputs needed to simulate each kind of transformation. Rotations and particularly translations are not trite without explicit calibration, but our work tries to keep the calibration implicit by restoring inputs from trilinear parameters [11, 12].

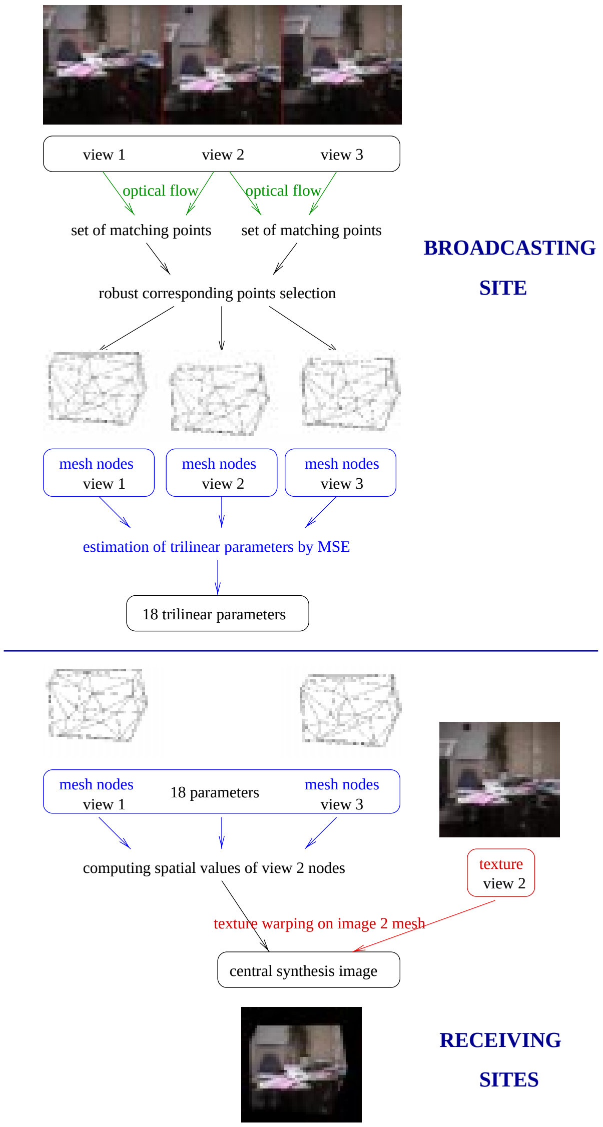


Figure 1: Real views regeneration procedure.

3 Results and Conclusion

3.1 Visual Results

A few virtual synthesized points of view are presented in figure 4.

3.2 TRAIVI Perspectives

Our work on video spatialization takes place in the larger TRAIVI[1] project, whose goal is to create a complete virtual teleconferencing system. In fact, the use of teleconferencing systems between multiple sites has considerably increased [13], because of industrial demands, but generally offers a poor quality of service [14]. The

[1] TRAIVI stands for "TRAItement des images VIrtuelles" (Virtual Images Processing)

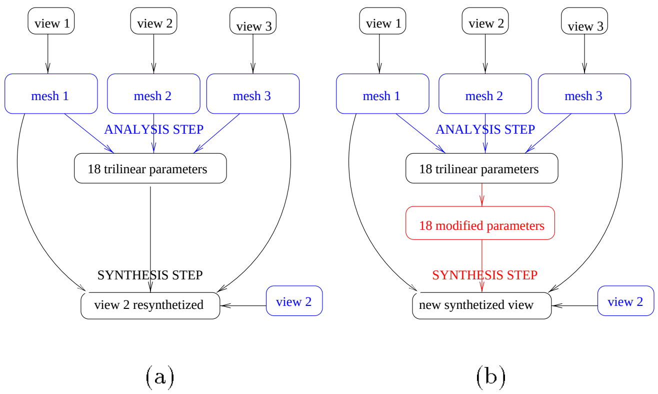


Figure 2: Comparison between real and virtual views synthesis methods: (a) sums up the real views regeneration method (b) outlines the procedure to synthesize a new point of view.

immersion of the participants in the same virtual environment, with the ability to move and look at the other participants, could make up for the lack of realism of classical systems and offer new ergonomic possibilities [15].

Video spatialization for background control is one of video processings we have to master in combination with model-based coding for participants control [16]. That is why we focus on the synthesis of meeting-room images, with a priority for real-time and realism of regenerated or unknown synthesized images, as opposed to the reconstruction accuracy. To that extent our “mesh-oriented” approach is fully justified in the context of the TRAIVI project.

Virtual views synthesis is particularly interesting for this application: we can now imagine a virtual meeting composed of a pre-processing stage preceding the session. During the pre-processing stage, information like the user’s position and the choice of the meeting area will be transmitted to a central site, which pre-computes, from a few real uncalibrated views, the corresponding vectors of trilinear parameters uploaded to each remote site. During the session each site, independently of each others, will be able to create locally by algebraïc processing new coherent points of view for its user based on his virtual position, motion parameters and domain of interest in the meeting room.

Our perspectives for the TRAIVI project are:

- the implementation of a complete room spatialization system, which requires a pick-up strategy to screen entirely and optimally the meeting space, dealing with the quantity of pre-downloaded textures and the user’s permitted motion granularity.
- the study of integration of 2D background images and 3D models of participants, which is still an open problem.

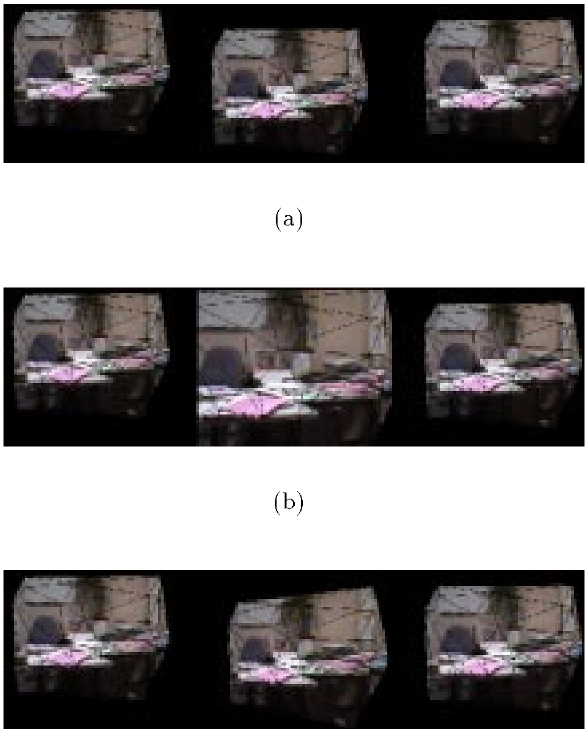
	Trilinear parameters	Foc.	Rot.
focal changes	$\begin{cases} \alpha'_i & = \alpha_i \\ & i = 1..6 \\ \alpha'_i & = c.\alpha_i \\ & i = 7..18 \end{cases}$		
camera rotation	$\begin{cases} \alpha'_1 = c_\eta \alpha_1 - \frac{s_\eta \alpha_{16}}{k_v^2 f^2} \\ \alpha'_2 = c_\eta \alpha_2 - \frac{s_\eta \alpha_{17}}{k_v^2 f^2} \\ \alpha'_3 = c_\eta \alpha_3 - \frac{s_\eta \alpha_{18}}{k_v^2 f^2} \\ \alpha'_4 = c_\eta \alpha_4 - \frac{s_\eta \alpha_{13}}{k_v^2 f^2} \\ \alpha'_5 = c_\eta \alpha_5 - \frac{s_\eta \alpha_{14}}{k_v^2 f^2} \\ \alpha'_6 = c_\eta \alpha_6 - \frac{s_\eta \alpha_{15}}{k_v^2 f^2} \\ \alpha'_i = \alpha_{i,i=7..12} \\ \alpha'_{13} = c_\eta \alpha_{13} + k_v^2 f^2 s_\eta \alpha_4 \\ \alpha'_{14} = c_\eta \alpha_{14} + k_v^2 f^2 s_\eta \alpha_5 \\ \alpha'_{15} = c_\eta \alpha_{15} + k_v^2 f^2 s_\eta \alpha_6 \\ \alpha'_{16} = c_\eta \alpha_{16} + k_v^2 f^2 s_\eta \alpha_1 \\ \alpha'_{17} = c_\eta \alpha_{17} + k_v^2 f^2 s_\eta \alpha_2 \\ \alpha'_{18} = c_\eta \alpha_{18} + k_v^2 f^2 s_\eta \alpha_3 \end{cases}$ $c_\eta = cos(\eta) \quad s_\eta = sin(\eta)$ $\eta =$ rotation angle	×	
camera translation	$\begin{cases} \alpha'_7 = \alpha_7 + k_u^2 f^2 r_{31}^1 c \\ \alpha'_8 = \alpha_8 + k_u^2 f^2 r_{32}^1 c \\ \alpha'_9 = \alpha_9 + k_u^2 f^2 r_{33}^1 c \\ \alpha'_{10} = \alpha_{10} + k_u^1 f^1 k_u^2 f^2 r_{11}^1 c \\ \alpha'_{11} = \alpha_{11} + k_u^1 f^1 k_u^2 f^2 r_{12}^1 c \\ \alpha'_{12} = \alpha_{12} + k_u^1 f^1 k_u^2 f^2 r_{13}^1 c \\ \alpha'_i = \alpha_{i,i=1...6,13...18} \end{cases}$	×	×

Figure 3: $(\alpha_i)_{i=1..18}$ are the initial trilinear parameters and $(\alpha'_i)_{i=1..18}$ stand for modified parameters.

References

[1] A Shashua. “Projective structure from uncalibrated images: structure from motion and recognition”. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, August 1994.

[2] A Shashua. “Trilinearity in visual recognition by alignment”. In *ECCV A*, pages 479–484, 1994.



(a)

(b)

(c)

Figure 4: Synthesized viewpoints after camera focal change or camera rotation: (a) initial sequence, (b) second camera focal change (c) second camera vertical rotation

[3] A Shashua. "Algebraic functions for recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, 1995.

[4] A Shashua & M Werman. "On the trilinear tensor of three perpective views and its underlying geometry". In *International Conference on Computer Vision*, Cambridge, June 1995.

[5] O Faugeras. "Quelques pas vers la vision artificielle en trois dimension". In *Technique et Science Informatique*, 1989.

[6] O Faugeras. *"Three-Dimensional Computer Vision: A Geometric Viewpoint"*. The MIT PRESS, 1993.

[7] A Shashua. "On geometric and algebraic aspect of 3D affine and projective structures from perspective 2D views". In A Zisserman & D Forsyth eds J-L Mundy, editor, *Applications of Invariance in Computer Vision.* Second European Workshop Invariants, Ponta Delagada, Azores, October 1993.

[8] K Fintzel & J-L Dugelay. "Défocalisation en Spatialisation Vidéo à partir de Trois Vues de Référence (Expressions Analytiques)". Technical report, EURECOM, Département Communications Multimédia, Sophia Antipolis, France, Fevrier 1996.

[9] J Blanc & R Mohr P Bobet. "Aspect cachés de la trilinéarité". In *Proc. RFIA'96 Conf.*, pages 137–146, Rennes, France, Janvier 1996.

[10] K Fintzel & J-L Dugelay. "Manipulations analytiques des paramètres trilinéaires pour la resynthèse d'images inédites". Technical report, EURECOM, Département Communications Multimédia, Sophia Antipolis, France, Novembre 1997.

[11] A Shashua & S Peleg B Rousso, S Avidan. "Robust recovery of camera rotation from three frames". In *CVPR 96*, June 1996.

[12] K Fintzel & J-L Dugelay. "Restitution des paramètres de rotations initiales à partir des paramètres trilinéaires d'un système de trois caméras". Technical report, EURECOM, Département Communications Multimédia, Sophia Antipolis, France, Décembre 1997.

[13] K Okada F Maeda Y Ichikawaa & Y Matsushita. "Multiparty videoconferencing at virtual social distance: MAJIC design". In *ACM'94*, pages 385–393, 3-14-1 Hiyoshi, Kohoku-ku, Yokohama, 223 Japan, October 1994.

[14] T Talley & F.D Smith K Jeffay, D.L Stone. "Adaptive, best effort delivery of audio and video across packet-switched networks". In *3rnd Intl. Workshop on Network and OS Support for Digital Audio and Video*, SanDiego, CA, November 1992.

[15] H Fuchs G Bishop K Arthur L McMillan R Bajcsy S Lee H Farid & T Kanade. "Virtual space teleconferencing using a sea of cameras". In *First International Symposium on Medical Robotics and Computer-Assisted Surgery 2*, pages 161–167, Pittsburgh, Pa, September 1994.

[16] J-L Dugelay S Valente. "Model-Based Coding and Virtual Teleconferencing". In *2nd Symposium "Advances in Digital Image Communication"*, Erlangen, Germany, April 1997.