Image Reconstruction and Interpolation in Trinocular Vision

J.-L. DUGELAY & K. FINTZEL

Institut EURECOM, MultiMedia Communications dept.
route des Crêtes, BP 193, F-06904 Sophia Antipolis Cedex
e-mail. dugelay@eurecom.fr
url. http://www.eurecom.fr/~image
tel. (+33) 93 00 26 66 - fax. (+33) 93 00 26 27

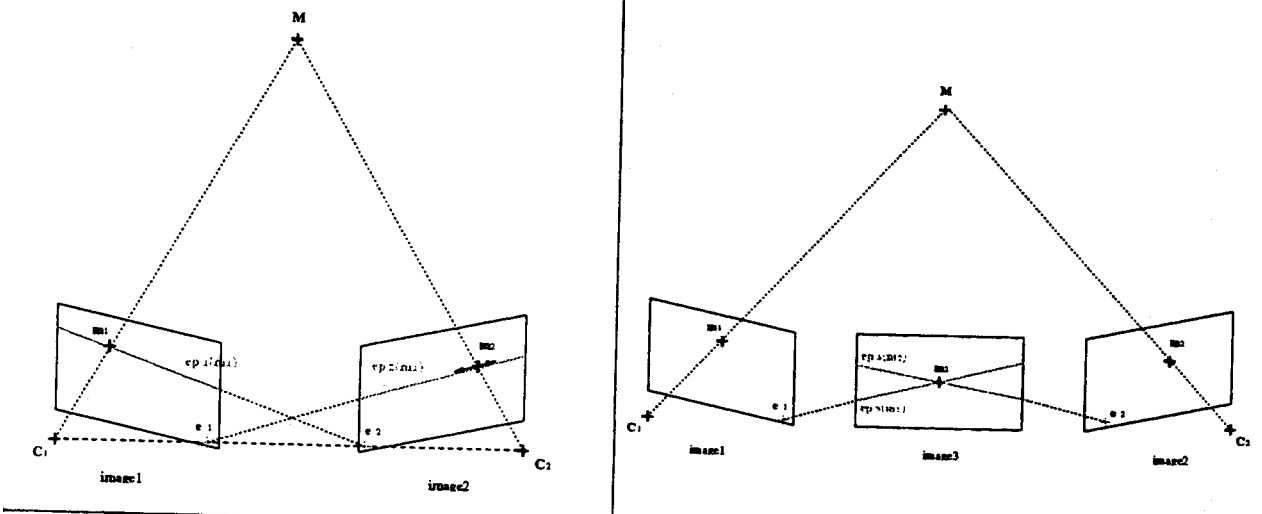
keywords. Tele-Virtuality, Video Spatialization, Trinocular Vision, Focal length change.

Abstract. The aim of "video spatialization" is to build virtual points of view of a real meeting room. This is achieved by interpolating a finite set of uncalibrated reference views of the considered scene. Hence, the ultimate goal consists in offering the possibility to visualize a meeting room from anywhere and toward any direction unlike currently available commercial teleconferencing systems which impose a unique point of view for every remote site participant.

I. INTRODUCTION

TRAIVI is a tele-virtuality project which intends to create and run virtual meeting places. Lots of techniques in audio and video processing [1], such as face cloning, audio and video spatialization, are essential to obtain a good degree of realism, in spite of low bit rate bindings.

The aim of "video spatialization", treated in this paper, is to build existent and non-existent views of a real 3D scene with at least two different, neighboring and uncalibrated views of the considered scene. In this work, no knowledge about the pick-up equipment is assumed: neither internal nor external calibration parameters. This study uses geometric aspects of computer vision in the trinocular case proposed by A. Shashua, recalled in **section II**. An effective algorithm able to reconstruct an existing view from two neighbouring other views is presented in **section III**. Several image processing techniques like matching and interpolation, already developed in the context of motion and disparity estimation-compensation, are used. The reconstruction of some virtual points of view from a set a three reference views is studied in **section IV**. In particular, the simulation of a virtual focal length change of one of the three reference video cameras is presented. In **section V**, video spatialization using more than three views, covering an entire meeting space, is discussed.



binocular vision, the epipolar geometry allows to define, r a given pixel m1 in one of the two views, a set of pixels the other view which defines a line called epipolar line 2(m1), along which the homologous pixel m2 of the first e is located.

By extending the binocular property to a set of three images, it is clear that from a couple of pixels m1 and m2, the localization of the third one m3 in the third view can be directly deducted from the intersection of the epipolar lines ep3(m1) and ep3(m2).

figure 1: Epipolar Geometry in binocular and trinocular vision

1st form,

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

2nd form,

$$\begin{cases} x''(\beta_1 x+\beta_2 y+\beta_3)+x''y'(\beta_4 x+\beta_5 y+\beta_6)+y'(\beta_7 x+\beta_8 y+\beta_9)+\beta_{10}x+\beta_{11}y+\beta_{12}=0 \\ y''(\beta_1 x+\beta_2 y+\beta_3)+y''y'(\beta_4 x+\beta_5 y+\beta_6)+y'(\beta_{13} x+\beta_{14} y+\beta_{15})+\beta_{16}x''+\beta_{17}y+\beta_{18}=0 \end{cases}$$

3rd form,

$$\begin{cases} x'(\gamma_1 x+\gamma_2 y+\gamma_3)+x'x''(\gamma_4 x+\gamma_5 y+\gamma_6)+x''(\gamma_7 x+\gamma_8 y+\gamma_9)+\gamma_{10}x+\gamma_{11}y+\gamma_{12}=0 \\ y'(\gamma_1 x+\gamma_2 y+\gamma_3)+y'x''(\gamma_4 x+\gamma_5 y+\gamma_6)+x''(\gamma_{13} x+\gamma_{14} y+\gamma_{15})+\gamma_{16}x+\gamma_{17}y+\gamma_{18}=0 \end{cases}$$

4th form,

$$\begin{cases} x'(\theta_1 x+\theta_2 y+\theta_3)+x'y''(\theta_4 x+\theta_5 y+\theta_6)+y''(\theta_7 x+\theta_8 y+\theta_9)+\theta_{10}x+\theta_{11}y+\theta_{12}=0 \\ y'(\theta_1 x+\theta_2 y+\theta_3)+y'y''(\theta_4 x+\theta_5 y+\theta_6)+y''(\theta_{13} x+\theta_{14} y+\theta_{15})+\theta_{16}x+\theta_{17}y+\theta_{18}=0 \end{cases}$$

where (x,y), (x',y') and (x",y") denote pixel positions inside image 1, 2 and 3.

Equation 1.

$$\begin{cases}
\alpha_1 = \gamma_7 = \theta_7 = & t_1^2 f^2 r_{31}^3 - r_{11}^2 f^2 t_3^3 \\
\alpha_2 = \gamma_8 = \theta_8 = & t_1^2 f^2 r_{32}^3 - r_{12}^2 f^2 t_3^3 \\
\alpha_3 = \gamma_9 = \theta_9 = & t_1^2 f^2 r_{33}^3 - r_{13}^2 f^2 t_3^3 \\
\alpha_4 = \beta_4 = \gamma_4 = \theta_4 = & r_{31}^2 t_3^3 - t_3^2 r_{31}^3 \\
\alpha_5 = \beta_5 = \gamma_5 = \theta_5 = & r_{32}^2 t_3^3 - t_3^2 r_{32}^3 \\
\alpha_6 = \beta_6 = \gamma_6 = \theta_6 = & r_{33}^2 t_3^3 - t_3^2 r_{33}^3 \\
\alpha_7 = \beta_7 = \gamma_1 = & t_3^2 f^3 r_{11}^3 - r_{31}^2 f^3 t_1^3 \\
\alpha_8 = \beta_8 = \gamma_2 = & t_3^2 f^3 r_{12}^3 - r_{32}^2 f^3 t_1^3 \\
\alpha_9 = \beta_9 = \gamma_3 = & t_3^2 f^3 r_{13}^3 - r_{33}^2 f^3 t_1^3 \\
\alpha_{10} = \gamma_{10} = & t_1^3 f^2 f^3 r_{11}^2 - r_{11}^3 f^2 f^3 t_1^2 \\
\alpha_{11} = \gamma_{11} = & t_1^3 f^2 f^3 r_{12}^2 - r_{12}^3 f^2 f^3 t_1^2 \\
\alpha_{12} = \gamma_{12} = & t_1^3 f^2 f^3 r_{13}^2 - r_{12}^3 f^2 f^3 t_1^2 \\
\alpha_{13} = \beta_{13} = \theta_1 = & t_3^2 f^3 r_{21}^3 - r_{31}^2 f^3 t_2^3 \\
\alpha_{14} = \beta_{14} = \theta_2 = & t_3^2 f^3 r_{22}^3 - r_{32}^2 f^3 t_2^3 \\
\alpha_{15} = \beta_{15} = \theta_3 = & t_3^2 f^3 r_{23}^3 - r_{33}^2 f^3 t_2^3 \\
\alpha_{16} = \theta_{10} = & t_2^3 f^2 f^3 r_{11}^2 - r_{11}^3 f^2 f^3 t_1^2 \\
\alpha_{17} = \theta_{11} = & t_2^3 f^2 f^3 r_{12}^2 - r_{12}^3 f^2 f^3 t_1^2 \\
\alpha_{18} = \theta_{12} = & t_2^3 f^2 f^3 r_{13}^2 - r_{13}^3 f^2 f^3 t_1^2 \\
\beta_1 = \gamma_{13} = \theta_{13} = & t_2^2 f^2 r_{31}^3 - r_{21}^2 f^2 t_3^3 \\
\beta_2 = \gamma_{14} = \theta_{14} = & t_2^2 f^2 r_{32}^3 - r_{22}^2 f^2 t_3^3 \\
\beta_3 = \gamma_{15} = \theta_{15} = & t_2^2 f^2 r_{33}^3 - r_{23}^2 f^2 t_3^3 \\
\beta_{10} = \gamma_{16} = & t_1^3 f^2 f^3 r_{21}^2 - r_{11}^3 f^2 f^3 t_2^2 \\
\beta_{11} = \gamma_{17} = & t_1^3 f^2 f^3 r_{22}^2 - r_{12}^3 f^2 f^3 t_2^2 \\
\beta_{12} = \gamma_{18} = & t_1^3 f^2 f^3 r_{23}^2 - r_{13}^3 f^2 f^3 t_2^2 \\
\beta_{16} = \theta_{16} = & t_2^3 f^2 f^3 r_{21}^2 - r_{21}^3 f^2 f^3 t_2^2 \\
\beta_{17} = \theta_{17} = & t_2^3 f^2 f^3 r_{22}^2 - r_{22}^3 f^2 f^3 t_2^2 \\
\beta_{18} = \theta_{18} = & t_2^3 f^2 f^3 r_{23}^2 - r_{23}^3 f^2 f^3 t_2^2
\end{cases}$$

where f^i denotes the focal length of the i^{th} vid camera; and t^i_j (r^i_j) denote the j^{th} translati (rotation) component associated with the i^{th} vid camera relatively to the reference coordinat system attached to the 1st video camera.

Equation 2.

II. BASIC EQUATIONS

II.1 Recalls about Epipolar Geometry

It is well known that in binocular vision, epipolar geometry (i.e. knowledge of the pick equipment parameters) allows to define, for a gi pixel in one of the two views, a set of pixels in other view which defines a line. Along this line, homologous pixel of the first one (i.e. correspond to the same 3D point) is located [2]. Generalizin a set of three images makes clear that fro couple of pixels, the localization of the third on the third view can be directly infered (see figure

II.2 Shashua's Equations

This geometry property has been modelled by Shashua [3,4]. The relation, in terms of trili constraints, can be expressed under four poss existing forms [5] (see Eq. 1).

The link between the four sets of parameters (α_i, γ_i, θ_i) on the one hand, and with the internal (f length) and external (translation and rotat calibration parameters of the pick-up equipmen the other hand is indicated on the left hand (see Eq. 2).
In the following sections, without loss of gener only the first form will be considered.

III. RECONSTRUCTION OF AN EXISTING POINT OF VIEW FROM TWO NEIGHBOURING OTHER VIEWS

III.1 Introduction

In this section, from three reference views, we propose and describe an effective algorithm, based on Shashua's equations which allows the reconstruction of one point of view from the two others (see figure 3). The algorithm can be divided into two parts (see figure 2) : an analysis stage (section II.2) and a synthesis stage (section II.3).

III.2 Analysis stage

In order to identify parameters (α_1,..., α_{18}), several significant sets of three pixels (p,p',p") corresponding to the same physical point are selected. This step is realized without any assumptions about the video acquizition conditions (no explicit calibration stage is required). Nevertheless, this stage can be seen as a pseudo-calibration stage. Pixels in the three views are matched using a modified optical flow algorithm [6]. This approach allows dense matching with subpixel precision. Then, according to a MSE criterion on luminance values, (α_1,..., α_{18}) parameters are estimated up to a scale factor (i.e. α_{12} is arbitrarily set to 1).

III.3 Synthesis stage

From the (α_1,..., α_{18}) parameters on the one hand and from the two other views on the other hand, which are matched in the same way as in the previous stage, the third view can be reconstructed using the following expression:

$$\begin{cases} x'' = -\dfrac{x'(\alpha_7 x + \alpha_8 y + \alpha_9) + \alpha_{10} x + \alpha_{11} y + \alpha_{12}}{(\alpha_1 x + \alpha_2 y + \alpha_3) + x'(\alpha_4 x + \alpha_5 y + \alpha_6)} \\ y'' = -\dfrac{x'(\alpha_{13} x + \alpha_{14} y + \alpha_{15}) + \alpha_{16} x + \alpha_{17} y + \alpha_{18}}{(\alpha_1 x + \alpha_2 y + \alpha_3) + x'(\alpha_4 x + \alpha_5 y + \alpha_6)} \end{cases}$$

Equation 3.

Several pre/post-processing techniques have been added in order to solve some problems such as the fact that some pixels receive no value or, on the contrary, two or more values.

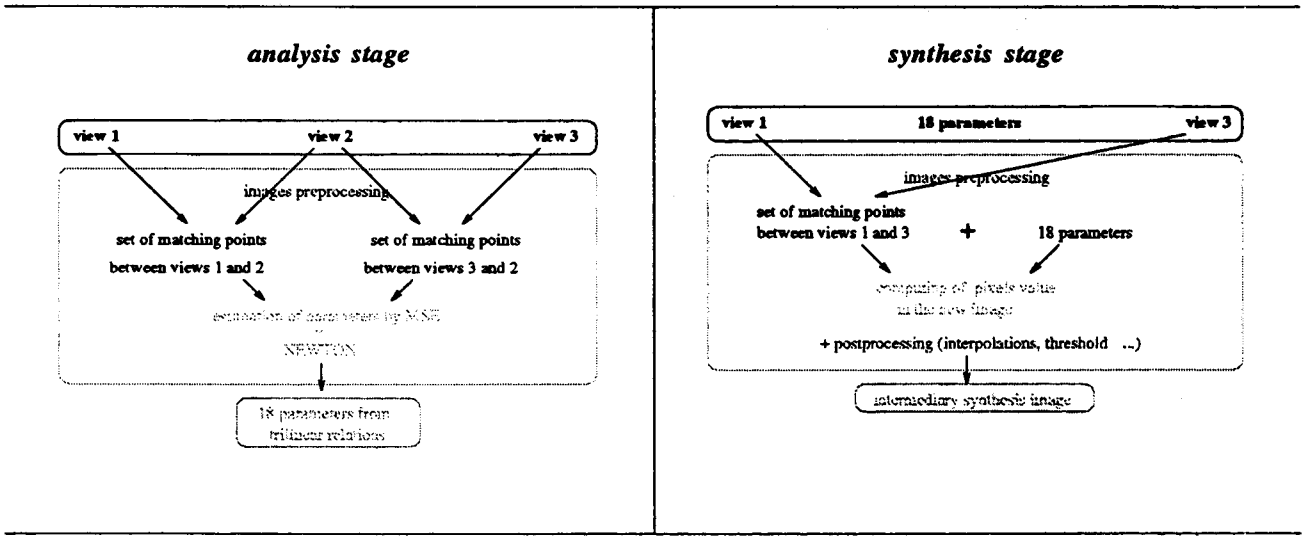


figure 2: Reconstruction of an existing view from two neighbouring views (flow chart)

IV. RECONSTRUCTION OF A VIRTUAL POINT OF VIEW FROM A SET OF THREE REAL REFERENCE VIEWS

Possible extensions are currently studied in order to reconstruct virtual points of view (i.e. other views than the three reference views used to define the set of pseudo-calibration parameters). This can be achieved by inverting the image indexes between the analysis stage and the synthesis stage and/or by modifying the set of parameter values just before the reconstruction stage [7,8].

We emphasize that this method uses neither a preliminary explicit 3D calibration stage nor an intermediate stage of depth map estimation of the scene before the reconstruction of a new view, including an in-between one [9]. For this reason, an

important issue inherent to this approach then consists in being able to interpret these 2 D parametric modifications in terms of 3D physical (i.e. real) transformations.

view 1

view 2

view 3

reconstructed view 1

figure 3. Reconstruction of an existing view from two neighboring views (result images)

At the synthesis stage, any of the three reference images can be reconstructed from the two other ones (i.e. not necessarily the intermediate one).

error of reconstruction

$$\begin{cases} x'' = -\dfrac{x'\left(\alpha_7^* x + \alpha_8^* y + \alpha_9^*\right) + \alpha_{10}^* x + \alpha_{11}^* y + \alpha_{12}^*}{\left(\alpha_1^* x + \alpha_2^* y + \alpha_3^*\right) + x'\left(\alpha_4^* x + \alpha_5^* y + \alpha_6^*\right)} \\ y'' = -\dfrac{x'\left(\alpha_{13}^* x + \alpha_{14}^* y + \alpha_{15}^*\right) + \alpha_{16}^* x + \alpha_{17}^* y + \alpha_{18}^*}{\left(\alpha_1^* x + \alpha_2^* y + \alpha_3^*\right) + x'\left(\alpha_4^* x + \alpha_5^* y + \alpha_6^*\right)} \end{cases}$$

where

$$\begin{cases} \alpha_i^* = \alpha_i & for\ i = 1...6 \\ \alpha_i^* = \lambda.\alpha_i & for\ i = 7...18 \end{cases}$$

Equation 4.

IV.1 Simulation of a focal length change

In this section, only the first form is used. Nevertheless, one can verify that starting from any form, the transformation to be applied on parameters to simulate a focus change is the same [10]. Nevertheless, according to the localization of the reference views, a given form can be more relevant than the others.

An unknown view, equivalent to a view which would have been obtained if the focal length had been different, can be simply obtained by modifying the focal length value. By substituting f^3 by $\lambda.f^3$ in Eq. 2, Eq. 3 becomes Eq. 4.
This way, another set of parameters (α^*_i) corresponding to a new (virtual) focal length can be directly derived from the original set of parameters (α_i).

IV.2 Simulations of other transformations

As seen in section IV.1, in addition to the reconstruction of an existing view, preliminary results are available to simulate some zooming action on points of view associated with one of the reference views (see figure 4).

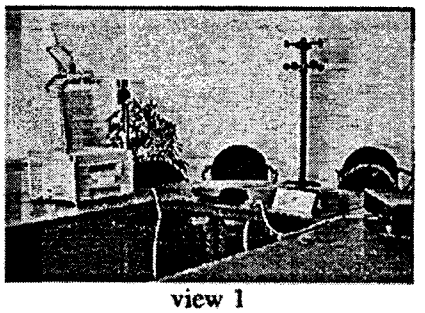
In order to access any point of view of the scene, studies on some other possibilities are currently in progress in order to reconstruct some translated or rotated points of view.

figure 4.

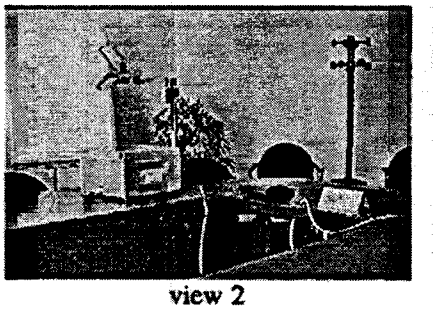
Reconstruction of view 3 from views 1 and 2, under different virtual focal length steps.

Due to the fact that no corresponding real view to the virtual one is available, the image has to be evaluated according to visual criteria.

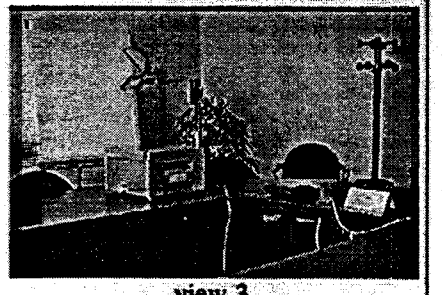
λ = 0.8



view 1



view 2



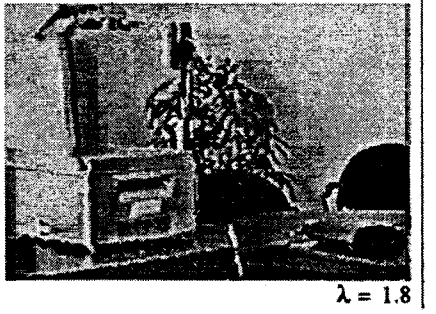
view 3

V. MORE THAN THREE REFERENCE VIEWS

The aim of virtual teleconferencing is to give several persons the opportunity of participating to a common remote meeting, with (nevertheless) the impression to be as much as possible physically in the same room. To this extend, the video spatialization of a meeting room has to be precise enough to offer the possibility for each virtual participant to visualize the room from a view point coherent with his gaze direction and/or the position he is assumed to have in the space. It is clear that, for this application, video spatialization will need more than three reference views. Current studies are in progress to define the number and the position of video cameras needed to achieve such a possibility. In the case of n (n >> 3) video cameras, a given point of view of the meeting room could be accessed using different ways (i.e. different sets of parameters and/or different sets of three video cameras), and will then probably lead us to introduce the notion of multi-trilinearity.

λ = 0.1.2

λ = 1.5



λ = 1.8

VI. CONCLUDING REMARKS

In this paper, a promising algorithm which allows to reconstruct an existing view from two neighbouring other views has been presented. Through this algorithm, some virtual view points, including virtual focal length changes, can also be reconstructed. Some further investigations are in progress to finely control other physical transformations (video camera virtual translations and/or rotations).

In order to obtain a complete range of meeting space overlapping views, algorithms developed in the limited context of a three video cameras set have to be extended to a larger number of video cameras.

Video spatialization, presented here for a virtual teleconferencing application [11], can more generally offer an effective alternative to artificial CAD models (which are more accurate but generally less realistic, unless some real textures are used to enhance the realism level [12]) and can be useful for some other applications, such as virtual house visiting or interactive television, that could enable viewers to choose the angle they would like to watch a basketball game under [13], without requiring the transmission of excessive additional informations.

REFERENCES

[1] J.-L. Dugelay, Traitements vidéo et espaces viruels de réunion, Conf. L'Interface des mondes réels et virtuels, Montpellier, France, Mai 1996.

[2] O. D. Faugeras, Quelques pas vers la visio artificielle en trois dimensions, Technique e Science Informatique, 1989.

[3] A. Shashua, Trilinearity in Visual Recognitio by Alignment, Proc. ECCV, Stockholm, Sweden May 1994.

[4] A. Shashua, Projective structure fro uncalibrated images: structure from motion an recognition, IEEE Trans. on PAMI, 1994.

[5] P. Bobet, J. Blanc & R. Mohr, Aspects caché de la trilinéarité. In Proc. Conf. RFIA'96, pp. 137 146, Rennes.

[6] J.-L. Dugelay and D. Pelé, Motion and Disparit Analysis of a Stereoscopic Sequence - Applicatio to 3DTV Coding -, Proc. EUSIPCO'92, Brussel Belgium, August 24-27, 1992.

[7] J. Blanc & R. Mohr, Calcul de vues de scèn 3D. Application à la compression vidé CORESA'95, Rennes.

[8] K. Fintzel & J.-L. Dugelay, Spatialisation vidé Conf. CORESA'96, Grenoble.

[9] G. Le Mestre and D. Pelé, Construction de cart de profondeur à partir de triplets d'image Application à la synthèse de points de v intermédiaires, Conf. CORESA'96, Grenoble.

[10] K. Fintzel & J.-L. Dugelay, Défocalisation spatialisation vidéo à partir de trois vues référence (Expressions analytiques), Intern Research Report no. RR-96-019.

[11] J. Ohya, Y. Kitamura, F. Kishino and Terashima, Virtual Space Teleconferencing: Re time reproduction of tridimensional human imag Journal of VCIR, 6 (1): 1-25, March 1995.

[12] P. Sander, ESSI VRnet Group, http://www.essi.fr/~sander/proj/EssiVR/EssiVR.htm

[13] T. Kanade, Virtualized Reality: Concepts a Early Results, IEEE Workshop on Representation Visual Scenes, June 1995, Massachusetts.