2D AND 3D MOTION ESTIMATION
FROM
BINOCULAR SEQUENCES
BY
COOPERATION BETWEEN MOTION AND DISPARITY
IN
3DTV CONTEXT

J-L. Dugelay [#], A. Tamtaoui [†], B. Choquet [‡], C. Labit [†]

‡ CCETT, 4 rue du Clos-Courtel, BP 59, 35512 Cesson-Sévigné
† IRISA/INRIA, Campus de Beaulieu 35042 Rennes Cedex
# EURECOM, 2229 rte des crêtes, Sophia Antipolis, 06560 Valbonne
FRANCE

This paper deals with the stereoscopic sequences analysis and compression by motion-disparity estimation-compensation. Works presented here principally concern the algorithm aspect of coding and are part of general research in 3-D television.

In a context of binocular sequences [Fig.1], such as stereoscopic television, there are two kinds of analysis in order to generate a coding scheme by estimation-compensation [Fig.2]. The first is called dynamic monocular analysis (or structure from motion, in 3D analysis) and consists in studying spatio-temporal changes of a single view. The second is called static binocular analysis (structure from stereo) and consists in studying projection differences between the two views at a given time.

Some recent studies try to combine these two approaches in a dynamic binocular analysis (rigid cameras with a moving scene). In this way, algorithms of motion estimation from binocular sequences, as well as in 2D analysis (i.e. : in the image plane) [1] as 3D analysis (i.e. : in 3D coordinates system) [2] have been developed and in part validated [3], [4]. We propose here a summary of these different methods and study the analytical links between them. First, we identify the common characteristics and then we underline the interest of each one.
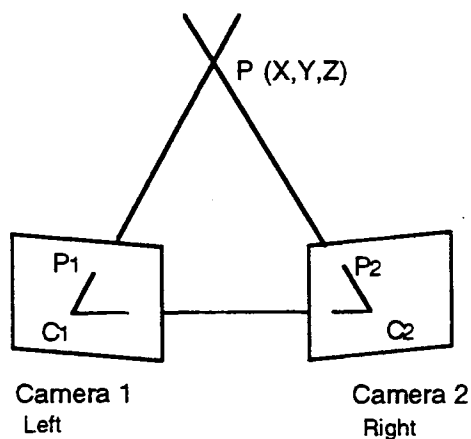

Fig.1

Fig.2

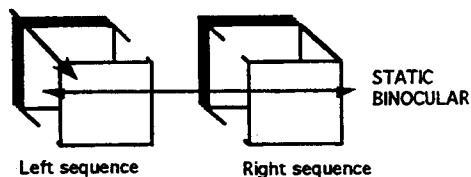# The author's contribution to this work was mainly performed while working on his thesis at CCETT.

A stereoscopic sequence is composed of two views, usually called left view and right view. The relative motion of an object in the scene (with regard to the pick-up equipment composed by two joint video cameras) perceived in each view is not the same because of the difference of the left and right video cameras' localisation. The link between the left motion with the right motion of a couple of homologous areas (i.e. : provided from the same physical entity's projection) can be established via the disparity (or static binocular analysis). More precisely, the link can be establised with the knowledge of the pick-up equipment calibration (i.e. : these parameters define on one hand the characteristics of the video cameras and on the other hand their localisation). The consideration of the binocular aspect of the studied sequences can be made at different levels :

In the 2D approach, the extrinsic parameters are considered before the identification stage of motion and enable coherency equations to be established from a relationship that has to verify the horizontal and vertical components of the left and right view apparent motions. In this appraoch, a joint estimation of the two movements is developed under constraints, issued from 3D modeling stage.

In the 3D approach, the motion estimation is independently realized in each view or in only one of the two views. The calibration parameters are then used, either to verify the coherency of the results between the two views or to deduce the motion parameters of one view from the other view parameters, by a change of marker, when the identification has been realized in only one of the two views.

In the two approaches, both the perspective projection aspect (from scene to image plane) and the relative localisation of the left and right video cameras are used in order to realize an estimation of the motion from binocular sequences. Although the two approaches use different primitives (2D or 3D), several problems are common, such as the simultaneous obtaining of motion and segmented maps (region based motion) or the matching of the left and right views at each instance of the sequence.

In fact, 2D and 3D approches mainly differ from the level in the analysis of the sequence where the passage between the scene and the image plane is taken into account. Nevertheless, one or other of the above approaches can be more or less adapted according to the objective (coding -and more precisely according to the possibility of an implementation in an effective coding scheme-, segmentation, interpretation, ...). Finally, we present some results in terms of maps of motion [5], of disparity [6] and of predicted images by compensation. We try to extract some common points of conclusion from these studies [1][2] for the motion and disparity analysis in the context of 3DTV.

Bibliography

[1] A. Tamtaoui, "Stereo-motion cooperation in view of stereoscopic sequences compression, for 3DTV application", Ph.D Thesis, n° 879 - IRISA - University of Rennes I - November 1992. [in french]

[2] J-L. Dugelay, "3D Motion Estimation and 3DTV Application", Ph.D Thesis, n° 893 - CCETT - University of Rennes I - November 1992. [in french]

[3] A. Tamtaoui and C. Labit, "Constrained disparity and motion estimators for 3DTV image sequence coding", Image Com., Special Issue on 3DTV, Vol. 4, no.1, Nov. 1991.

[4] All Race-II R2045 DISTIMA Partners, "State of the Art Report on Stereoscopic Imaging", R2045/SIE/SN6/DS/P/003/b1, 1992.

[5] A. M. Waxman and J. H. Ducan, "Binocular image flows : Steps toward stereo-motion vision", IEEE Trans. PAMI, Vol. 8, no. 6, pp. 715-729, November 1986.

[6] M. R. Jenkin, "Techniques for disparity measurement", CVGIP Image Understanding, Vol. 53, no. 1, January 1991.