

Motion Segmentation Initialization Strategies for Bi-Directional Inter-Frame Prediction

Ashek Ahmmed #, Rui Xu *, Aous Thabit Naman *, Md. Jahangir Alam #, Mark Pickering #, and David Taubman *

*School of Engineering and Information Technology,
The University of New South Wales, Canberra, Australia.*

* *School of Electrical Engineering and Telecommunications,
The University of New South Wales, Sydney, Australia.*

Abstract—Experimental results and the latest standards have proved that segmentation based video coding systems can outperform the traditional block-based video coding systems. However, this approach requires the simultaneous estimation of both the shape and motion of moving objects in a video scene. In most of the cases neither the shape nor the motion are known initially. Another critical aspect of this tightly-coupled relationship is that inaccurate motion estimation may cause poor segmentation and erroneous segmentation may negatively impact motion estimation. While some of the existing approaches require user intervention and some use clues such as depth, color or occlusion to separate the foreground from the background, we propose to use motion reliability information for this purpose. This is because the ingredients necessary for the calculation of motion reliability are the by-product of block-based motion estimation and compensation between the reference frames. Therefore, they require very little or no increase in the computational overhead. In this paper, we explore several motion segmentation initialization strategies based on motion reliability. The performances of these initialization approaches are investigated, in terms of the PSNR, for the predicted intermediate frames.

I. INTRODUCTION

In traditional video coding, motion estimation and compensation play crucial roles in granting high compression gains. Spatially adjacent pixels are grouped into blocks that are predicted via equally-shaped blocks that are to be found in the previously-decoded frames. However, the adaptation of block shapes can significantly improve the final coding performance since it is possible to match the blocks to the objects in the scene. This adaptation partially mitigates the compression inefficiency derived from the unnatural division of the image to be coded into square or rectangular blocks.

Recent video coding standards allow the adoption of block partitioning involving variable sized blocks that can be adapted according to the rate-distortion performance on the current frame. As an example, the H.264/AVC standard [1] supports several types of block partitions from 4×4 to 16×16 pixels. More recently, the HEVC [2] standard allows prediction blocks sized up to 64×64 pixels. Careful partitioning of motion

blocks in the vicinity of object boundaries represents a crude yet important way of segmenting the motion vector field into disjoint regions, with a smooth (typically constant) motion model within each block [3-5].

A more flexible partitioning can be obtained via segmentation. The possibility of characterizing individual objects in a coded scene was standardized within the MPEG-4 video coding standard [6], but it was never widely adopted and so far, few coding systems take full advantage of the object-oriented coding tools offered by MPEG-4. Building on the idea proposed in [7], Tagliasacchi *et al.* [8] proposed a motion estimation algorithm using a quadtree structure which produces a region based motion representation. A prune-merge scheme is used to segment the input image into regions. Blocks characterized by the same motion model are grouped together in order to reduce the amount of bits allocated to motion. In the pruning phase each 16×16 block is partitioned into sub-blocks optimally in a rate-distortion sense. While in the merging phase blocks representing the same moving object are represented by the same motion vector i.e. these blocks are merged into a single block.

In [9] an implicit block segmentation approach is proposed where segmentation is performed on the difference of the two predictors. This segmentation is based on the fact that for a 16×16 block each predictor may reduce the matching error non-uniformly inside that block. Their approach showed encouraging results for the *Foreman* sequence where illumination mismatches are not shown. Milani *et al.* [10] proposed a segmentation-based video coding system that partitions each frame into arbitrarily-shaped segments for a more effective motion compensation. Their scheme has shown a significant improvement (up to 4 dB) with respect to the H.264/AVC standard depending on the amount of motion in the sequence and on the size of the generated segments.

A novel approach was proposed in [11] that uses motion hints for inter-frame prediction. Motion hints provide a global description of motion over specific domains. Fundamentally this is related to the segmentation of foreground from background regions where the foreground and background motions are the motion hints. The appealing thing about motion hints



Fig. 1. The two reference frames of the *Foreman* sequence used herein as a running example.

is that they are continuous and invertible, even though the observed motion field for a frame will be discontinuous and non-invertible. It has been shown that, with reasonably accurate motion, inter-frame predictions with good subjective quality and high PSNR can be generated [12].

Leveraging on the promising results shown by segmentation-based video coding and inter-frame prediction using motion hints, a bi-directional segmentation-based motion compensated prediction paradigm that employs motion hints has been developed. In this approach each reference frame is partitioned into arbitrarily shaped foreground-background regions based on motion and thereby their motion hints are generated. These segmented foregrounds and backgrounds are then projected onto the current frame and predictions for its foreground and background are formed. Finally, by fusing the predictions through a weighted scheme, a prediction of the current frame is generated. However to function properly, this approach requires an initial estimate of the foreground-background shapes so that the algorithm can gradually refine these initial shapes through successive motion estimation and compensation stages and produce highly accurate foreground-background segmentation and motion hints.

In [13] a robust bi-layer segmentation algorithm to extract moving objects from videos captured by hand-held cameras was proposed. This approach finds clues from depth and motion to estimate camera configurations and then can warp one frame against its reference frame. The difference between the reference frame and the warped version of it acts as the initial segmentation mask. Xiong *et al.* [14] used forward and backward occlusion maps generated by camera motion to find foreground and background seeds. Then an interactive image segmentation algorithm was used to separate the foreground from the background using those seeds.

In this paper, we generate the foreground-background seeds based on motion reliability information and then by clustering the reliable motion vectors. A major advantage of this approach is that the ingredients necessary for determining if the estimated motion of a block is reliable or not are the by-product of the block-based motion estimation and compensation procedure between the reference frames. Hence only a small increase in the overall computational overhead

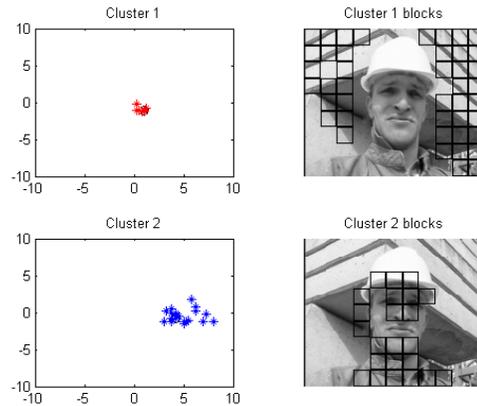


Fig. 2. Clustered reliable MVs and the foreground-background seeds over the *Foreman* sequence.

is required for our segmentation based inter-frame prediction paradigm. In the work presented herein we measure the performance of a number of initialization approaches in terms of the PSNR for the predicted intermediate frames.

The rest of this paper is organized as follows: in section II we describe the segmentation based inter-frame prediction paradigm in brief. Section III contains descriptions of the considered initialization approaches. The prediction performance of the segmentation based prediction algorithm over different video sequences when using the seeds produced by different initialization approaches is then investigated. Finally, in section V, we present our conclusions from these results.

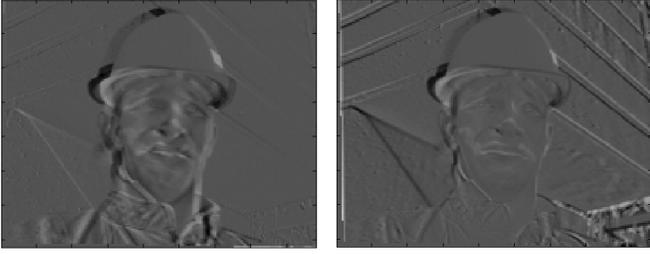
II. SEGMENTATION-BASED INTERMEDIATE FRAME PREDICTION PARADIGM

The prediction architecture has three main parts: the first part performs the forward foreground-background motion segmentation, the second one performs the backward foreground-background motion segmentation and the third part generates the prediction of the current frame by projecting these foregrounds and backgrounds, computed in the previous steps, on to the current frame. In the following subsections, these parts are discussed briefly.

A. Forward Motion Segmentation

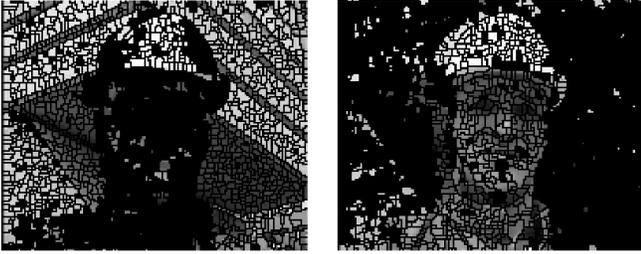
In the bi-directional motion estimation setting the previous and future reference frames are denoted by R_i and R_j respectively herein. A typical example of them is shown in Figure 1.

1) *Initialization*: The initial foreground-background shapes work as seeds for the actual foreground-background regions. Given all the blocks of R_j each of the addressed initialization approaches produces a set of blocks that have reliable estimated motion. The reliable motion vectors are then clustered into groups where the blocks corresponding to the motion vectors belonging to the same group together form the foreground-background seeds. This is shown in Figure 2. How each approach finds its own set of reliable blocks and the dependency of our inter-frame prediction approach on distinctive enough foreground-background seeds will be described in detail in Sections III and IV respectively.



(a) Prediction error due to $M_1^{(R_i \rightarrow R_j)}$ compensation (b) Prediction error due to $M_2^{(R_i \rightarrow R_j)}$ compensation

Fig. 3. Forward motion compensated prediction errors.

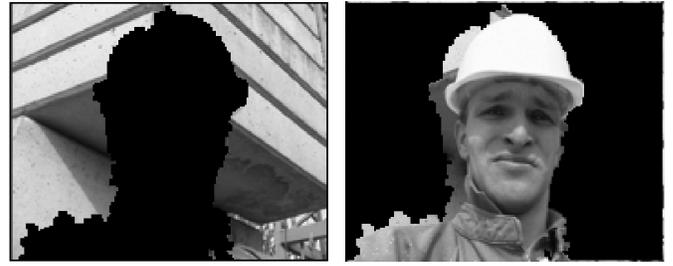


(a) Region 1 of R_j (b) Region 2 of R_j

Fig. 4. Improved foreground-background segmentation for R_j . Some super-pixels are common to both the regions e.g. inside the hat of the *Foreman*; both motion models compensated their motion very well. Some super-pixels are missing from both the regions e.g. the left portion of the hat's rim; here the prediction errors from both the models are higher than an error tolerance threshold.

2) *Shape Refinement*: Having the initial foreground-background shapes, it is possible to estimate the optimal 6-parameter motion models i.e. the affine motion models associated with these shapes. And once these initial values of the forward motion hints i.e. of $M_1^{(R_i \rightarrow R_j)}$ and $M_2^{(R_i \rightarrow R_j)}$ are known, motion compensated predictions of R_j can be generated by warping R_i using these hints. In Figure 3, the existence of two predominant motions between R_i and R_j can be detected by observing that one motion model compensates the motion in part of the frame while the other one fails there and vice-versa. The precise location and boundary of such regions can be estimated with the help of a color-based segmentation of R_j that partitions R_j into super-pixels. The performance of each motion model in compensating the motion within each super-pixel is then investigated and finally the super-pixels are grouped into homogeneous motion model groups. This is shown in Figure 4.

3) *Iterative Motion Estimation-Shape Refinement*: The proposed approach adopts an iterative 2-step strategy where, in the first step, the motion hints of the available shapes are estimated and used to generate motion compensated prediction errors. In the next step these prediction errors are used to group the super-pixels of R_j into homogeneous motion model groups and thereby refine the shapes. The approach toggles between these 2 steps until the motion segmentation becomes stable. Along with the prediction errors, an estimate of the



(a) Region 1 of R_j (b) Region 2 of R_j

Fig. 5. Converged foreground-background segmentation for R_j .



(a) Region 1 of R_i (b) Region 2 of R_i

Fig. 6. Converged foreground-background segmentation for R_i .

smoothness of the motion hint fields is used to regulate this optimization stage. The outputs of this iterative multi objective approach are the converged forward motion hints as well as the shapes shown in Figure 5.

B. Backward Motion Segmentation

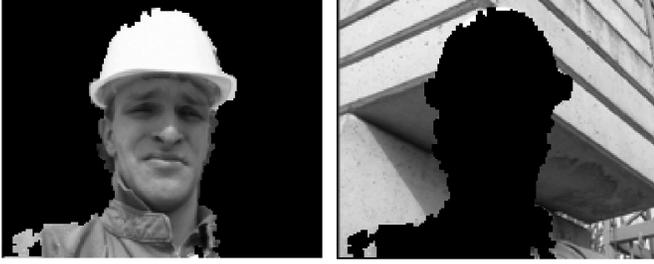
After performing the forward motion segmentation, the algorithm basically repeats the same procedure it used to find the forward foreground-background segmentation with the roles of frames R_i and R_j interchanged.

1) *Initialization*: The initial motion segmentation is more accurate in the backward case than it was in the forward case. This is because of the availability of the forward motion hints. The initial values of the backward motion hints are generated as follows.

$$M_n^{(R_j \rightarrow R_i)} = (M_n^{(R_i \rightarrow R_j)})^{-1} \quad n = 1, 2.$$

Unlike in Section II-A1 where we started with foreground-background seeds and then obtained the initial forward motion hints, the backward initialization starts with motion hints and then finds the shapes using motion compensated prediction errors and a color based segmentation of R_i that partitions this frame into super-pixels.

2) *Iterative Motion Estimation-Shape Refinement*: The estimated seeds in the backward case are composed of super-pixels. Therefore they can be fed into the iterative motion estimation-shape refinement paradigm discussed in Section II-A3 straightaway with the same parameter values used in the forward case and it is possible to obtain the converged backward motion hints and the segmentation of Figure 6.

(a) $F_j^* = f_j^* \cdot R_j$ (b) $B_j = b_j \cdot R_j$ Fig. 7. Foreground and background of R_j respectively.(a) $F_i^* = f_i^* \cdot R_i$ (b) $B_i = b_i \cdot R_i$ Fig. 8. Foreground and background of R_i respectively.

C. Prediction of the Current Frame

1) *Bi-directional Foreground Correction*: The estimated segmentations of R_i and R_j are now improved using the segmentation information available in both reference frames. At this stage the algorithm requires the knowledge of which region is the foreground of R_i and R_j . In the remainder of this example Region 2 is considered to be the foreground and Region 1 to be the background. The algorithm makes use of binary images, which we refer to as masks. These masks have a value of 1 to indicate that a pixel belongs to a region and 0 elsewhere in the image.

The foreground correction approach starts with the existing foreground mask for example of R_j which will be denoted by f_j herein and then forms a new foreground mask f_j^* using the intersection between the existing mask f_j and the forward foreground motion compensated foreground mask from R_i .

$$f_i^* = M_2^{(R_j \rightarrow R_i)}(f_j) \cap f_i$$

where $M_2^{(R_j \rightarrow R_i)}(f_j)$ denotes applying the motion hint $M_2^{(R_j \rightarrow R_i)}$ to the foreground mask f_j . Similarly, f_i^* is the new foreground mask formed in R_i and is given by:

$$f_j^* = M_2^{(R_i \rightarrow R_j)}(f_i) \cap f_j$$

Note that the new foreground masks f_i^* and f_j^* are now related through the forward/backward foreground motion hints. For example, $f_j^* = M_2^{(R_i \rightarrow R_j)}(f_i^*)$, apart from numerical approximations introduced by the warping process and any differences which might exist between $M_2^{(R_i \rightarrow R_j)}$ and $(M_2^{(R_i \rightarrow R_j)})^{-1}$.

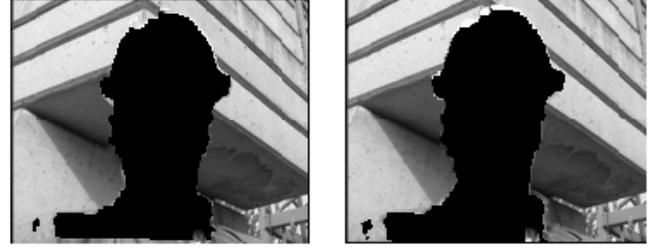
(a) B_j^* (b) B_i^*

Fig. 9. The modified backgrounds with uncovered regions.

The remaining parts of R_i and R_j are declared to be background masks and are given by:

$$b_i = 1 - f_i^*$$

$$b_j = 1 - f_j^*$$

The foreground and background regions, denoted by F_i^* , F_j^* , B_i and B_j , found by multiplying the reference frames with these masks are shown in Figures 7 and 8.

2) *Uncovered Regions Determination and Background Modification*: New background masks b_i^* and b_j^* are now formed by determining the uncovered regions in R_j and R_i and then adding them to the existing background regions b_i and b_j respectively. The modified backgrounds are shown in Figure 9.

3) *Projection onto the Current Frame*: To predict the current frame C , the proposed approach now has all the required ingredients. It starts by estimating the motion hints $M_2^{(R_i \rightarrow C)}$ and $M_2^{(R_j \rightarrow C)}$. The predicted foreground mask \hat{f}_c is generated by taking the intersection between these motion compensated foreground masks as follows:

$$\hat{f}_c = \left(M_2^{(R_i \rightarrow C)}(f_i^*) \right) \cap \left(M_2^{(R_j \rightarrow C)}(f_j^*) \right)$$

The predicted background mask \hat{b}_c is generated by using the old background masks to find the motion hints and then warping the new ones with these hints.

$$\hat{b}_c = \left(M_1^{(R_i \rightarrow C)}(b_i^*) \right) \cup \left(M_1^{(R_j \rightarrow C)}(b_j^*) \right)$$

Using these background and foreground masks, the prediction \hat{C} is generated in the following way:

$$\hat{F}_c = \left(M_2^{(R_i \rightarrow C)}(R_i) + M_2^{(R_j \rightarrow C)}(R_j) \right) \cdot \hat{f}_c$$

$$\hat{B}_c = \left(M_1^{(R_i \rightarrow C)}(B_i^*) + M_1^{(R_j \rightarrow C)}(B_j^*) \right) \cdot \hat{b}_c$$

$$\hat{C} = \hat{F}_c + (1 - \hat{f}_c) \cdot \hat{B}_c$$

III. FOREGROUND-BACKGROUND SHAPE INITIALIZATION

Due to the tightly coupled nature of the relationship between motion and shape it is necessary to initialize the algorithm described in the previous section with distinctive enough foreground-background shapes. We propose to do this by deriving the initial forward motion hints using only blocks where the motion vectors are deemed reliable. However this objective should be accomplished in a computationally simple way and thereby not significantly increase the overall complexity of the prediction paradigm. In this Section we investigate 4 different motion reliability estimators that produce a set of blocks with reliable motion which are then clustered based on their motion vectors to obtain the shape seeds.

Long *et al.* [15] showed that the ratio r_l between the low frequency energy in a phase-matched frame difference (PMFD) [15] image block and the total energy in that block can be used as a measure of motion reliability for the corresponding block of R_j . Blocks with high r_l are deemed to be unreliable. We have found in our experimental evaluation that, when applied to motion reliability detection, the performance of the PMFD image based r_l measure can be very well matched by the performance of r_l calculated from the displaced frame difference (DFD) image and the DFD based r_l is comparatively easy to calculate because the DFD image is a direct consequence of motion estimation and compensation between R_i and R_j . The r_l measure of a DFD image block is calculated in the following way:

$$r_l = \frac{E_l}{E_t}$$

where E_l and E_t denote the amount of low frequency energy and the total energy in a DFD block respectively.

The next approach for motion reliability that we consider is also a by-product of the block-based motion estimation and compensation between the frames R_i and R_j . While performing the block based translational motion estimation for each block to be matched in the previous frame, the sum of absolute differences (SADs) between that block and all the candidate blocks within the search window are stored. The kurtosis of the distribution of these SADs can be used as an indicative measure of the estimated local motion vector's reliability [16]. Since blocks in flat image regions will be matched to many blocks in the neighbourhood, their SAD distribution are expected to have low kurtosis which in turn suggests they have unreliable motion. On the other hand blocks with corners or texture patterns can have reliable local motion estimation which is evident from the high kurtosis of their SAD distributions.

In [17] motion vectors which are found in regions with little or no texture or a moving object boundary and regions with repetitive texture patterns are deemed as "noisy". A filter is used to examine the magnitude and phase difference between a motion vector and its 8-adjacency neighbors and then a fraction of the available motion vectors was removed based on spatial dissimilarity. For our work, we consider the blocks that help to keep the motion vector field smooth as reliable.

And to keep things simple we only consider the motion vector magnitude difference and use a threshold to count the number of neighbouring blocks, in the 8-adjacency neighborhood, with motion that is dissimilar to the current block. If the majority of a block's neighbours are found to have dissimilar motion, the motion vector of the block is assumed to be unreliable and is not used in the subsequent clustering process.

Xu *et al.* [18] employed a multi-scale block-based approach for motion estimation that generates a dense motion field. Their approach involves applying a 2-bit transform to each detail band in the Laplacian pyramid representation of each video frame; motion search is then performed on the transformed detail bands at each detail level, employing different apertures, to generate a matching score for each location and amount of spatial shift. These scores are then combined across the detail levels, using the structural measure proposed in [19], to generate an overall score for each location and amount of shift, where the motion estimate with the highest score is chosen. This approach tends to produce motion fields that better reflect the true underlying motion than those obtained using MSE as a matching criterion, while also being more robust to noise and illumination variations, as shown in [18]. For our work we use the matching score of the center pixel of a block to determine its reliability.

Next we investigate the performance of the discussed initialization approaches on the QCIF sequences *Foreman*, *Hand held Mobile phone* where a person mimics video conferencing on a mobile phone by recording himself talking and at the same time moving while the mobile phone is in his right hand and *Stair* [13] where a person is walking and another person is following him and recording him with a hand-held camera.

IV. EXPERIMENTAL RESULTS

To evaluate the performance of the foreground-background seed generation approaches, we clustered the reliable blocks returned by each approach and thereby generated the initial foreground-background segmentation. Block sizes of 4×4 , 8×8 and 16×16 were tested. This initial segmentation information is fed into the inter-frame prediction paradigm which performed motion estimation and compensation between pairs of frames from the aforementioned video sequences. The reference frame pairs were used to predict two intervening frames and 10 such pairs were used i.e. we predicted 20 intervening frames. The prediction quality is measured in terms of the PSNR and the average results are reported in Tables I-III. Figure 10 shows an example of the current frame and prediction frame produced by the algorithm when using the 2-bit transform-based initialization.

TABLE I
Average PSNR (in dB) between the predicted and current frames for the *Foreman* sequence

| | 4×4 | 8×8 | 16×16 |
|-------------------------|--------------|--------------|----------------|
| r_l | 36.66 | 36.85 | 37.07 |
| SAD _{kurtosis} | 36.63 | 37.01 | 37.61 |
| MV _{smooth} | 37.21 | 37.37 | 37.47 |
| 2bit | 37.72 | 37.24 | 36.35 |



Fig. 10. The outcome of the prediction paradigm with the 2-bit transform based initialization strategy (used block size = 4×4).

TABLE II
Average PSNR (in dB) between the predicted and current frames for the *Hand held Mobile phone* sequence

| | 4×4 | 8×8 | 16×16 |
|-------------------------|--------------|--------------|----------------|
| r_l | 35.19 | 35.65 | 35.72 |
| SAD _{kurtosis} | 35.73 | 35.89 | 35.98 |
| MV _{smooth} | 35.05 | 35.98 | 35.72 |
| 2bit | 35.95 | 36.27 | 35.12 |

TABLE III
Average PSNR (in dB) between the predicted and current frames for the *Stair* sequence

| | 4×4 | 8×8 | 16×16 |
|-------------------------|--------------|--------------|----------------|
| r_l | 34.07 | 34.54 | 34.85 |
| SAD _{kurtosis} | 34.45 | 34.52 | 35.01 |
| MV _{smooth} | 34.06 | 34.21 | 34.55 |
| 2bit | 35.14 | 34.46 | 34.41 |

Firstly, it can be observed from the results that the 2bit transform-based approach performs the best for all the 3 video sequences. This is expected because of its ability to estimate the true underlying motion better than the traditional motion estimation approaches that uses MSE or SAD as the matching criterion. Note that this measure performed the best when smaller block sizes were utilized. Alternatively, the r_l and kurtosis-based initialization approaches perform better with bigger block sizes. This is because, for smaller blocks, the ratio r_l would most of the time be high therefore many false positives may happen. For the kurtosis-based approach, a smaller current block may match too many smaller blocks in the neighborhood and hence the SAD distribution is expected to have low kurtosis. The kurtosis-based approach using 16×16 block sizes performed second best to the 2-bit transform-based approach. It has an advantage over the 2-bit transform based approach in that it is computationally simpler and produces a smaller number of reliable blocks to be clustered.

V. CONCLUSION

In this paper, we have investigated the performance of a variety of motion reliability based segmentation initialization strategies for a segmentation-based prediction algorithm. The prediction performance of this algorithm when using initial segmentations generated by these motion reliability based

strategies is taken as the criterion to compare their relative performance. We have found in our experiments that the 2-bit transform based motion reliability approach showed the best performance in terms of high PSNR.

REFERENCES

- [1] Joint Video Team (JVT) of ISO/IEC MPEG and IUT-T VCEG, "Joint final committee draft (JFCD) of joint video specification (ITU-T Rec. H.264-ISO/IEC 14496-10 AVC)," in *Joint Video Team, 4th Meeting*, Klagenfurt, Germany, July 2002.
- [2] T. Davies, K. R. Andersson, R. Sjberg, T. Wiegand, D. Marpe, K. Ugur, J. Ridge, M. karczewicz, P. Chen, G. Martin-Cocher, K. McCann, W.J. Han, G. Bjontegaard, and A. Fuldseth, "Suggestion for a test model," in *Joint Collaborative Team on Video Coding (JCT-VC), 1st Meeting*, Dresden, Germany, Apr. 15-23, 2010.
- [3] A. A. Muhi, M. R. Pickering, and M. R. Frater, "A fast approach for geometry-adaptive block partitioning," *Picture Coding Symposium (PCS)*, 2009, pp. 413-416.
- [4] A. A. Muhi, M. R. Pickering, and M. R. Frater, "Motion compensation using geometry and an elastic motion model," *IEEE Int. Conf. Image Processing (ICIP)*, Nov 2009, pp. 621-624.
- [5] R. Mathew and D. S. Taubman, "Scalable modeling of motion and boundary geometry with quad-tree node merging," *IEEE Trans. CSVT*, vol.21, no.2, pp.178-192, Feb. 2011.
- [6] I. E. G. Richardson, *H.264 and MPEG-4 Video Compression*, John Wiley and Sons, Sept. 2003.
- [7] R. D. Forni and D. Taubman, "On the benefits of leaf merging in quadtree motion models," *Proc. IEEE Int. Conf. Image Processing*, vol. 2, pp. 858-861, September 2005.
- [8] M. Tagliasacchi, M. Sarchi, and S. Tubaro, "Motion estimation by quadtree pruning and merging," *IEEE Int. Conf. Multimedia Expo (ICME)*, July 2006, pp. 1861-1864.
- [9] J. Kim, A. Ortega, P. Yin, P. Pandit, and C. Gomila, "Motion compensation based on implicit block segmentation," *IEEE Int. Conf. on Image Processing (ICIP)*, 2008, pp. 2452-2455.
- [10] S. Milani and G. Calvagno, "Segmentation-based motion compensation for enhanced video coding," *IEEE Int. Conf. on Image Processing (ICIP)*, 2011, pp. 1685-1688.
- [11] A.T. Naman, D. Edwards, and D. Taubman, "Efficient communication of video using metadata," *18th Proc. IEEE Int. Conf. Image Proc.* 2011, pp. 589592, September 2011.
- [12] A.T. Naman, Rui Xu, and D. Taubman, "Inter-frame prediction using motion hints," *20th Proc. IEEE Int. Conf. Image Proc.*, September 2013, submitted for publication, available at <http://dstn.ee.unsw.edu.au/~aous/pdfs/aous-2013-b.pdf>.
- [13] G. Zhang, J. Jia, W. Hua, and H. Bao, "Robust bilayer segmentation and motion/depth estimation with a handheld camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 3, pp. 603617, 2011.
- [14] H. Xiong, Z. Wang, R. He, and D. Feng, "Video object segmentation with occlusion map," *International Conference on Digital Image Computing Techniques and Applications (DICTA)*, Dec. 2012.
- [15] L. To, M. Pickering, M. Frater, J. Arnold, "A motion confidence measure from phase information," *IEEE Int. Conf. on Image Processing (ICIP)*, 2004, pp. 2583-2586.
- [16] Y. C. Chung and Z. He "Reliability Analysis for Global Motion Estimation," *IEEE Signal Processing Letters*, VOL. 16, NO. 11, NOVEMBER 2009.
- [17] Y.-M. Chen and I. V. Bajic, "A joint approach to global motion estimation and motion segmentation from a coarsely sampled motion vector field," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 21, no. 9, pp. 13161328, Sep. 2011.
- [18] R. Xu and D. Taubman, "Robust Dense Block-Based Motion Estimation Using a 2-Bit Transform on a Laplacian Pyramid," *20th Proc. IEEE Int. Conf. Image Proc.*, September 2013, submitted for publication, available at <http://dstn.ee.unsw.edu.au/~rui/2bitMotion.pdf>.
- [19] A.T. Naman and D. Taubman, "A Soft Measure for Identifying Structure from Randomness in Images," *20th Proc. IEEE Int. Conf. Image Proc.*, September 2013, submitted for publication, available at <http://dstn.ee.unsw.edu.au/~aous/pdfs/aous-2013-a.pdf>.