

AN EFFICIENT AND DIRECT NON-PLANAR ROTATION ESTIMATION ALGORITHM FOR VIDEO APPLICATIONS

M. García and H. Nicolas

IRISA/INRIA, Campus de Beaulieu
35042 Rennes Cedex, France
e-mail:mgarciav,hnicolas@irisa.fr

Abstract

We present a new method for the estimation of non-planar rotations, i.e. rotations around axis parallel to the image plane, in the context of video compression applications. This method is based on a non planar rotation model which assumes that the moving objects have a planar surface. The proposed block-based motion estimation approach is performed between consecutive or non-consecutive images, which may contained large displacements, and aims at minimizing the motion compensation error. The efficiency of the method has been compared to the results obtained with the classical full search block matching approach. Experimental results have been done on real video sequences. These results show a significant gain in term of PSNR for the motion compensated P or B frames, compared to the classical full search block matching approach, while the coding cost of the additional motion information is very low, which demonstrates the interest of the proposed rotation model in the context of motion compensation for video compression applications.

Keywords: Non-planar rotation estimation, motion estimation.

1. INTRODUCTION

Motion estimation (ME) has proven to be effective to exploit the temporal redundancy of video sequences and is therefore a central part of the ISO/IEC MPEG-1, MPEG-2, MPEG-4 and the CCIT H.261 / ITU-T H.263, H.26L [1] video compression encoder algorithms. These video compression standards are based on a block based hybrid coding concept, which was (among other improvements) extended to support arbitrarily shaped video object within MPEG-4 [2] [3] [4].

Motion estimation algorithm have attracted some attention within research and industry because of these reasons:

- First, it is the computational most demanding algorithm of a video encoder (about 60-80% of the total commutation time) which limits the performance of the encoder in terms of encoding speed.
- Second, the motion estimation algorithm has a high impact on the visual quality of an encoder for a given bit-rate.
- Finally, the method to extract motion vectors from the video material is not standardized, thus being open to competition.

It is important to indicate that the motion estimation algorithms in these compression standards allow only the use of a translational motion model. As a consequence, the block-based motion estimation can theoretically compensate only 2-D translational displacements. In practice, it may be possible to compensate more complex motions, such as zoom or rotations, only if the amplitude of the motion is low or if the texture is homogeneous in the considered block. Many papers have proposed and developed efficient algorithms for an efficient estimation of these translational parameters. Nevertheless, if the images contain large non-translational motion in textured areas, the motion compensation process may not be efficient.

In order to obtain an efficient motion compensation in areas containing non-translational displacements, different solutions have been investigated. A first approach involves in reducing the size of the blocks. This case is exemplified in the under development H26L [2] compression scheme where the block size may be reduced down to 4x4 pixels. A second approach involves the use of more complex motion models. Classically, the use of an affine model allows an efficient compensation of motion such as zoom or 2D rotations. Nevertheless, Non Planar Rotation (NPR) [5], i.e. rotations around an axis parallel to the image plane are not taken into account by such a model. Alternatively, other methods such as the control grid interpolation [6] or geometric transformation motion estimation [7] have also been developed. They are based on a warping process which allows the distortion of each block in order to warp it onto the reference picture. If any kind of distortion of the blocks may be allowed, this model does not provide explicit NPR modeling.

In this paper, we propose a model of NPR which allows a better motion compensation efficiency when this kind of motion occurs. This model contains four motion parameters: two translational ones, one angle which defines the rotation axis, and the rotation angle.

2. NON PLANAR MOTION MODELING

The 2D apparent displacement of video objects in an original video sequence is generated by the relative 3D object/camera displacement [5]. The two following main categories of motion can be distinguished according to a variation of orientation criterion.

Planar Motion. Planar motions represent the relative camera/object displacements which do not modify the relative orientation object/camera. They corresponds to a combination of translation (low amplitude compared to the distance camera/object), divergence or rotation around an axis parallel to the image plane. Under reasonable assumptions such kinds of motion can be represented using affine or homographic motion models.

Non-planar Motion. Non-planar motions correspond to rotations around axis parallel to the image plane, to large amplitude (compared to the object-camera distance) translation, and eventually combined with a planar motion. With the exception of particular configurations (when the object can be regarded as plane or when the optical center of the camera is fixed relatively to the scene), this type of motions cannot be represented using a simple model such as the affine or homographic ones, often use in classical motion estimation algorithms.

In the context of block-based coding applications, the geometric model of the scene can be considered as a patchwork of rigid planar surfaces, one for each block, which can closely approximate 3D rigid bodies. Under this assumption, the 3D motion can be described by a rigid 3D motion model. This hypothesis is justified by the fact that the blocks are usually relatively small. Consequently, a representation of the 2D

motion in the image plane can easily be derived by the projection of the 3D motion. Mathematically, for each point P of an object, its position at time t' can be expressed as:

$$\vec{P}' = R\vec{P} + \vec{D} \quad (1)$$

$\vec{D} = (D_x, D_y, D_z)^T$ is the translational displacement, and R denotes the rotation matrix which is defined as:

$$R = R_x R_y R_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & \sin \alpha \\ 0 & -\sin \alpha & \cos \alpha \end{bmatrix} \cdot \begin{bmatrix} \cos \beta & 0 & -\sin \beta \\ 0 & 1 & 0 \\ \sin \beta & 0 & \cos \beta \end{bmatrix} \begin{bmatrix} \cos \gamma & \sin \gamma & 0 \\ -\sin \gamma & \cos \gamma & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

where R_x, R_y, R_z are the rotation matrix around X, Y, and Z axis, respectively, and α, β and γ the corresponding rotation angles. R_z represents the 2D rotations, while R_x and R_y are the non-planar rotations. In practice and for video compression purposes, it may be sufficient to consider only one rotation. If only the 2D rotation R_z is considered, the model leads to the rotation parameter defines by the affine model. If we consider only one of the two non-planar rotations, and assuming that inside a block, the object is plane and a perspective projection for the camera from the point of view of the projection plane:

$$x = f \frac{X}{Z+f} \quad \text{and} \quad y = f \frac{Y}{Z+f}$$

where f denotes the focal length, and (x,y) the coordinates of point $P(X,Y,Z)$ in the image plane, the projection of Eq. (1) in the image plane leads to:

$$x_2 - x_{g2} = \frac{(x_1 - x_{g1}) \cos \alpha}{1 - \frac{\sin \alpha}{f} (x_1 - x_{g1})}$$

$$y_2 - y_{g2} = \frac{(y_1 - y_{g1})}{1 - \frac{\sin \alpha}{f} (x_1 - x_{g1})}$$

where the coordinates (x_1, y_1) and (x_2, y_2) represents the position of each pixel at time t and t+1 respectively, (x_g, y_g) denotes the gravity center of a block and as demonstrated in [8], the camera focal f can be expressed as:

$$f = \frac{CCD \text{ sensor width (pixels)}/2}{\tan(\text{horizontal field of view}/2)},$$

where the horizontal field of view is approximately equal to 50°. It should be pointed out that the use of the gravity center as a reference point means that this gravity center is considered to be located on the rotation axis. If it is not the case, a translation term needs to be added.

This representation can easily be generalized to any axis Φ parallel to the image plane as follows:

$$\begin{pmatrix} x_2(\phi) \\ y_2(\phi) \end{pmatrix} = \begin{pmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix}$$

where $x_2(\phi)$ and $y_2(\phi)$ denotes the coordinates of pixel (x_2, y_2) in the coordinate system (Φ, Φ^\perp) , and ϕ is the angle between the X and Φ axis. As a consequence, the displacement can be computed if the two translational terms, the axis Φ and the rotation angle α are estimated.

3. MOTION ESTIMATION

In order to validate the non planar rotation model proposed in the previous section, a comparison of this model with the classical translational one has been done. The translational parameters can be estimated using a full search block matching algorithm in order to obtain the best possible result, in term of minimization of the Mean Square Reconstruction Error (MSRE). For the non-planar motion estimation model, four parameters have to be estimated. It is therefore not reasonable, from a computational complexity point of view, to perform a full search on these four parameters. A sub-optimal approach has therefore to be defined. Furthermore, an efficient estimation of the rotation parameters ϕ and α can be obtained only if the translational parameters have been previously obtained. This is due to the fact that the rotation is arbitrary considered to rotate around the block gravity center. As a consequence, the estimation method proposed here is performed in the three following steps:

Step 1. Rough estimation of the translational parameters using a full search block matching algorithm. This first estimation is performed on a sub-sampled image (by a factor of 2 in each direction) in order to get a rough and fast estimation of the translational parameters. The goal is to get a rough match between the block which should be predicted and the reference image in order to allow a correct estimation of the rotation parameters.

Step 2. Rough estimation of the two rotational parameters using a full search method. The angles precision is fixed to an angle step of 5° in order to achieve a fast estimation.

Step 3. Refinement stage. Once a first estimation for the four parameters have been obtained with the two first stages, a refinement stage is used to get a more precise estimation. A full search is therefore performed on the four parameters with a maximum value of 4 for the translational parameters, and 5° for the rotation angles. The final precision is fixed to a half-pixel for the translation parameters, and 1° for the angles.

Furthermore, this motion estimation phase is embedded in a GOP-based compression algorithm at it is usually done in MPEG and H.26X video compression standards. This means that the motion estimation is performed only between two successive I or P frames, while the intermediate frames, the B ones, are predicted using previous I or P frames (see Figure 1). This prediction is done here using a bi-directional motion compensation technique based on the NPR model. The encoder then computes the two predictions obtained with the previous and next I or P frame respectively. Decision on whether the forward or backward prediction is retained is based on the minimization of the motion compensated error. As can be seen in Fig 3. the quality (in term of PSNR) of the bi-directionally motion compensated image B with the NPR model, is always superior to the quality obtained with the BM model.

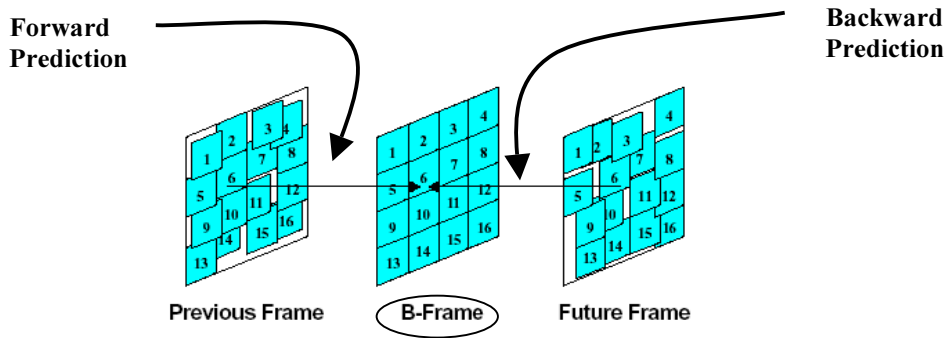


Figure 1: Forward and backward prediction

4. EXPERIMENTAL RESULTS

Experimental tests were performed in order to assess the performance of the presented method in sequences containing non-planar rotations. Figure 2 shows some original frames from the tested sequences: “*Tai*”, “*Foreman*” and “*Car*” sequences (CIF format). In the *Tai* sequence the head has a non-planar rotation of around 180 degrees during the sequence. In the *Foreman* sequence, the camera has a panoramic displacement, which generates a non-planar rotation of the scene relatively to the camera. The *Car* sequence shows a rigid non-planar rotation for the car composed with a moving camera (translational + zoom).

Experiments were carried out using 16x16 blocks and 8x8 blocks. The maximal search range was set to ± 16 pixels for the translational motion parameters and ± 40 degrees for the rotation angles. Figure 3 shows the gain in term of PSNR obtained for the B frames. Figure 4.a shows the blocks for which a gain of at least 1db is obtained with the NPR model and with the bi-directional MC compared to the use of the translational model. The quality improvement is shown on the images displayed in Figure 4.b. In a general way, a gain of around 1dB is obtained on rotating and textured blocks for which the translation model are not efficient. At the opposite, the gain is obviously very low for blocks in which the quality prediction is very high.

As a consequence, it is not interesting to test the NPR model for blocks well predicted with the translational model. In practice, the NPR model is therefore tested only in the blocks for which the PSNR obtained with the translational model is lower than 40 dB. This represents typically between 30% and 50% of the blocks, which allows a significant reduction of the computational complexity.

Furthermore, for each block, the NPR is validated only if the gain, in term of PSNR reduction, is lower than a predefined threshold. In term of coding cost, the overhead generated by the model is therefore reduced. Two rotational parameters have to be coded for each NPR block, and a flag must indicate to the decoder which model has been used. For 16x16 blocks, if this information is entropy coded, it represents less than 0.01 bit/pixels.

4. CONCLUSION

This paper proposes a four parameters model of non-planar rotation and its use in the context of motion compensation for video compression applications. It has been shown that a significant gain can be obtained, in term of PSNR for P and B frames, compared to the use of a translation model in sequences containing non planar moving objects, or having a rotating camera. One of the main perspective of this work is the use of a

block-based adaptive motion model representation, including translation, non-planar rotation and affine models, to improve the motion compensation process.



Figure 2. Original frames from sequences. Left to right: Tai(a), Foreman(b) and Car (c).

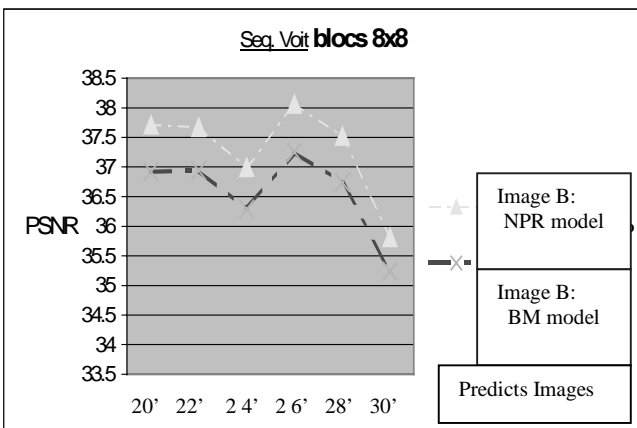
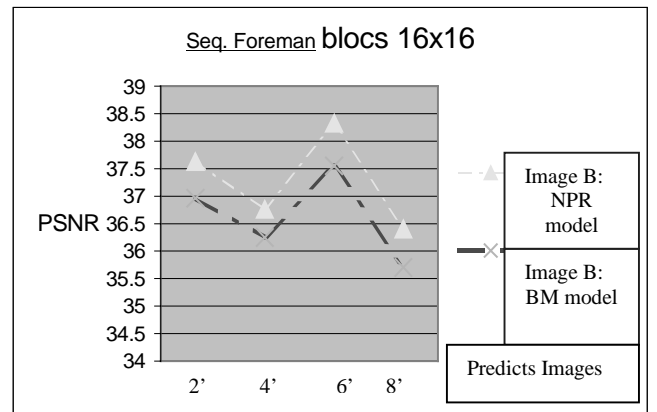
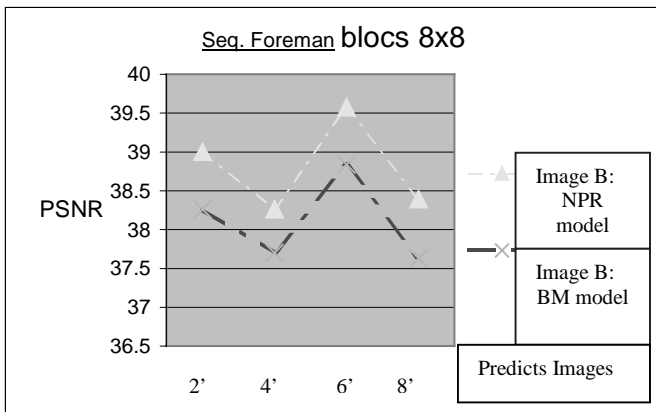
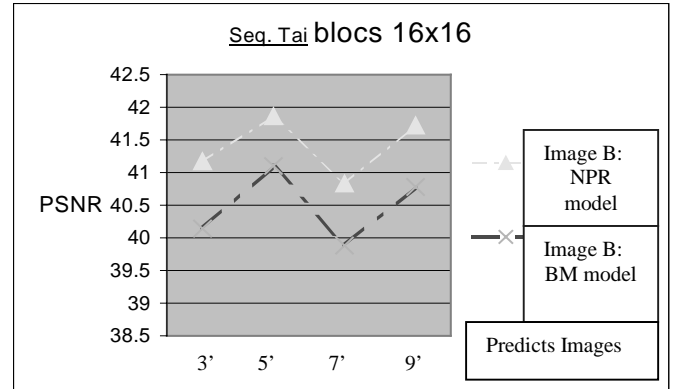
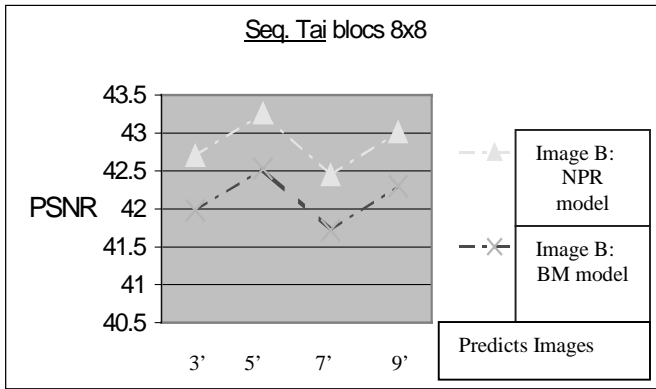
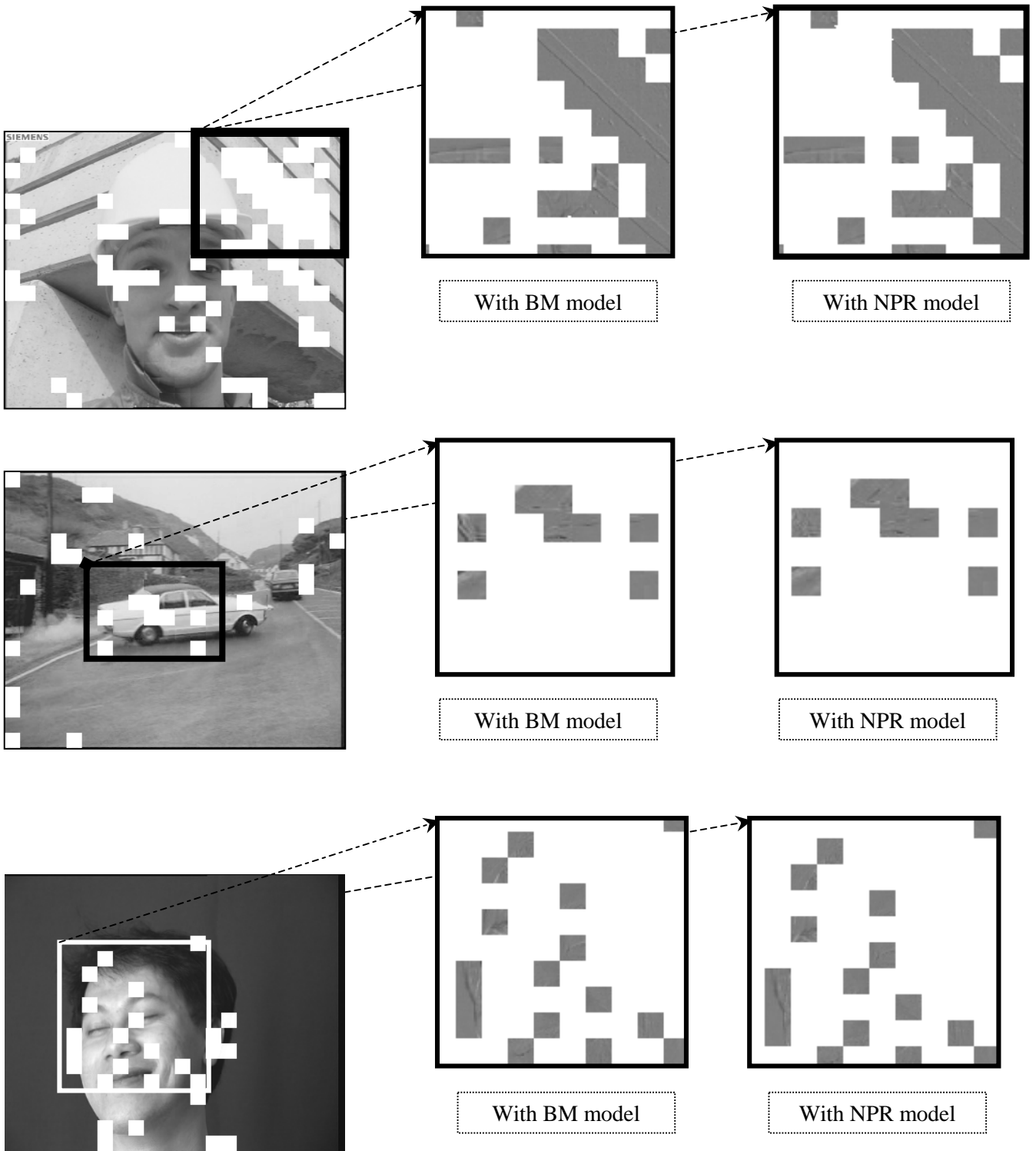


Figure 3. PSNR improvement of bi-directional MC with NPR model versus BM model. A gain of 0.5 to 1.3 dB is obtained with the NPR model compared to the translational one.



(a)

(b)

Figure 4.(a) The blocks for which a gain of at least 1db is obtained with the NPR model . (b) The quality improvement is shown in (b).

5. REFERENCES

- [1] <http://www.cselt.it/mpeg>.
- [2] www.ubvideo.com/public/h261-white_paper.pdf
- [3] T. Sikora: "The MPEG-4 Video Standard Verification Model", IEEE Trans. Circuits and Systems for Video Technology, Vol. 7, No.1, pp 19-31, Feb 1997.
- [4] T. Sikora: "MPEG Digital Video Coding Standards", IEEE Signal Processing Magazine, Vol. 14, No.5, pp 82-100, Sept. 1997.
- [5] M. García and H. Nicolas. Video Object Trajectory Analysis. Proceedings of ICIP 2002, Rochester, USA.
- [6] G.J. Sullivan & R. L. Baker, "Motion compensation for video compression using video grid interpolation", Proc. of ICASSP 91, pp.2713-2716, Toronto, May. 1991.
- [7] S. Maciel de Faria, Very low bit rate video coding using geometric transform motion compensation, thesis, Univ. of Essex 1996.
- [8] D. Scharstein. View Synthesis Using Stereo Vision. Lecture Notes in Computer Science (LNCS), volume 1583. Springer Verlag, 1999.