

MOTION ESTIMATION AND IMAGE WARPING FOR VIDEO COMPRESSION

B Henricksen

ERA Technology Ltd., England

ABSTRACT

Image warping is now commonly seen on television as a special effect, bending an image into a cylinder or such like. But image warping can also have uses in video compression.

As future trends head towards higher quality digital broadcast television and increased service traffic on transmission networks, the need for levels of higher compression and fidelity is driving research into techniques beyond MPEG2. These second generation coding techniques are aiming for a human visual model approach where objects in an image sequence are recognised as cohesive, but perhaps flexible, units. The motion of these objects and their apparent deformation in shape can be estimated and characterised by a number of descriptive parameters, which can then be transmitted. At the receiving end, a reconstruction of the moving object can be obtained from the object in a previous frame by using the parameters to define the necessary motion and deformation. This approach to motion estimation/compensation allows for a more natural movement of the object than purely translational models. This paper describes the attempts to produce a minimal but sufficient estimate of the motion of a flexible object. It is contrasted with traditional block matching techniques and the improvements in the prediction image obtained with warping motion estimation and compensation are shown.

INTRODUCTION

The application of motion estimation and compensation to image compression is now a standard technique incorporated into MPEG and associated video coding standards. The main objective is to predict the current frame to be encoded by transforming the previous frame or frames but using as little information as possible to describe the transformation. The prediction image is then subtracted from the current image and the small resultant error image is encoded and transmitted along with the transformation description. The current standard technique of motion estimation and compensation is block matching. The image is divided into a grid of typically 16 by 16 pel blocks. Each block is searched for in the previous image within a search area, with pel or even sub-pel accuracy until the minimum absolute difference (MAD) block of the same size, shape and

orientation is found. A motion vector is then stored for that block which is the displacement between the blocks in the two frames. The error between the two blocks and their motion vector are transmitted to the decoder. The motion compensation at the decoder consists of blocks in a grid being reconstructed from blocks in the previous frame, offset by the associated motion vector and the addition of the corresponding error block.

The motion vector describes the motion for the centre point of the block, assuming a purely translational model of motion. The motion vectors for all the blocks form a partial estimate of the motion vector field for the two images assuming a simple translational model.

The motion model fails when the motion is not purely translational or when objects with different motion vectors occupy the same block forming a motion discontinuity.

Block matching generally works well producing a low dynamic range error image, even for cases where the motion model fails, because of its matching criteria. The displacement of each block is determined so as to minimise the error between the current and the predicted block. If purely translational motion is present in the block then this criteria will provide an accurate estimate of the motion.

WARPING

Within most natural scenes the motion is not purely translational. Objects can rotate on various axes or deform. The motion vector field within these objects are continuous unless occlusions occur. For example when a camera zooms in on a scene the whole image has a smooth motion vector field increasing out from the centre of the zoom. Warping attempts to provide a better model of motion by allowing deformation, rotation and translation of a region in order to find an improved match and hence smaller error.

This is realised by allowing the deformation of the vertices of a square into any other quadrilateral. Each pel inside the four vertices is given a motion vector by bi-linear interpolation producing a smooth vector field. Since they are interpolated, these individual pel motion vectors do not have to be transmitted.

BASIC TECHNIQUE

Theory

The basis of this work was a paper by Nieweglowski and Haavisto (1). Motion estimation/compensation using a warping model has been developed in different ways using the triangular based affine transforms described in references Nakaya and Harashima (2) , Dudon et al (3) and Fuh and Maragos (4).

In the basic technique, the current frame which is to be predicted is divided up into a regular grid with nodes at the intersections. A low resolution estimate of the motion vector field is made for these nodes. Each node is assigned a motion vector by performing traditional block matching with the centre of the block corresponding to the position of a node. See Fig. 1a. This is the motion estimation process.

Motion compensation is provided by assigning the transmitted motion vectors to the relevant nodes. The nodes in the previous frame are then displaced by their respective motion vectors. Thus a square with a node at each corner in the current frame is mapped to a quadrilateral in the previous frame. It should be noted that the deformed square is not the square used to find the motion vector of the node by block matching. See Fig.1 b and c.

The pels in between the four nodes are assigned a motion vector by interpolating the vectors corresponding to the four nodes. The prediction of the current frame is then constructed by copying the relevant pels with their calculated shifts from the previous frame. Using the nomenclature of Figs 1b and 1c the formulae for the interpolation is given below.

$$\begin{bmatrix} 1 & u_0 & v_0 & u_0v_0 \\ 1 & u_1 & v_1 & u_1v_1 \\ 1 & u_2 & v_2 & u_2v_2 \\ 1 & u_3 & v_3 & u_3v_3 \end{bmatrix} = UV \quad \begin{bmatrix} x_0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix} = X,$$

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ y_3 \end{bmatrix} = Y, \quad \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ a_3 \end{bmatrix} = A, \quad \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix} = B$$

$$[A, B] = inv(UV)[X, Y] \quad (1)$$

$$[x_i, y_j] = [u_i v_j, v_j, u_i, 1] \begin{bmatrix} a_3 & b_3 \\ a_2 & b_2 \\ a_1 & b_1 \\ a_0 & b_0 \end{bmatrix} \quad (2)$$

For current frame vertices (u , v) and their associated previous frame vertices (x , y) the bi-linear warping coefficients (a , b) can be calculated using equation 1.

The interpolated co-ordinates of the pels in the previous frame are calculated by inserting the enclosed pel co-ordinates of the current frame as (u_i, v_j) into equation 2.

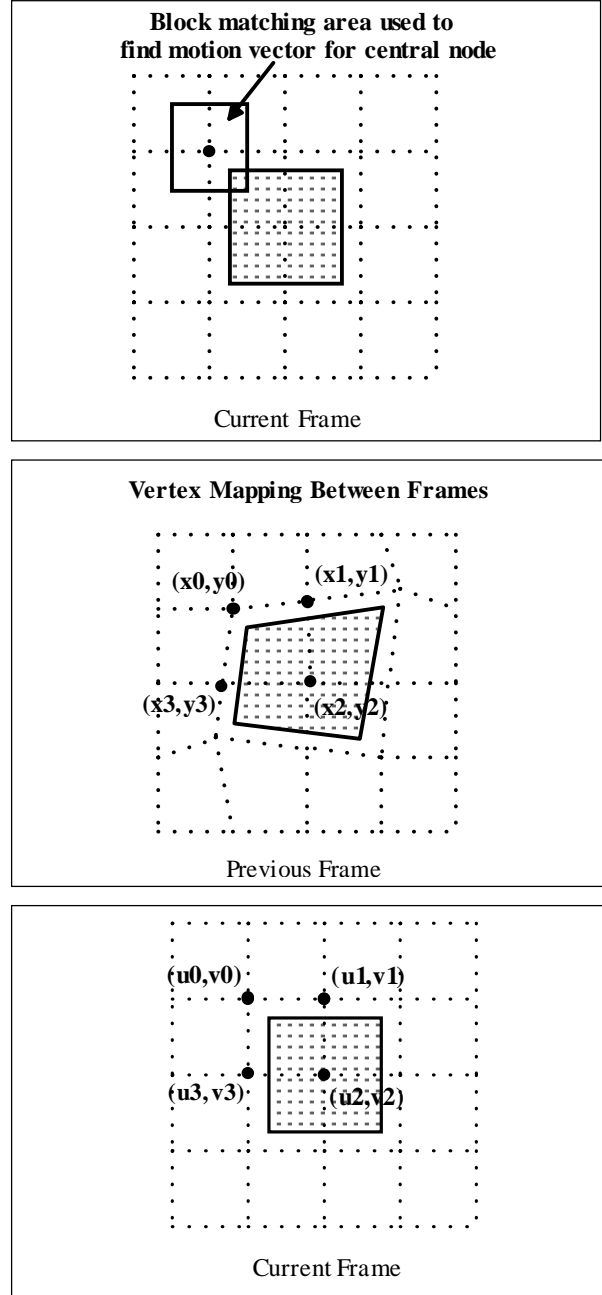


Fig.1a, b & c : Vertex mapping between frames

In this way the warping can model translation, rotation, stretching and shearing. This results in a smooth motion vector field as generally observed within objects. No blocking artefacts are present in the predicted image as seen with translational motion models. See Fig.2 for a comparison of a block matching prediction of an image and a warping prediction. The warping technique demonstrated is the final enhanced version discussed later. Discontinuities in the motion vector field are not dealt with by this model but more will be said about this later.

Application of Warping Motion Compensation

By having a square in the current frame and the arbitrary shaped quadrilateral in the previous frame, the transformation coefficients a and b can be calculated in advance saving computation time. This was pointed out by Nakaya and Harashima (2) in their paper on using affine triangles for warping. They also found that the peak signal to noise ratio (PSNR) of the prediction was similar if not better than calculating the coefficients for the transformation in the opposite direction.

The technique in the previous section is an initial description of the warping compensation concepts. If typical scenes are processed with this method as described above an increase in the prediction error is noticed over that achieved by block matching (see Table 1, key WRP1). This seems to come from two sources. The motion of the nodes is estimated by attempting to match a rigid square region to a square region in the previous frame. If the object within the current frame's region has deformed, a poor match will be achieved. Also at motion discontinuities the four vertices may have differing motion vectors, each belonging to different objects within the same block. This will produce a warped block with the majority of the inner pels having incorrect values. Conversely, with block matching the whole block is assigned the motion vector of the dominating area within the block. Thus the larger part of the block will have a good estimate of the motion and produce a small error. Hence block matching performs slightly better overall.

Nieweglowski & Haavisto recognised the difficulties experienced with warping and block matching at motion vector field discontinuities. Finding that block matching results in a lower prediction error, they adaptively switched between the two techniques at blocks having vertices with radically differing motion vectors.

When the interpolated motion vectors are calculated for each pel in a block, (u, v) , fractional motion vectors are produced. If these are rounded to the nearest pel the results shown in Table 1 Key WRP1 are achieved. If however the intensity value of the fractional position between pels is estimated by bi-linear interpolation of the four adjacent pels, it is possible to improve the warping procedure. In a simple test, a reference image was warped to a new quadrilateral and then warped back to the original vertices and the result compared with the original image. Improvements of up to 2.5 dBs were achieved for some deformations by using an interpolated intensity value rather than the nearest pel value. The results for block matching followed by warping compensation using interpolation for fractional motion vectors on a sequence are given as a comparison in Table 1 with key WRP 2.

DEVELOPMENTS

To improve the motion estimate and to partially deal with the motion discontinuities the following ideas have been incorporated and tested.

The block matching motion estimate is not satisfactory since it relies on matching a pair of square regions, which will not fit well if the contents of one have deformed. This can be improved by varying the shape of the region to be matched to the square in the current frame. To achieve this the vertices of the current block are moved through a search pattern in the previous frame. At each position the contained pels are found by interpolating the four vertices' vectors and the error of this prediction block found. The node offsets which give the minimum error are the modified motion vectors. In this way a smaller minimum can be found when the motion is not purely translational.

For an optimal solution when performing this estimation, every combination of vertex in each search position, with all other vertices in every search position should be tried. For a reasonable search size (± 16 pels horizontally and vertically is normal) the number of combinations would be unreasonably high to compute. To reduce this an initial estimate of the motion vector is made for each node by block matching. Then the first vertex of a block is then moved through a search pattern (± 2 pels), the warping applied and all the pels of the warped block are calculated along with the error for each search location. The minimum error displacement is found. This vertex is set to this modified displacement and then the search is conducted for the subsequent vertices. Note that the error minimum is located for a single vertex at a time. This is not an optimal solution as all combinations of vertex positions within the search ranges are not tried.

It is important to note that in this way a motion vector is assigned to each vertex of a block without consideration of the neighbouring blocks. Thus four motion vectors have to be assigned to every block. This is where the technique differs from the Nieweglowski approach. This allows blocks to partially deal with discontinuities without affecting neighbouring blocks. If a motion discontinuity from two moving objects within the block is present, the deformed area can be taken from the larger object area only, giving small error for that region and hence a smaller overall block error.

The benefits of this improved estimation can now be seen with a smaller prediction error image being produced. See Figs 3 and 4 and Table 1 Key WRP1. BM16 is the traditional block matching technique. WRP 7 is the perturbed warping estimation followed by warping compensation.

In this warping estimation process, when the first vertex is perturbed to find its best position, the other vertices

have not moved. It can be assumed that its best position could change when the other vertices are optimally positioned. Similarly for the other vertices. To improve this a second iteration of the search is performed. This gives a prediction improvement as shown in Figs 3 and 4 and Table 1 with key WRP 6. The benefits of further searches have been found to be minimal.

RESULTS

Table 1 lists the average peak signal to noise ratios of predicted images. The predictions are produced by a number of methods which are given a key to refer to Fig 3 & 4. Varied standard test sequences were used to compare the techniques. The original previous frames from the sequences are used to generate the predictions of the current frame.

Key to Table 1 and Figs. 3 & 4	
BM16	Standard block matching motion estimation followed by block matching motion compensation with 16 by 16 pel block size.
BM8	Block matching motion estimation and compensation with 8 by 8 pel blocks
WRP1	Block matching motion estimation based on 16 by 16 pel blocks followed by warp motion compensation but with out interpolation of fractional motion vectors
WRP2	Block matching motion estimation based on 16 by 16 pel blocks followed by warp motion compensation with interpolation of fractional motion vectors
WRP8	Warp motion estimation based on 16 by 16 pel blocks but without interpolation of fractional motion vectors followed by warp motion compensation without interpolation of fractional motion vectors
WRP7	Warp motion estimation based on 16 by 16 pel blocks with interpolation of fractional motion vectors followed by warp motion compensation with interpolation of fractional motion vectors
WRP6	Warp motion estimation based on 16 by 16 pel blocks with interpolation of fractional motion vectors with a refinement iteration followed by warp motion compensation with interpolation of fractional motion vectors.

The WRP1 and WRP2 methods could be used as warp compensating decoders for MPEG2 coders using block matching but only the Suzie sequence for WRP1 gave an improvement over block matching compensation.

However, the results in Table 1 and Figs 3 and 4 indicate a clear gain for the warping estimation/compensation over the block matching technique with a number of differing sequences. This gain is up to 2.7 dBs. A

comparison with block-matching for 8 by 8 pel blocks with equivalent number of motion vectors is also given (Key BM8). The technique gives up to 0.7 dB gain over this. The motion estimation refinement stage is shown to give an appreciable gain to the warping prediction accuracy of up to 0.3 dBs.

DISCUSSION

The benefits of motion estimation and compensation by warping are based on its ability to model a smooth motion vector field such as found within an object (which does not have any occlusions) and to deal with rotational movement or apparent deformations within the object. Considering this, block sizes can be adapted to give less error along boundaries with a quad-tree type approach. This can also give a higher density of motion vectors where required to describe the warping area more accurately within an object. If motion segmentation has been used to separate the image into objects, warping is an ideal way to model motion within the objects. Also, there should be less need to have four vectors per node within the object's motion field. These approaches are to be explored next.

CONCLUSIONS

The technique developed has been shown to give significant gains in the accuracy of the prediction image over block matching with equivalent block sizes, and over block matching with equivalent number of motion vectors. The gains made are sufficient to overcome inadequacies in the motion model at motion discontinuities compared with block matching. The motion models give significant coding gain and are sufficiently accurate to be ripe for exploring with second generation coding techniques.

ACKNOWLEDGEMENTS

This work has been performed under the RACE 2110 HAMLET project on behalf of the Independent Television Commission.

REFERENCES

1. Nieweglowski, J., Haavisto, P., 1993, "A method for motion compensated video prediction", International Workshop on HDTV '93, Ottawa, Canada, October 26-28.
2. Nakaya, Y., Harashima, H., 1991, "An iterative motion estimation method using triangular patches for motion compensation", SPIE VCIP '91.
3. Dudon, M., Avaro, O., Eude G., 1994, "Object-oriented motion estimation", PCS '94.
4. Fuh, C.S., Maragos, P., 1991, "Motion displacement estimation using an affine model for image matching", Optical Engineering, Vol.30, No.7.

Prediction Image Comparison - Block Matching vrs Warping

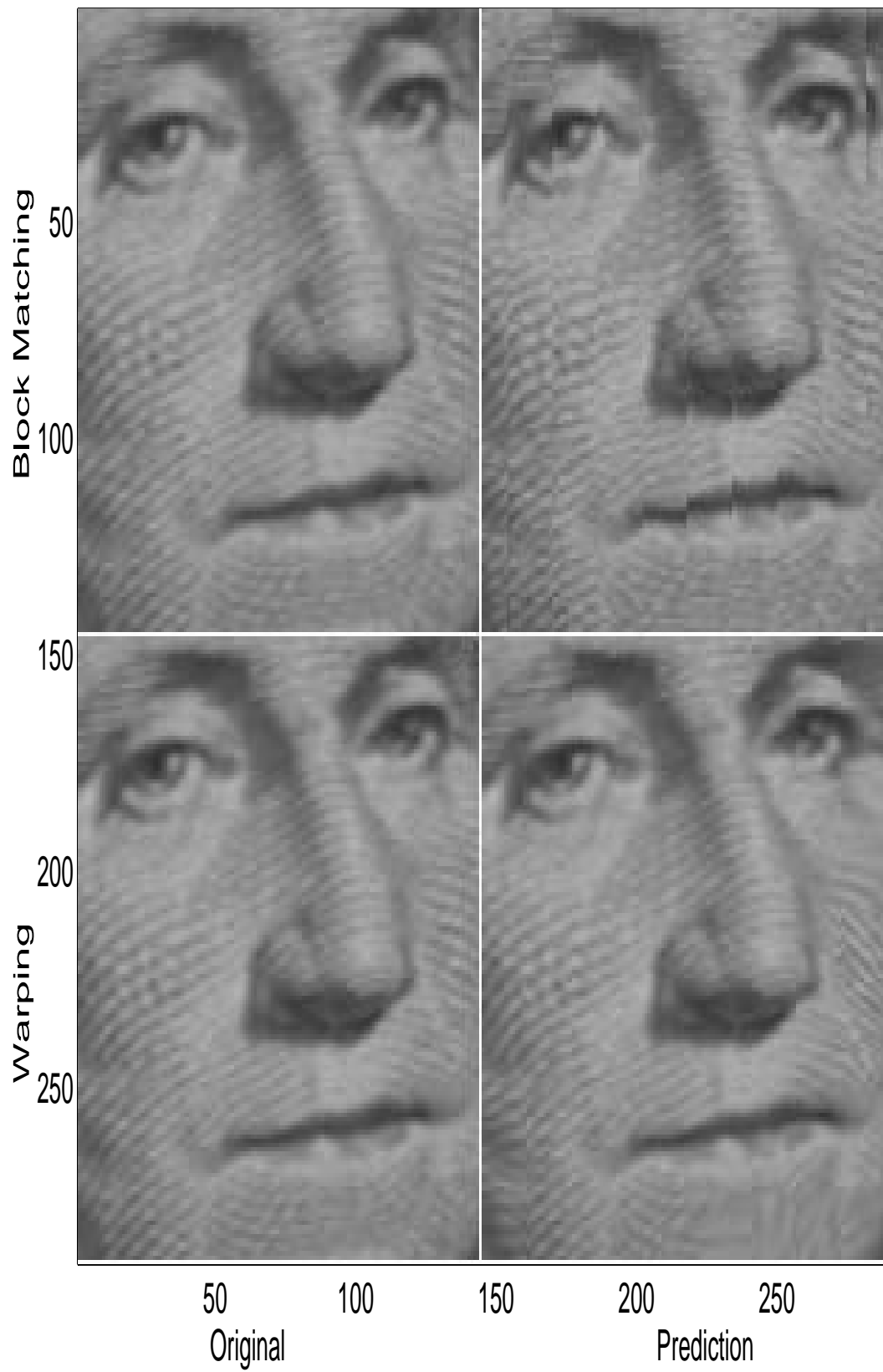


Figure 2

TABLE 1 - Comparison of Prediction Errors of the Techniques				
Average Peak Signal to Noise Ratio of Predicted frame	Sequence Title			
Technique	Mobile & Calendar (25 frames)	Kiel Harbour (25 frames)	Flower Garden (5 frames)	Suzie (5 frames)
BM16: BM Est + BM Comp 16 by 16 blocks	22.44	24.04	25.2	34.48
BM8: BM Est + BM Comp 8 by 8 blocks	23.15	24.47	25.96	36.51
WRP1: BM Est + Warp Comp, no fractional motion vector interpolation	-	-	24.71	35.02
WRP2: BM Est + Warp Comp	-	-	24.09	33.73
WRP8: Warp Est, no fractional motion vector interpolation + Warp Comp, no fractional motion vector interpolation	-	-	22.1	31.16
WRP7: Warp Est + Warp Comp	23.74	25.33	25.78	36.93
WRP6: Warp Est + refinement + Warp Comp	23.91	25.41	25.99	37.24

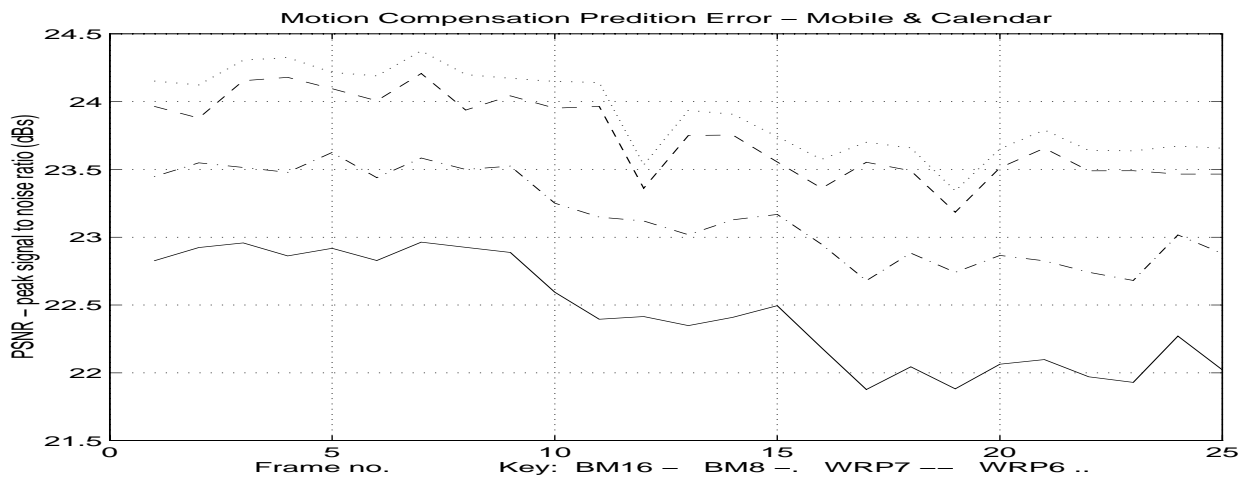


Figure 3

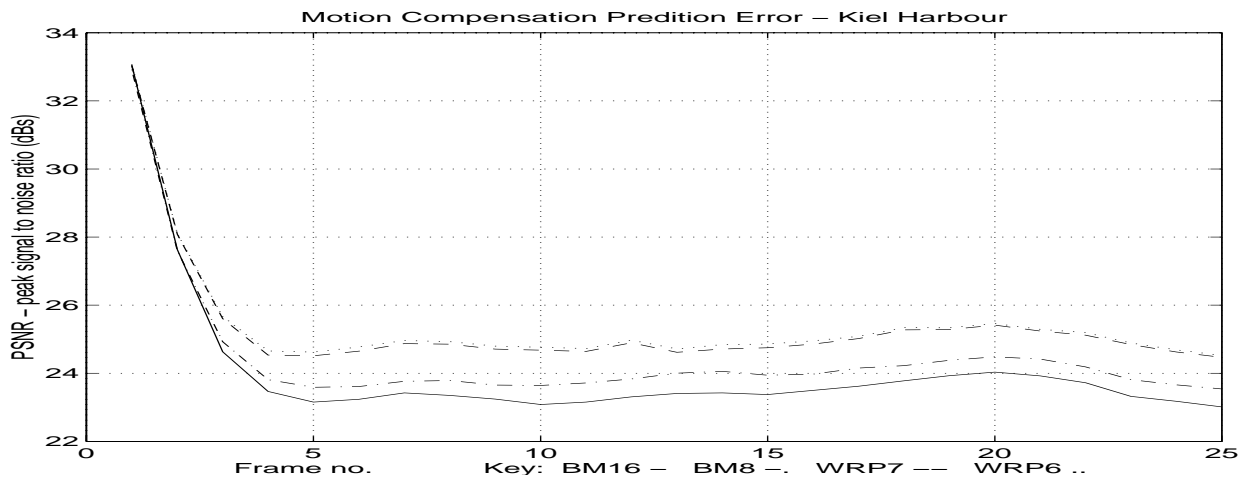


Figure 4