Bidirectional Hierarchical Anchoring of Motion Fields for Scalable Video Coding

Dominic Rüfenacht #1, Reji Mathew #2, and David Taubman #3

# *Interactive Visual Media Processing Lab (IVMP), School of EE&T, UNSW*
Australia
1 d.ruefenacht@unsw.edu.au
2 reji.mathew@unsw.edu.au
3 d.taubman@unsw.edu.au

***Abstract*—The ability to predict motion fields at finer temporal scales from coarser ones is a very desirable property for temporal scalability. This is at best very difficult in current state-of-the-art video codecs (*i.e.*, H.264, HEVC), where motion fields are anchored in the frame that is to be predicted (*target* frame). In this paper, we propose to anchor motion fields in the *reference* frames. We show how from only one fully coded motion field at the coarsest temporal level as well as breakpoints which signal discontinuities in the motion field, we are able to reliably predict motion fields used at finer temporal levels. This significantly reduces the cost for coding the motion fields. Results on synthetic data show improved rate-distortion (R-D) performance and superior scalability, when compared to the traditional way of anchoring motion fields.**

I. Introduction

The last years have shown a rapidly growing demand for consuming multimedia over networks with varying bandwidths, with a large heterogeneity of end user devices (from smartphones to high definition displays). Acceptable decoding quality can only be achieved if the server can instantaneously adapt the bit-rate. The aim of scalable video coding is to encode the video once at the highest quality level in an embedded way such that partial streams can be decoded at lower spatial resolution and temporal resolution, as well as bit-rate. For anything other than a small number of operating points, this requires the decoder to work independently from the encoder [1]. The explicit communication of motion parameters between the coder and the decoder of current state-of-the-art codecs make them ill-suited for fully scalable video coding [2]. Wavelet-based approaches using a feedforward system are a more natural way of achieving full scalability.

Various *wavelet-based scalable video coders* (WSVC) have been proposed, which mainly differ in the order the spatial and temporal wavelet decompositions are applied: spatial domain motion compensated temporal filtering ("$t+2D$") [3], in-band motion compensation ("$2D+t$") [4], and adaptive schemes [5] have all been explored. While there has been significant progress in scalable video coding, the rate-distortion performance of scalable video coders on generic video data is still inferior to their non-scalable counterparts.

One of the main obstacles to highly scalable video is scalable motion; a scalable video transform is best served by a single highly scalable motion field, yet multiresolution motion descriptions have difficulty in providing valid descriptions in the vicinity of moving object boundaries [6]. In particular, while bandlimited sampling of the texture data is a reasonable imaging model, bandlimited sampling of the accompanying motion is not actually helpful. These problems can be greatly mitigated if the geometry of the scene is correctly modelled. Recent results by Lalgudi *et al.* [7] on volume rendered images show superior compression performance compared to H.264/AVC by incorporating the underlying geometric relationship of the volumetric data into the lifting steps of a temporal wavelet transform.

Mathew *et al.* [8] propose a fully scalable representation of discontinuities in both resolution and precision. Discontinuities are determined in a rate-distortion optimization framework, and signalled using *breakpoints*. These breakpoints are used to avoid wavelet bases from crossing discontinuity boundaries. Experimental results on depth maps show how the resulting breakpoint-adaptive DWT effectively reduces the magnitude of subband samples in the vicinity of depth discontinuities, which improves compression performance.

Most attempts at scalable video coding are using block-based motion fields, which are known to have problems at discontinuities in the motion field. Young *et al.* [9] advocate a *compression regularized optical flow*, where the motion field and breakpoints are jointly discovered. The resulting piecewise smooth motion field can then be efficiently encoded using the breakpoint-adaptive DWT.

In this paper, we propose a bidirectional motion compensation framework, involving a hierarchical description of motion that is highly novel with regard to what is commonly found in the literature. Additionally, we propose methods for composing and inverting motion fields with the aid of breakpoint information, including methods for estimating motion in disoccluded regions, and resolving ambiguities in regions of motion folding.

The only similar work we are aware of is [10], which considers only uni-directional motion compensation and does

2014 IEEE 16th International Workshop on Multimedia Signal Processing (MMSP), Sep. 22-24, 2014, Jakarta, Indonesia.
978-1-4799-5896-2/14/$31.00 ©2014 IEEE.

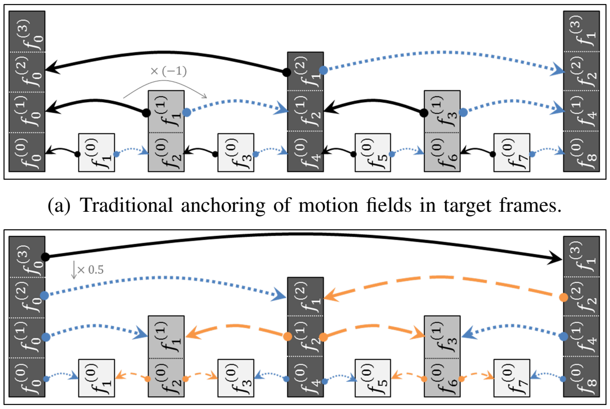
not encounter a number of the key issues that are resolved in this work.

We present the bidirectional hierarchical anchoring of motion fields in Sect. II. In Sect. III, we explain how we warp motion fields anchored at reference frames to the target frames involved in the prediction. Experimental results are presented in Sect. IV, and the paper is concluded in Sect. V.

II. Bidirectional Hierarchical Anchoring of Motion Fields

In this paper we work with the Spline 5/3 temporal wavelet. At each temporal level, this means that the odd indexed frames are predicted using the preceding and proceeding even indexed frames, while even indexed frames are updated using the prediction residuals.

In the following, a motion field *anchored* at frame i and *pointing* to frame j means that each pixel in frame i has a motion vector associated which points to a location in frame j. By anchoring motion fields at the (even indexed) reference frames as opposed to anchoring them at the (odd indexed) target frame as employed in current state-of-the-art codecs (*i.e.*, H.264, HEVC), we are able to infer motion fields at finer temporal levels from coarser level motion anchored at the same frame. This is a very desirable property for temporal scalability, and is very difficult if the motion fields are anchored in the target frames. We refer to the latter way of anchoring motion fields as *traditional anchoring*, which is essentially what is done in H.264/SVC using hierarchical B-frames. Fig. 1 shows the traditional and our proposed *bidirectional hierarchical motion field anchoring* schemes.



(a) Traditional anchoring of motion fields in target frames.

(b) Proposed anchoring of motion fields at the reference frames.

Fig. 1. Two ways of anchoring motion fields: (a) Traditional anchoring at target frames; (b) The proposed *bidirectional hierarchical anchoring* at reference frames. Solid black arrows are *full* motion fields, dotted blue are *scaled* motion fields, and dashed orange indicates *inferred* motion fields.

We use the terms *scaled* and *inferred* to refer to motion fields that are used as prediction references for motion coding. The scaled motion fields are obtained by applying a constant scaling factor to the motion vectors found in other motion fields at the same or a coarser level of the hierarchy. In the traditional approach, each target frame has an independent motion field (solid black arrows in the Fig. 1), which is scaled (typically by -1) to form a scaled reference (dotted blue arrows) for coding the reverse motion field. In the proposed scheme, the scaled motion fields are obtained by scaling coarser level motion by 0.5, as shown Fig. 1(b). As prediction references, these scaled motion fields can be expected to be most efficient under constant (non-accelerated) motion.

The *inferred* motion fields (dashed orange arrows in Fig. 1(b)) are specific to our proposed hierarchical motion anchoring scheme, being obtained through composition and inversion of other motion fields at the same and coarser levels of the hierarchy. Importantly, the inferred motion fields can be highly effective in predicting actual motion, even under accelerating conditions. Table I shows the number of each type of motion field, at each temporal level t. Evidently, in the proposed motion model roughly half of all the motion fields are inferred.

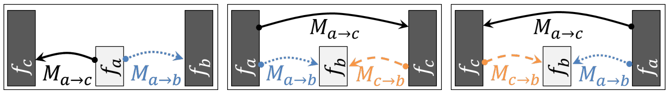
TABLE I
Required Motion Fields for a given number of temporal decompositions t.

	Full	*Scaled*	*Inferred*
Traditional anchoring	$2^t - 1$	$2^t - 1$	0
Hierarchical anchoring	1	$2^t - 1$	$2^t - 1$

We use $M_{i \to j} (= M_{f_i \to f_j})$ to denote the motion field anchored in frame f_i and pointing to frame f_j. We remind the reader that the *scaled* and *inferred* motion fields serve as references $\hat{M}_{i \to j}$ for a predictive coding scheme that encodes the actual motion field $M_{i \to j}$ as $\Delta_{M_{i \to j}} = M_{i \to j} - \hat{M}_{i \to j}$.

Clearly, the quality of these references has a large impact on the motion coding cost. Moreover, as bits are discarded from a scalable bit-stream, small prediction residuals will be quantized to zero so that the motion obtained by the scaling and inference algorithms comes to dominate the visual properties of the reconstructed video.

To facilitate the discussion, we label the frames and motion fields involved in the bi-directional prediction process at any given temporal level t as shown in Fig. 2.



(a) Trad. anchoring (b) Prop. anchoring 1 (c) Prop. anchoring 2

Fig. 2. Frame naming conventions. The target frame (in the middle) is predicted from its temporal left and right neighbour.

Both the traditional and proposed schemes involve a “full” motion field that is either independently coded, or differentially coded at a coarser temporal level. As shown in the figure this full motion field is anchored at frame labelled as f_a and points to a frame that is labelled as f_c. The other frame involved in prediction at level t is labelled f_b. Fig. 2 shows three arrangements of these frames f_a, f_b and f_c, corresponding to the traditional anchoring approach of Fig. 1(a) and the proposed anchoring of Fig. 1(b), where the latter involves two different arrangements depending on the index of the target frame.

A. *Scaling of Motion Fields*

If two motion fields are anchored at the same frame, only one needs to be fully coded, and the other one can be coded as a *scaled* version of the full reference. In particular, $M_{a\to b}$ can be predicted from $M_{a\to c}$ as:

$$\hat{M}_{a\to b} = \alpha M_{a\to c}, \tag{1}$$

where $\alpha = 0.5$ in the proposed, and $\alpha = -1$ in the traditional anchoring scheme are the most natural choices. For the remainder of this paper, we refer to these motion fields as *scaled* motion fields.

B. *Inferring of Motion Fields*

In the proposed scheme, the fact that the motion fields are anchored at reference frames allows us to notionally *infer* $M_{c\to b}$ from $M_{a\to c}$ and $M_{a\to b}$ as follows:

$$\hat{M}_{c\to b} = M_{a\to b} \circ (M_{a\to c})^{-1}. \tag{2}$$

Note that none of the motion fields are likely to be truly invertible. However, we propose a well-defined breakpoint dependent procedure for inferring these motion fields. With the aid of Fig. 3, we explain how $\hat{M}_{c\to b}$ can be obtained.

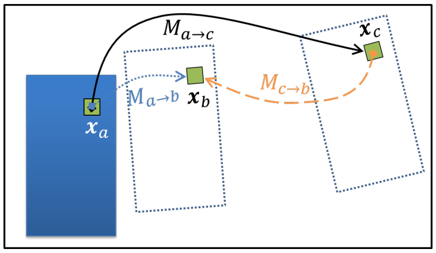


Fig. 3. Inferring $M_{c\to b}$ from $M_{a\to c}$ and $M_{a\to b}$.

For each location $\mathbf{x}_a$ in frame f_a, motion fields $M_{a\to c}$ and $M_{a\to b}$ provide corresponding locations $\mathbf{x}_c = M_{a\to c}(\mathbf{x}_a)$ and $\mathbf{x}_b = M_{a\to b}(\mathbf{x}_a)$ in frames f_c and f_b, respectively. For each $\mathbf{x}_c$ that arises in this way, these relationships provide at least one candidate for the motion $M_{c\to b}(\mathbf{x}_c) = \mathbf{x}_b - \mathbf{x}_c$. Because of occlusions at moving object boundaries, some locations $\mathbf{x}_c$ get hit multiple times, while other locations in f_c never get hit. In Sect. III-C and Sect. III-B, respectively, we explain how we resolve double mappings and avoid holes with the aid of breakpoints.

For the rest of this paper, we denote these motion fields as *inferred* motion fields. It is important to note that regardless of the accuracy of $M_{a\to c}$ and $M_{a\to b}$, the predictions of the target frame f_b formed using the inferred motion field $M_{c\to b}$ should always be geometrically consistent with those formed using $M_{a\to b}$.

III. Motion Field Warping

In the proposed scheme, we need to warp complete motion fields between frames for both: 1) inferring $M_{c\to b}$, as well as 2) obtaining $(M_{a\to b})^{-1}$ and $(M_{c\to b})^{-1}$, which are used to predict f_b. The warped motion fields then get assigned reverse motion vectors to form the "inverse" motion fields.

The main challenges during this warping process are to resolve double mappings in folded regions, as well as to avoid holes in disoccluded regions. In the following, we present a procedure which uses discontinuities in the motion field (encoded using breakpoints) to fill in reasonable information in disoccluded regions, as well as resolving ambiguities in folded regions.

A. *Cellular Affine Warping with Motion Type Identification*

We first propose a procedure to generate $M_{j\to i}$ from the available "inverse" motion field $M_{i\to j}$. For this, we completely partition $M_{i\to j}$ into small triangles in f_i, and assign each triangle an affine motion flow. In the present work, we use cell sizes of 1×1 pixels.[1]

Aliasing in regions of local contraction can be mitigated by using an upsampled representation of the warped motion field in the target frame. We extend the motion field in the reference frame by one pixel, assigning it all zero motion vectors. This guarantees that the warped triangles represent a complete distorted affine mesh in the target frame.

We label triangles according to the nature of the motion they are undergoing as they are warped from frame i to frame j, and distinguish three cases (see Fig. 4): (1) Visible in both frames (cyan), (2) disoccluded in target frame (green), and (3) folded in target frame (magenta). Such a labelling is very useful because the labels form a *disocclusion and folding map*, which can be used to guide the temporal prediction (*e.g.,* don't predict in disoccluded regions if they are visible in the other (reference) frame) and update steps (*e.g.,* skip update if disoccluded) of the temporal 5/3 Spline wavelet.

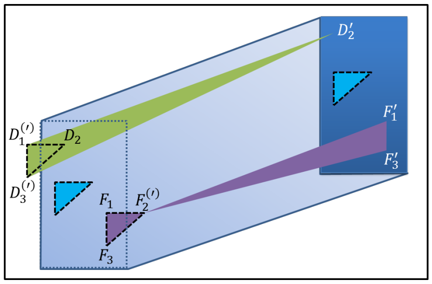


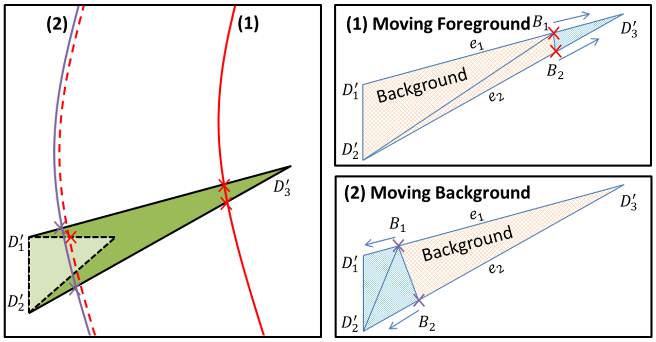
Fig. 4. Different types of warped triangles: Visible (cyan), disoccluded (green), and folded (magenta).

The triangles are labelled by observing that whenever a triangle gets warped to a disoccluded/folded region, the area of the triangle is increasing as well as it contains a breakpoint on one of its arcs in the reference frame. To differentiate between disocclusion and folding, we observe that if the orientation of the vertices of the triangle does not change, the triangle is mapped to a disoccluded region, and if the orientation changes, it is folding over. In the following, we show how the motion for triangles labelled as *disoccluded* or *folded* is determined.

[1] Clearly, this scheme could be accelerated by using larger cells in regions of smooth motion.

B. Handling of Disoccluded Regions

For disoccluded triangles, the "correct" motion to be assigned in the target frame is not observable in the reference frame. Direct application of the cellular affine warping operation described above is guaranteed not to leave any holes in the target frame, but achieves this by linearly interpolating between background and foreground motions. Unfortunately, this has a number of adverse consequences for the integrity of the resulting motion. Instead, we propose to extrapolate the background motion in regions of disocclusion. Whenever a triangle is stretched due to disocclusion, we expect that the stretched triangle intersects with a motion discontinuity in the target frame, which partitions the triangle into another triangle and a quadrilateral. With the aid of Fig. 5, we describe how we identify which part of the stretched triangle is in the background.



(a) Disoccluded triangle

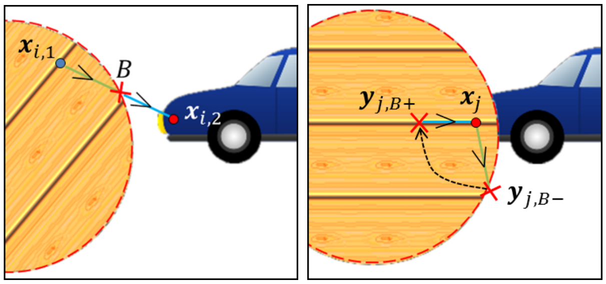
(b) Two different cases

Fig. 5. Reasoning about geometry of object boundaries is used to find out which part of the triangle lies in the background.

In Fig. 5(a), we focus on one triangle that is stretched in the target frame. The red dashed line indicates the location of the motion discontinuity in the reference frame. The red line shows where the discontinuity moves if the foreground object is moving on top of the background (case (1) in Fig. 5), while the magenta line assumes that the foreground is still and the background is moving (case (2) in Fig. 5). We observe that the motion discontinuity boundary "moves" with the foreground object. The procedure of identifying which part is background tries to intersect the three lines of the triangle formed by connecting any two of the three vertices of the triangle in the target frame with motion discontinuities; if such an intersection is found, we record which vertex lies closer to the intersection (indicated by blue arrows in Fig. 5(b)). Two edges of the triangle should intersect with motion discontinuities (i.e., breaks). Let e_1 and B_1 denote the first such edge and break location and e_2 and B_2 the second such edge/break pair, writing D'_3 for the vertex that shares these two edges. As shown in Fig. 5(b), one of two situations should occur: 1) D'_3 is closer to both B_1 and B_2 than the other vertex on edges e_1 and e_2, respectively; or 2) D'_3 is further from both B_1 and B_2 than the other vertex on each of the respective edges. In both situations, the motion of D'_3 is extrapolated in the triangle formed by D'_3, B_1, and B_2. The quadrilateral is broken up into two triangles (D'_1, D'_2, $B1$), and (D'_2, B_1, $B2$), and the motion of D'_1 and D'_2 is extrapolated in the respective triangles.

C. Handling of Folded Regions

As the cellular affine warping process visits triangles in f_i in order to map motion vectors from $M_{i\to j}$ into frame f_j, it can happen that a location $\mathbf{x}_j$ already has an assigned motion. That is, there are two locations $\mathbf{x}_{i,1}$ and $\mathbf{x}_{i,2}$ such that $M_{i\to j}(\mathbf{x}_{i,1})$ and $M_{i\to j}(\mathbf{x}_{i,2})$ are both equal to $\mathbf{x}_j$. This occurs as a result of folding in the motion field. Our proposed approach for disambiguating these double mappings is based on the observation that the motion discontinuity "moves" with the foreground object. Fig. 6 illustrates the proposed procedure.



(a) Reference frame f_i

(b) Target frame f_j

Fig. 6. A disk moves on top of a stationary car. Two points in the reference frame ($\mathbf{x}_{i,1}$ and $\mathbf{x}_{i,2}$) map to the same point $\mathbf{x}_j$ in the target frame. Breakpoints are used to identify the foreground moving object.

Consider the line segment that connects $\mathbf{x}_{i,1}$ with $\mathbf{x}_{i,2}$ in f_i. This line has to intersect with (at least) one motion discontinuity. In our example, there is just one intersection (B) at the motion boundary between foreground (*i.e.,* the disk) and background (*i.e.,* the stationary car and white background) objects. Let $\mathbf{y}_{i,s} = (1-s)\mathbf{x}_{i,1} + s\mathbf{x}_{i,2}$ be a parametrisation of the points on this line segment, where $s \in [0,1]$, and consider the behaviour of $\mathbf{y}_{j,s} = M_{i\to j}(\mathbf{y}_{i,s})$ as s transitions from 0 to 1. When $\mathbf{y}_{i,s}$ arrives at the break location $\mathbf{y}_{i,B}$, the mapped location $\mathbf{y}_{j,s}$ will exhibit a discontinuous jump (black dotted arrow in Fig. 6(b)). Let $\mathbf{y}_{i,B-}$ and $\mathbf{y}_{i,B+}$ denote the locations immediately before and after the break, having mapped locations $\mathbf{y}_{j,B-}$ and $\mathbf{y}_{j,B+}$. We expect one of these two mapped locations to align (at least very closely) with a break location in the target frame f_j. Accordingly, we conclude that $M_{j\to i}(\mathbf{x}_j) = \mathbf{x}_{i,1} - \mathbf{x}_j$ if $\mathbf{y}_{j,B-}$ lies closer to a motion discontinuity in the target frame than $\mathbf{y}_{j,B+}$, else $M_{j\to i}(\mathbf{x}_j) = \mathbf{x}_{i,2} - \mathbf{x}_j$. In practice, we measure distances from motion boundaries through a local one–dimensional search along the lines that run between $\mathbf{x}_j$ and each of $\mathbf{y}_{j,B-}$ and $\mathbf{y}_{j,B+}$.

IV. Experimental Results and Discussion

In this section, we present qualitative and quantitative results of the proposed method. We focus on synthetically generated scenes in order to control the type and amount of motion, which allows us to better study the behaviour of the transform.

Fig. 7 shows two frames of the synthetic sequence (800 × 512 pixels) used in the discussion. A rotating ball is translating north and east, and a plane is moving west and south, intersecting with the ball from frame 2 onward. The motion of both objects is *accelerated*, which results in an error in prediction of the *scaled* motion fields. Nevertheless, predictive coding of the motion with respect to the scaled fields is still beneficial.

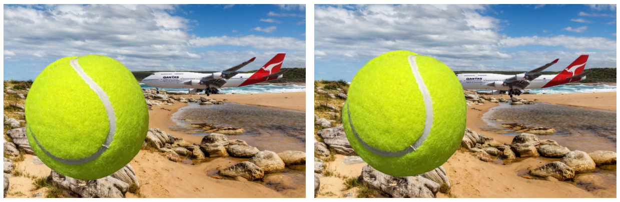

(a) Frame 0
(b) Frame 4

Fig. 7. Synthetic sequence used in the discussion.

A. Qualitative Evaluation

The interesting motion fields are the *inferred* ones, which do not rely on a specific motion model. Fig. 8 shows the ground truth and *inferred* motion field $\hat{M}_{4\to2}$, which is obtained by first warping $M_{0\to4}$ to frame 4, and then using $M_{0\to2}$ to infer the motion vectors $\hat{M}_{4\to2}$ (*e.g*, $\hat{M}_{4\to2} = M_{0\to2} \circ (M_{0\to4})^{-1}$).

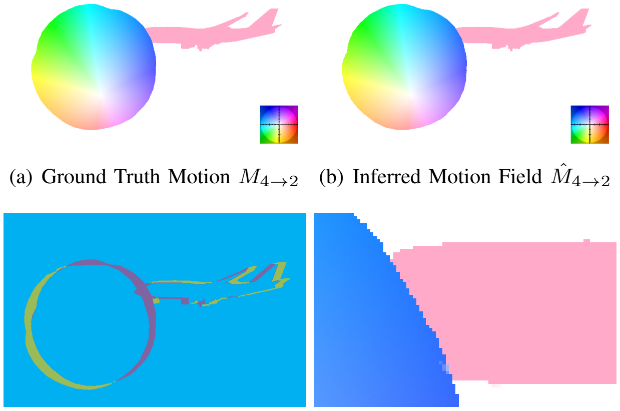

(a) Ground Truth Motion $M_{4\to2}$
(b) Inferred Motion Field $\hat{M}_{4\to2}$
(c) Folding and Disocclusion Map
(d) Crop of (b)

Fig. 8. Example of an *inferred* motion field. The backward pointing motion field $M_{4\to2}$ is *inferred* from the two forward-pointing motion fields $M_{0\to4}$ and $M_{0\to2}$. (c) shows the disocclusion and folding map obtained by warping $M_{0\to4}$ to frame 4, using the same colour scheme as in Fig. 4.

We can see that the proposed warping scheme correctly extrapolates the background motion in disoccluded regions (green). Double mappings (magenta) are also resolved correctly most of the times. Of particular interest is the region where the plane gets covered by the ball (see Fig. 8(d)), where the double mappings are accurately resolved.

As explained in Sect. III-A, during the cellular affine warping process, we extrapolate the motion of the background object in disoccluded regions. Fig. 9 shows an example where this does not assign the “correct” motion.

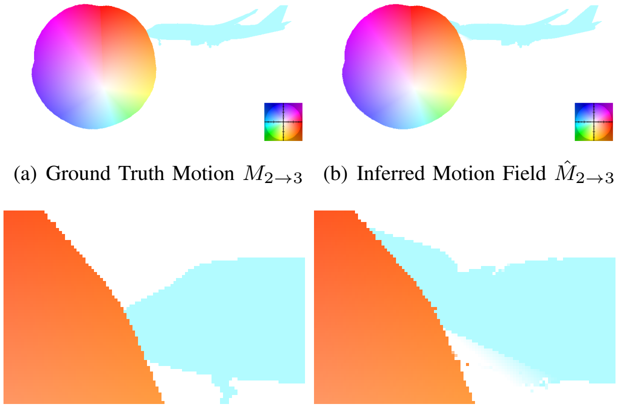

(a) Ground Truth Motion $M_{2\to3}$
(b) Inferred Motion Field $\hat{M}_{2\to3}$
(c) Crop of (a)
(d) Crop of (b)

Fig. 9. An *inferred* motion field where there is a new motion appearing in the disoccluded region, which is not handled properly by the proposed motion extrapolation.

In order to obtain the *inferred* motion field $\hat{M}_{2\to3}$, we need to warp $M_{4\to2}$ to frame 2. This means that the plane and the ball are moving away from each other, which unveils the background (with a different motion), which is not seen in the reference frame 4. We can see in Fig. 9(d) that the procedure behaves as it should, extrapolating the motion of the plane, even though this turns out not to be the correct motion, due to the complexity of the covering and uncovering processes involved.

B. Quantitative Results

We present R-D graphs for the proposed bidirectional hierarchical anchoring of motion fields, and compare it with the traditional way of anchoring motion fields. Note that any block-based motion estimation strategy would have problems at object boundaries as well as representing the rotation of the ball. We therefore use the ground truth motion data for both anchoring schemes. For the proposed hierarchical as well as the traditional way of anchoring motion fields, the sample data is compressed using two levels of temporal decomposition. The temporal subband frames as well as the differentially coded motion fields are then subjected to a five level spatial DWT, followed by embedded block coding of the quantized wavelet coefficients. Motion and temporal subband frames are coded using JPEG2000, while breakpoints are coded using the method described in [8]. Note that the motion DWT is breakpoint adaptive. Fig. 10 shows the observed rate-distortion performance, expressed in terms PSNR – only the luminance component of the video is processed in these experiments.

The horizontal axis in the figure corresponds to the number of bits decoded for a group of five frames. Since all elements of the coded representation are fully scalable, the motion, breakpoint and texture subband data can all be truncated independently to obtain a huge family of potential operating points for a decoder. In this work we consider just three different operating points for the motion, while sweeping the rate associated with the texture data to generate a set of R-D curves. We label the motion operating points as “low”, "medium" and "high", having made an effort to ensure that these labels are consistent between the two motion anchoring schemes.

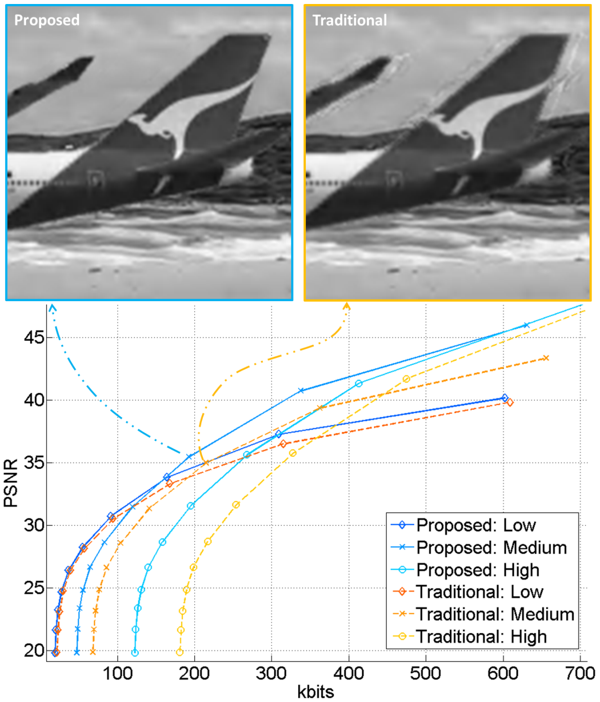


Fig. 10. R-D comparison of the proposed *bidirectional hierarchical anchoring* (blue curves) of motion fields and the traditional way (orange curves). The graph shows for both schemes three curves which are obtained by fixing the motion quality to low (diamonds), medium (crosses), and high (circles).

The fact that we can predict the *inferred* motion fields very well means that we have much less motion to code than in the traditional setup, which results in better R-D performance. The two crops of a reconstructed frame 3 in Fig. 10 show an example for the same texture quality, almost identical PSNR, as well as comparable motion field qualities. Even though our scheme operates at a lower bitrate, it contains many fewer artifacts in the vicinity of moving object boundaries, which shows the scalability of motion.

An interesting point to highlight is that the proposed scheme is able to create a highly credible reconstruction at frame rates higher than what was available at the encoder. In that case, the *scaled* motion fields will assume constant motion between frames and the proposed scheme "invents" a frame that sits in between the coarser level parent frames. Importantly, the *inferred* motion field will always be geometrically consistent with the *scaled* one, so that credible motion interpolated predictions are naturally formed at any point in time, even if the actual motion is not known. For this to work, we have to warp breakpoint information from the frames that are available to those that are being interpolated. This can be done using the breakpoint warping scheme proposed in [10].

While the primary focus of this paper has been that of breakpoint-assisted motion warping, we would like to stress here that the experimental results have been obtained in the context of a fully motion compensated temporal wavelet transform, involving motion compensated temporal prediction and update steps, each of which is spatially adapted in accordance with the disocclusion and folding information that is deduced during the motion warping process.

V. Conclusions and Future Work

This paper proposes a hierarchical anchoring scheme for motion fields that facilitates scalable and efficient bi-directional motion compensation, for use with motion compensated temporal wavelet transforms. Hierarchical bi-directional motion compensation schemes involve roughly twice as many motion fields as original frames. The primary benefit of the proposed motion anchoring scheme is that half of these can be *inferred* in a manner that preserves geometric consistency with the remaining motion fields, most of which can be reasonably predicted by scaling coarser level motion. The paper proposes a robust technique to resolve double mappings in regions where the motion field is folding and a method for extrapolating background motion in *disoccluded* regions, by leveraging an underlying motion description that is piecewise smooth, mediated by the scalable breakpoint description in [8]. Experimental results suggest that the *inferred* motion fields can be of very high quality, even with complex motion in multi-layered scenes. Even though these *inferred* fields are used only as prediction references for the actual motion, they are of sufficient quality that the prediction errors are best omitted at all but the very highest bit-rates, which effectively halves the overall motion field cost.

In future work, we plan to investigate a hierarchical breakpoint warping scheme, which is expected to further improve the scalability attributes of the proposed scheme.

References

[1] H. Schwarz, D. Marpe, and T. Wiegand, "Overview of the Scalable Video Coding Extension of the H.264/AVC Standard," *IEEE Trans. on Circ. and Syst. for Vid. Tech.*, vol. 17, no. 9, pp. 1103–1120, 2007.

[2] T. Sikora, "Trends and Perspectives in Image and Video Coding," *Proc. of the IEEE*, vol. 93, 2005.

[3] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression." *IEEE Trans. on Im. Proc.*, vol. 12, pp. 1530–42, 2003.

[4] Y. Andreopoulos, A. Munteanu, J. Barbarien, M. Van der Schaar, J. Cornelis, and P. Schelkens, "In-band motion compensated temporal filtering," *Sig. Proc.: Im. Comm.*, vol. 19, no. 7, pp. 653–673, Aug. 2004.

[5] N. Mehrseresht and D. Taubman, "An efficient content-adaptive motion-compensated 3-D DWT with enhanced spatial and temporal scalability." *IEEE Trans. on Im. Proc.*, vol. 15, no. 6, pp. 1397–412, 2006.

[6] N. Adami, A. Signoroni, and R. Leonardi, "State-of-the-Art and Trends in Scalable Video Compression With Wavelet-Based Approaches," *IEEE Trans. on Circ. and Syst. for Vid. Tech.*, vol. 17, no. 9, pp. 1238–1255, 2007.

[7] H. G. Lalgudi, M. W. Marcellin, A. Bilgin, H. Oh, and M. S. Nadar, "View compensated compression of volume rendered images for remote visualization." *IEEE Trans. on Im. Proc.*, vol. 18, no. 7, pp. 1501–11, 2009.

[8] R. Mathew, D. Taubman, and P. Zanuttigh, "Scalable coding of depth maps with R-d optimized embedding." *IEEE Trans. on Im. Proc.*, vol. 22, no. 5, pp. 1982–95, 2013.

[9] S. Young, R. Mathew, and D. Taubman, "Joint estimation of motion and arc breakpoints for scalable compression," *IEEE Global Conf. on Signal and Information Proc. (Global SIP)*, 2013.

[10] D. Rüfenacht, R. Mathew, and D. Taubman, "Hierarchical Anchoring of Motion Fields for Fully Scalable Video Coding," *IEEE Int. Conf. on Im. Proc. (ICIP)*, Oct. 2014.