A Coding Scheme for Stereoscopic Television Sequences based on Motion Estimation-Compensation using a 3D Approach

B. Choquet[†], J.-L. Dugelay[‡], D. Pelé[†]

[†] CCETT, Centre Commun d'Etudes de Télédiffusion et Télécommunications.
[‡] Institut EURECOM, FRANCE.

The implementation of a new television service (in this case stereoscopic television or 3DTV) requires research into four fundamental areas : pick-up equipment, transmission (including, bandwidth reduction), display and picture quality (1).
This paper focuses on algorithm aspects for the definition of an optimal coder well-suited to 3DTV. We present a 3D approach linking 2D left and right motion in order to define a 3D coding scheme based on motion estimation-compensation. This 3D scheme is more efficient than current 2D schemes because it allows for the possibility of using a unique channel for transmitting motion estimation-compensation information instead of two or three channels (2D left motion, 2D right motion, and eventually 2D disparity). A 3D approach could then offer new possibilities for increasing the compression rate (by about a factor 2 for additional motion information) by merging all 2D information in an unique set of 3D information.

The proposed algorithm takes the pick-up conditions used in 3DTV into account and consists of two steps: the first one is a dynamic monocular analysis of one the two views and is described in the next section. The second one is a static binocular analysis. Having performed steps 1 and 2, the analytical link between 3D left and right parameters (structure and motion) are establised. Some improvements to this scheme are envisaged, in the last section of this paper, by merging the two stages of this algorithm into a unique stage of dynamic binocular analysis.

FIRST STEP:DYNAMIC MONOCULAR ANALYSIS

The first step is a 3D dynamic monocular analysis stage. This step could be conducted either in the left or in the right sequence.
The movement of an object in the scene, with respect to the camera, expressed in a three-dimensional coordinates system, may be divided into two components, an instantaneous translation rate $T(t_x,t_y,t_z)$ and an instantaneous rotation rate $W(w_x,w_y,w_z)$.
Let p_i be the picture element (camera plane) associated with $P_i(X_i,Y_i,Z_i)$, with coordinates (x_i,y_i) and displacements (u_i,v_i).
Using known hypotheses (2), the equations expressing the link between the apparent motion in the image plane and the real motion of the object are:

$$u_i = +y_i.w_z - (1+(x_i)^2).w_y + x_i y_i.w_x + (1 / Z_i).t_x - (x_i / Z_i).t_z$$
$$v_i = -x_i.w_z + (1+(y_i)^2).w_x + x_i y_i.w_y + (1 / Z_i).t_y - (y_i / Z_i).t_z \qquad (1)$$

In the case of planar surfaces (3), the depth Z_i can be expressed as $a.X_i+b.Y_i+c$, where (X_i,Y_i,Z_i) are the coordinates in the 3-D space. By projection in the image plane, $1 / Z_i$ can be expressed as

$$k_x.x_i+k_y.y_i+k_z. \qquad (2)$$

If we substitute $1 / Z_i$ by its expression into the basic equations of 3D motion (eq.1), we can rewrite the preceeding system into two systems. The displacement (u_i,v_i) is given by,

$$u_i = a_1 + a_2.x_i + a_3.y_i + a_7.x_i y_i + a_8.x_i^2$$
$$v_i = a_4 + a_5.x_i + a_6.y_i + a_8.x_i y_i + a_7.y_i^2 \qquad (3)$$

Image Processing And Its Applications, 4–6 July 1995
Conference Publication No. 410, © IEE 1995.

$$\text{with, } \begin{cases} a_1 = -w_y + k_z.t_x \\ a_2 = k_x.t_x + -k_z.t_z \\ a_3 = w_z + k_y.t_x \\ a_4 = w_x + k_z.t_y \\ a_5 = -w_z + k_x.t_y \\ a_6 = k_y.t_y + -k_z.t_z \\ a_7 = w_x + -k_y.t_z \\ a_8 = -w_y + -k_x.t_z \end{cases} \qquad (4)$$

The algorithm jointly performs the segmentation of the images into a set of arbitrary shaped areas which are supposed to correspond to a planar facet model, and the estimation of associated parameters. This stage is described in (4). It is well-known that at the end of this step, there remains an unknown scale factor between the structure (depth) and translation (amplitude). So, only the following parameters may be computed at this stage,

$$\begin{cases} t'_x = k_z.t_x \\ t'_y = k_z.t_y \\ t'_z = k_z.t_z \\ k'_x = k_x / k_z \\ k'_y = k_y / k_z \end{cases} \qquad (5)$$

SECOND STEP:STATIC BINOCULAR ANALYSIS

The pick-up conditions in the context of 3DTV are the following (5): The system is composed of two video cameras. The left and the right video cameras are separated by an angle α (convergence angle) and by an horizontal shift B (baseline). The internal parameters of the video cameras are assumed known. Relative calibration defined by the 3D rotation and the translation matrices W^{stereo} and T^{stereo} (to get from one video camera to the other one) is given by:

$$W^{stereo} \cong \begin{pmatrix} 1 & 0 & -\alpha \\ 0 & 1 & 0 \\ \alpha & 0 & 1 \end{pmatrix} \text{ and } T^{stereo} = \begin{pmatrix} B \\ 0 \\ 0 \end{pmatrix} \qquad (6)$$

Due to the different locations of video camera (6), the apparent 2D motion of P is not the same in the left image plane (i.e motion of P^{left}) and in the right image plane (i.e motion of P^{right}). For this reason, classical coding schemes for stereoscopic sequences based on 2D motion estimation-compensation are obliged to use two independent channels: one for the left view and one other for the right view.

In this second step, left and right sequences are matched according structure criteria. The matching stage is realized by using a prunning tree approach (7). This step yields the left and right areas associated to the same physical object, as well as the unknown depth factor k_z mentioned above.

Disparity can be considered as an optical flow resulting from moving one camera from one position to the other. As soon as the relative calibration is known and the planar facet model for Z is maintained as for monocular analysis, disparity can be described by a simplified quadratic model. Letting $d_x(x_i,y_i)$ and $d_y(x_i,y_i)$ be the disparity values of pixel (x_i,y_i) between the left and right views measured in the coordinates system attached to one the two views, we have that

$$\begin{cases} d_x(x_i,y_i) = a+b.x_i+c.y_i+d.x_i^2 \\ d_y(x_i,y_i) = d.x_iy_i \end{cases} \qquad (7)$$

where (a,b,c,d) are expressed in terms of the relative calibration parameters (convergence angle α and baseline B) and structure (k_x, k_y, k_z) as follows

$$\begin{cases} a = k_z.B - \alpha \\ b = k_x.B \\ c = k_y.B \\ d = -\alpha \end{cases} \qquad (8)$$

or, equally with:

$$\begin{cases} a = k_z.B - \alpha \\ b = k_z.(k_x/k_z).B \\ c = k_z.(k_y/k_z).B \\ d = -\alpha \end{cases} \qquad (9)$$

For each region determined by the dynamic monocular analysis of the right view (we assume here that this view has been chosen as the reference view), we calculate the vector P(a,b,c,d) using a differential method which minimizes the quadratic prediction error by compensation of disparities from the right view towards the left view. This is done in a similar way to the identification of the parametric vector $A(a_1,\dots,a_8)$ in the previous part. If α and B are known from the calibration stage, and if (k'_x,k'_y) were calculated in the monocular dynamic analysis stage, the estimation of the model parameters of disparity (a, b, c, d) yields the depth k_z, by the following expression,

$$\begin{aligned}k_z = \{&(a + \alpha) / B \\ &+ x_g^2.\{ [b / k'_x.B] \\ &+ y_g^2.\{ [c / k'_y.B]\} \\ &/ \{(1 + x_g^2 + y_g^2)\}\end{aligned} \tag{10}$$

where (x_g,y_g) is the centre of gravity of the region studied. The system (eq.9) is weighted in this way in order to take into account the term degree in which each element of the disparity vector P in (eq.7) is found and the removal of the zone considered at the image centre.

At the end of this second stage, we have completely determined the 3D apparent motion and structure parameters. These parameters after projection in image planes could be used for motion compensation for either sequence by using analytical link between 3D left and right motion and structure parameters.

ANALYTICAL LINK BETWEEN 3D LEFT AND RIGHT PARAMETERS

With an absolute coordinate system, for any point P, belonging to a facet of the scene animated by a movement (W,T), and accompanying a new position P' in the following instant, we have the relationship :

$$P' = W.P + T, \tag{11}$$

$$\text{with } W = \begin{pmatrix} 1 & w_z & -w_y \\ -w_z & 1 & w_x \\ w_y & -w_x & 1 \end{pmatrix}$$

$$\text{and } T = \begin{pmatrix} t_x \\ t_y \\ t_z \end{pmatrix}$$

Measured in the coordinate system attached to the left video camera, (eq.11) rewritten :

$$p'^{left} = W^{left}.p^{left} + T^{left}, \tag{12}$$

$$\text{with } W^l = \begin{pmatrix} 1 & w_z{}^l & -w_y{}^l \\ -w_z{}^l & 1 & w_x{}^l \\ w_y{}^l & -w_x{}^l & 1 \end{pmatrix}$$

$$\text{and } T^l = \begin{pmatrix} t_x{}^l \\ t_y{}^l \\ t_z{}^l \end{pmatrix}$$

Measured in the coordinate system attached to the right video camera, equation (eq.11) rewritten :

$$p'^{right} = W^{right}.p^{right} + T^{right}, \tag{13}$$

$$\text{with } W^r = \begin{pmatrix} 1 & w_z{}^r & -w_y{}^r \\ -w_z{}^r & 1 & w_x{}^r \\ w_y{}^r & -w_x{}^r & 1 \end{pmatrix}$$

$$\text{and } T^r = \begin{pmatrix} t_x{}^r \\ t_y{}^r \\ t_z{}^r \end{pmatrix}$$

Moreover, with instants t and t':

$$P^r = W^s.P^l + T^s \tag{14}$$

and

$$P'^r = W'^s.P'^l + T'^s \tag{15}$$

We assume here that the extrinsic calibration parameters remain unchanged between the times t and t' (i.e. : $W'^s = W^s$ et $T'^s = T^s$), equation (eq. 15) rewritten :

$$P'^r = W^s.P'^l + T^s \tag{16}$$

From the relationships 12, 14 and 16, we can therefore establish the following relationship :

$$\begin{aligned}W^s.P'^l + T^s &= P'^r \\ &= W^r.P^r + T^r \\ &= W^r.(W^s.P^l + T^s) + T^r\end{aligned}$$

that is to say,

$$P'^l = W^{-1s}.[W^r.(W^s.P^l + T^s) + T^r - T^s] \tag{17}$$

with equation (eq. 12), we obtain :

$$\begin{aligned}\mathbf{W^l}.P^l + \mathbf{T^l} = &\ \mathbf{(W^{-1s}.W^r.W^s)}P^l \\ &+\mathbf{W^{-1s}.[\ (W^r - Id_3).T^s + T^r]}\end{aligned} \tag{18}$$

Rotation components. From (eq. 18), we can deduce that the rotation components verify,

$$\mathbf{W^{l} = W^{-1s}.W^{r}.W^{s}},$$

that is to say :

$$\begin{pmatrix} w_x{}^{l} \\ w_y{}^{l} \\ w_z{}^{l} \end{pmatrix} = \begin{pmatrix} w_x{}^{r} \\ (1-\alpha^2).w_y{}^{r} \\ -\alpha.w_x{}^{r} + w_z{}^{r} \end{pmatrix} \tag{19}$$

Translation components. Those of the translation verify the relationship :

$$\mathbf{T^{l} = W^{-1s}.W^{r}.T^{s} + W^{-1s}.T^{r} - W^{-1s}.T^{s}},$$

that is to say :

$$\begin{pmatrix} t_x{}^{l} \\ t_y{}^{l} \\ t_z{}^{l} \end{pmatrix} = \begin{pmatrix} t_x{}^{r} + \alpha.[t_z{}^{r} + B.w_y{}^{r}] \\ t_y{}^{r} - B.w_z{}^{r} \\ t_z{}^{r} + B.w_y{}^{r} - \alpha.t_x{}^{r} \end{pmatrix} \tag{20}$$

Structure components. In a way similar to that carried out for the motion, we can establish from the expressions:

$$\begin{cases} 1 / Z^{l} = k_z{}^{l} + k_y{}^{l}.y^{l} + k_x{}^{l}.x^{l} \\ 1 / Z^{r} = k_z{}^{r} + k_y{}^{r}.y^{r} + k_x{}^{r}.x^{r} \end{cases} \tag{21}$$

the following relationships :

$$\begin{pmatrix} k_x{}^{l} \\ k_y{}^{l} \\ k_z{}^{l} \end{pmatrix} = \begin{pmatrix} (\alpha.k_z{}^{r} + k_x{}^{r}) / (1 - B.k_x{}^{r}) \\ k_y{}^{r} / (1 - B.k_x{}^{r}) \\ (k_z{}^{r} - \alpha.k_x{}^{r}) / (1 - B.k_x{}^{r}) \end{pmatrix} \tag{22}$$

These relations between left and right structure components have been establised as follow

From

$$1 / Z^{r} = k_z{}^{r} + k_y{}^{r}.y^{r} + k_x{}^{r}.x^{r},$$

$$\begin{cases} X^{r} = X^{l} - \alpha.Z^{l} + B \\ Y^{r} = Y^{l} \\ Z^{r} = \alpha.X^{l} + Z^{l} \end{cases}$$

and

$$\begin{cases} x^{r} = X^{r} / Z^{r} \\ y^{r} = Y^{r} / Z^{r} \end{cases}$$

we can write :

$$1/Z^{r} = k_z{}^{r} + k_y{}^{r}.[(Y^{l} / Z^{r})] + k_x{}^{r}.[(X^{l} / Z^{r}) - \alpha.(Z^{l} / Z^{r}) + B/Z^{r}]$$

That is to say, multiplying by Z^{r},

$$1 = Z^{r}.k_z{}^{r} + k_y{}^{r}.(Y^{l}) + k_x{}^{r}.(X^{l} - \alpha.Z^{l} + B)$$

By substituting Z^{r} by $(\alpha.X^{l} + Z^{l})$, we obtain :

$$1 = k_z{}^{r}.(\alpha.X^{l} + Z^{l}) + k_y{}^{r}.(Y^{l}) + k_x{}^{r}.(X^{l} - \alpha.Z^{l} + B)$$

By making the term $1 / Z^{l}$ appear,

$$1 / Z^{l} = k_z{}^{r}.(a.x^{l} + 1) + k_y{}^{r}.(y^{l}) + k_x{}^{r}.(x^{l} - a + B / Z^{l})$$

that is to say,

$$1 / Z^{l}\ (1 - k_x{}^{r}.B) = k_z{}^{r}.(\alpha.x^{l} + 1) + k_y{}^{r}.(y^{l}) + k_x{}^{r}.(x^{l} - \alpha)$$

By factoring by x^{l} and y^{l} :

$$1 / Z^{l}\ (1 - k_x{}^{r}.B) = (k_z{}^{r} - \alpha.k_x{}^{r}) + (k_y{}^{r}).y^{l} + (\alpha.k_z{}^{r} + k_x{}^{r}).x^{l}$$

By identifying the preceeding expression with $1 / Z^{l} = k_z{}^{l} + k_y{}^{l}.y^{l} + k_x{}^{l}.x^{l}$, we obtain the relationship (22).

CONCLUSION AND PERSPECTIVES

CONCLUSION

From these developments, a complete coding-decoding algorithm can be designed as:

coding stage.

1. Computing by dynamic monocular analysis right 3D parameters ($t'_x{}^{r}$, $t'_y{}^{r}$, $t'_z{}^{r}$, $k'_x{}^{r}$, $k'_y{}^{r}$) and associated segmentation.
2. Computing by static binocular analysis for each area of the right view: $k_z{}^{r}$ and then ($t_x{}^{r}$, $t_y{}^{r}$, $t_z{}^{r}$, $k_x{}^{r}$, $k_y{}^{r}$, $k_z{}^{r}$).

transmission.

Intrinsic parameters (α and B), spatial description and location with associated parameters of structure and motion for each area of the right view must be transmitted. Morover, the result of spatial matching between left and right views must be done in order to know, during the decoding stage, to which area in the other view corresponds a given area.

Decoding stage.

For each right area, compute 2D motion (i.e. $(a_1{}^r, \ldots, a_8{}^r)$ for each area, from (eq.4), and then (u^r, v^r) for each pixel, from (eq.1)) from 3D parameters for excuting the motion compensation procedure for this view. For the homologous area of the left view, compute with eq. (19), (20) and (22) 3D parameters, then 2D left parameters.

Results.

Preliminary results obtained using this algorithm allow for the computation of the left optical flow from the right optical flow and calibration parameters. The obtained left optical flow is very close to the original left optical flow (i.e. directly computed on the view itself).

FUTURE WORK

Future work is oriented toward cooperation between stages 1 and 2 of the algorithm, in order to simultanously realize both dynamic monocular analysis and static binocular analysis in a unique stage of dynamic binocular analysis (8, 9). This could be improve the quality and coherency of the segmentation left and right, using a symmetric matching between the two views.
Another point concerns modeling for object structure representation. A more complete model than planar facet such as curved facet (10) could be used in order to reduce the number of areas identified during the segmentation stage.

ACKNOWLEGMENTS

This work has been partly realized in the context of the European project RACE 2045 "DISTIMA" (Digital Stereoscopic Imaging and Applications).
Thanks to R. Knopp for reviewing contents of this paper.

REFERENCES

1. F. Chassaing, B. Choquet and D. Pelé, 1991, "A stereoscopic television system (3D-TV) and compatible transmission on a MAC channel (3D-MAC)", Signal Processing: Image Communication, pp. 33-43.
2. G. Adiv, 1985, "Determining three-dimentional motion and structure from optical flow generated by several moving objects", IEEE, Vol. PAMI-7, no. 4.
3. R.Y. Tsai and T.S. Huang, 1981, "Estimating Three-dimensional Motion Parameters of a Rigid Planar Patch, IEEE Trans. on ASSP, Vol. 29, no.6.
4. J.-L. Dugelay & H. Sanson, 1995, "Differential methods for the identification of 2D and 3D motion models in image sequences", to appear in Signal Processing: Image Communication.
5. D. Pelé and A. Kopernik, 1993, "Disparity estimation for stereo compensated 3DTV coding", PCS'93, Lausanne, section 16.5.
6. A. Tamtaoui and C. Labit, 1991, "Constrained disparity and motion estimators for 3DTV image sequence coding", Signal Processing: Image Communication, pp. 45-54.
7. A. Gagalowich and L. Vinet, 1989, "Region Matching for Stereo Pairs", the 6th SCIA, Oulu, Finland, PP. 63-70.
8. W. Richards, 1985, "Structure from stereo and motion", J. Opt. Soc. Am., pp.343-349.
9. A.M. Waxman and J.H. Ducan, 1986, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion", IEEE Trans. on PAMI, Vol.8, no.6.
10. M. Subbarao, 1988, "Interpretation of Image Flow : Rigid curved Surfaces in Motion", International Journal of Computer Vision, pp. 77-96.