LIFTING-BASED ILLUMINATION ADAPTIVE TRANSFORM (LIAT) USING MESH-BASED ILLUMINATION MODELLING

Maryam Haghighat, Reji Mathew, Aous Naman, and David Taubman

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

ABSTRACT

State-of-the-art video coding techniques employ block-based illumination compensation to improve coding efficiency. In this work, we propose a Lifting-based Illumination Adaptive Transform (LIAT) to exploit temporal redundancy among frames that have illumination variations, such as the frames of low frame rate video or multi-view video. LIAT employs a mesh-based spatially affine model to represent illumination variations between two frames. In LIAT, transformed frames are jointly compressed, together with illumination information, into a layered rate-distortion optimal codestream, using the JPEG2000 format. We show that the LIAT framework significantly improves compression efficiency of temporal subband transforms for both predictive and more general transforms with predict and update steps.

***Index Terms*—** Illumination compensation, Scalable video coding, Affine mesh, Wavelet-based coding.

1. INTRODUCTION

For video compression, inter-frame prediction is critical for exploiting temporal redundancy between frames. A prediction algorithm that can exploit illumination changes among frames, such as lighting changes over the course of the day in low frame-rate video scenarios, can be expected to achieve better coding efficiency than algorithms that are not illumination adaptive. For this reason, illumination-aware compensation has been explored by many researchers [1–5], and is supported in H.264 [6,7] and HEVC [8] standards, in the form of *weighted prediction*.

Illumination change is usually modelled by a scale and an offset [1–8]; in this model, prediction is generated by scaling the sample values of the reference frame and adding an offset to them. Kamikura and Watanabe [1] modified the MPEG-4 video codec to include a global illumination compensation mode. Both [2] and [3] estimate illumination changes in the macroblock being predicted at the target frame from illumination changes in causally adjacent macroblocks compared to their reference macroblocks. In [4] block-based illumination compensation is employed to improve coding efficiency of multi-view video.

The H.264 video coding standard [6] incorporates weighted prediction [7]; scale and offset parameters can be specified per frame or per slice. A slightly modified weighted prediction is available in HEVC [8].

All this prior work employs block-based models to represent illumination variations. Such models attempt to represent a mostly smooth illumination surface (or field) using a piecewise constant representation with artificially-positioned discontinuities; this is inherently problematic, and can produce abrupt and visually-annoying changes in illumination across block boundaries [9]. Furthermore, block-based coding approaches are less amenable to highly scalable coding representations, such as the ones presented in [10] and [11]. In this work we employ a mesh-based spatially affine model which can better represent slowly varying surfaces such as illumination and is more suited for highly scalable coding.

Prior work has mostly focused on utilizing illumination compensation within a framework that employs temporal *prediction only* schemes, ignoring the potential benefits of richer transforms that include a temporal update step. In this study we compare temporal transforms that are *prediction only* with transforms that employ both predict and update steps and show that the update step improves coding efficiency.

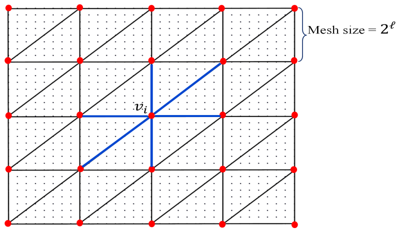
To provide a highly-scalable representation of the encoded frames, we encode illumination information and illumination-compensated, temporally-transformed frames into a layered embedded codestream where each layer is a rate-distortion optimal representation of the original frames at that layer's rate.

This work primarily focuses on low frame rate applications, typically found in surveillance video where a frame is captured for example at intervals of 5 minutes or greater. In such applications cameras are typically stationary and the prevalent motion in resulting video sequences can be classified as being of two types. One is global motion due to small camera motions (i.e. camera shaking) and the other is due to scene object motion. The former can be compensated as a preprocessing step as done in this paper while the latter is not amenable to exploitation due to the large temporal frame separation.

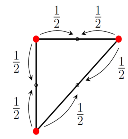
The contributions of this work are: (i) a lifting-based illumination adaptive transform (LIAT) to exploit inter-frame redundancy in the presence of illumination variations; (ii) modelling illumination changes with a mesh-based affine model which does not impose artificial block boundaries; and (iii) a highly scalable, embedded compression framework with rate-distortion optimal termination points for frames and illumination information.

2. ILLUMINATION MODELLING

As in prior work, we employ a scale and an offset to model the change in illumination between two frames; however, we use a mesh-based spatially affine representation instead of a block-based representation, as explained before. The spatially affine transform with corresponding scale α and offset β parameters is applied to a reference frame f_0 to produce an illumination compensated estimate



(a) Affine mesh



(b) Affine interpolation

Fig. 1. Affine mesh model. The 6 edges incident to a vertex v_i are highlighted in (a).

$\tilde{f}_1$ of the target frame f_1.

$$\tilde{f}_1[\mathbf{x}] = \alpha[\mathbf{x}] \cdot f_0[\mathbf{x}] + \beta[\mathbf{x}] \quad (1)$$

The uniformly spaced mesh structure is illustrated in Fig. 1a, where vertices (red dots) and edges (solid lines) of the resulting graph are shown. Within a given triangle, α and β values at each pixel location are derived by interpolating the α and β values defined at the three vertices of the triangle. An example of affine interpolation for half grid locations is shown in Fig. 1b. Estimation of the vertex illumination parameters is discussed in Section 5. From a compression perspective it is important that the representation of α and β be more compact than the representation of the frames. In our proposed scheme we achieve this by (i) using a coarse mesh to model α and β, thereby limiting the number of values needed for their representation; and (ii) incorporating regularization into the estimation algorithm described in Section 5. It should be noted that, in this work, we only consider the luminance component "Y", but the proposed approach can be extended to full-color representation, in which α and β take the form of a matrix at each pixel location.

3. LIFTING-BASED ILLUMINATION ADAPTIVE TRANSFORM

The proposed illumination adaptive temporal transform is realized through a sequence of two lifting steps, a predict step and an update step, which are given by

$$h[\mathbf{x}] = f_1[\mathbf{x}] - P(f_0[\mathbf{x}]) \quad (2)$$
$$l[\mathbf{x}] = f_0[\mathbf{x}] + b[\mathbf{x}]h[\mathbf{x}] \quad (3)$$

Here $P(\cdot)$ represents an illumination compensation operator and h and l refer to the high-pass and low-pass subband frames with b corresponding to the update factor. In this work, the illumination compensation operator $P(f_0[\mathbf{x}])$ is given by

$$P(f_0[\mathbf{x}]) = \alpha[\mathbf{x}]f_0[\mathbf{x}] \quad (4)$$

Although we find β to be useful for estimation, we choose not to code it separately as it is an additive offset to the residue. If β is included in (4) then separate code streams would be required for the residue data h and the additive offset β; from a compression perspective this arrangement is clearly wasteful. For this reason, β is not included in the illumination compensation operator.

Here, we move our attention to finding $b[\mathbf{x}]$. The synthesis steps that correspond to the analysis steps in (2) and (3) are

$$f_0[\mathbf{x}] = l[\mathbf{x}] - b[\mathbf{x}]h[\mathbf{x}]$$
$$f_1[\mathbf{x}] = h[\mathbf{x}] + P(f_0[\mathbf{x}])$$

It is straightforward to calculate the synthesis energy gain for the high-pass subband frame h. For a given location $\mathbf{x}$, let $h[\mathbf{x}] = 1$ and $l[\mathbf{x}] = 0$; using the synthesis steps above $f_0[\mathbf{x}] = -b[\mathbf{x}]$ and $f_1[\mathbf{x}] = 1 - \alpha[\mathbf{x}]b[\mathbf{x}]$. The squared summation of these terms provide the synthesis energy gain $G_h[\mathbf{x}]$ as

$$G_h[\mathbf{x}] = (b[\mathbf{x}])^2 + (1 - \alpha[\mathbf{x}]b[\mathbf{x}])^2$$

In this work we follow the approach pioneered by Girod and Han in [12] to choose an update factor $b[\mathbf{x}]$ so as to minimize $G_h[\mathbf{x}]$; this results in the following expression.

$$b[\mathbf{x}] = \frac{\alpha[\mathbf{x}]}{1 + \alpha[\mathbf{x}]^2}$$

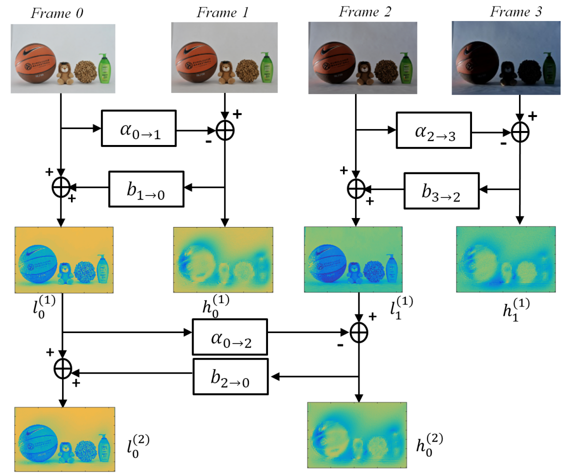


Fig. 2. The proposed illumination adaptive temporal transform

Fig. 2 shows the proposed LIAT scheme with two levels of temporal decomposition. Illumination fields $\alpha_{0\to1}$, $\alpha_{0\to2}$, and $\alpha_{2\to3}$ as well as temporal subband frames $l_0^{(2)}$, $h_0^{(1)}$, $h_0^{(2)}$, and $h_1^{(1)}$ undergo spatial wavelet decomposition and embedded block coding to generate highly scalable codestreams for both texture and illumination information. For convenience we denote these illumination fields and subband frames by y_s and refer to them as "local product images."

4. RATE-DISTORTION OPTIMIZATION

Each local product image y_s is encoded into a highly scalable embedded codestream with finely-spaced termination points $\theta_{s,i}$; by design, all termination points $\theta_{s,i}$ belong to the rate-distortion convex hull of y_s. Each termination point $\theta_{s,i}$ increments the codestream length by $\Delta R_{s,i} = R_{s,i} - R_{s,i-1}$ bytes, where $R_{s,i}$ is the codestream length at the end of termination point $\theta_{s,i}$. This additional data decrements reconstruction distortion of the codestream by $\Delta D_{s,i} = D_{s,i-1} - D_{s,i}$, where $D_{s,i} = \|y_s - \hat{y}_{s,i}\|_2^2$ is the distortion in a reconstructed local product image $\hat{y}_s$ using codestream data up to termination point $\theta_{s,i}$. Thus, each termination point has an associated distortion-length slope $\Delta_{s,i} = \Delta D_{s,i}/\Delta R_{s,i}$.

We denote an original frame by f_n, and the distortion in a reconstructed frame $\hat{f}_n$ by D_n, where $D_n = \|f_n - \hat{f}_n\|_2^2$. To model the contribution of distortion D_s in a local product image y_s on the global distortion $D = \sum_n D_n$, we employ an additive distortion model with energy gain factors G_s. These gain factors G_s express the amount by which quantization errors in local product images y_s are amplified in the synthesis (or reconstruction) process. With this model, global distortion D can also be written as $\sum_s G_s D_s$. The additive distortion model is valid if errors in reconstructed frames $\hat{f}_n$

arise from quantization errors that are independent; while statistical independence does not hold exactly, this model has been found to work quite well in practice, and is widely employed in the literature.

For the case of illumination compensation, the reconstruction process is not completely linear, so additional approximations are required in order to work with the additive model. Specifically, the reconstruction of any given frame f_n has many terms that are products of one or more illumination fields and subband frames. Local linearization is required, and this produces gain factors G_s that are signal dependent and therefore spatially varying. To see this, consider the reconstruction of frame f_2, shown in Fig. 2; f_2 is reconstructed from a few terms, one of which is $b_{3\to 2} h_1^{(1)} = \alpha_{23} \cdot h_1^{(1)}/(1+\alpha_{23}^2)$, where α_{23} is a shorthand for $\alpha_{2\to 3}$. The contribution of this term to the global reconstruction distortion D (remember that $D = \sum_n D_n$), is approximately given[1] by

$$\sum_{\mathbf{x}} \Bigg[\underbrace{\frac{\hat{\alpha}_{23}\hat{h}_1^{(1)}}{1+\hat{\alpha}_{23}^2} - \frac{\alpha_{23} h_1^{(1)}}{1+\alpha_{23}^2}}_{g\left(\hat{\alpha}, \hat{h}_1^{(1)}\right)} \Bigg]^2$$

Using the Taylor series expansion to linearize the two variable function $g(\hat{\alpha}, \hat{h}_1^{(1)})$ around $(\alpha_{23}, h_1^{(1)})$, we get

$$\sum_{\mathbf{x}} \left[\frac{1-\alpha_{23}^2}{(1+\alpha_{23}^2)^2} h_1^{(1)} (\hat{\alpha}_{23} - \alpha_{23}) + \frac{\alpha_{23}}{1+\alpha_{23}^2} (\hat{h}_1^{(1)} - h_1^{(1)}) \right]^2 \tag{5}$$

Evidently, the contribution of quantization errors in α_{23} depends on the spatial variations in $h_1^{(1)}$, and vice versa; in general, quantization errors in subband frames depend on, and are dependent on, spatial variations in illumination fields. To address this, our energy gain factor are obtained by taking expectations (averaging) over entire frames; it is possible to take averages over smaller domains, but we choose not to pursue this option here. In the example above, (5) is simplified by exploiting two approximations, employing spatial averages and assuming that quantization errors among subband frames and illumination fields are independent. Together, these yield the following contributions to the overall distortion D:

$$D_{\alpha_{23}} \cdot \underbrace{\frac{1}{|\mathbf{x}|} \sum_{\mathbf{x}} \left(\frac{1-\alpha_{23}^2}{(1+\alpha_{23}^2)^2} h_1^{(1)} \right)^2}_{\text{part of } G_{\alpha_{23}}} + D_{h_1^{(1)}} \cdot \underbrace{\frac{1}{|\mathbf{x}|} \sum_{\mathbf{x}} \left(\frac{\alpha_{23}}{1+\alpha_{23}^2} \right)^2}_{\text{part of } G_{h_1^{(1)}}}$$

Here, $|\mathbf{x}|$ is the number of pixels in a reconstructed frame $\hat{f}_n$. The expression above reveals contributions to two of the energy gain factors, $G_{\alpha_{23}}$ and $G_{h_1^{(1)}}$, in the additive model. Following this linearization and averaging process, we analyze all contributions to D, assembling respective contributions to the energy gain factors G_s.

Thus, the global distortion-length slopes $\Delta_{s,i}^g$ associated with local product image y_s and termination point $\theta_{s,i}$ are scaled values of their local counterparts, $\Delta_{s,i}$; i.e., $\Delta_{s,i}^g = G_s \Delta_{s,i}$.

The rate-allocation problem is then solved by "approximately" minimizing a Lagrangian cost functional $J(\lambda_j) = D(\lambda_j) + \lambda_j \sum_s R_s(\lambda_j)$ at progressively decreasing Lagrange multiplier values λ_j, each corresponds to one quality layer in the generated highly scalable codestream. Quality layer j for the highly scalable codestream of y_s contains all contributions $\Delta R_{s,i}$, possibly none, for which $\lambda_{j-1} > G_s \Delta_{s,i} \geq \lambda_j$. Many approaches exist for choosing a progression for λ_j, but discussing them is beyond the scope of this work.

[1] Here, we have assumed that quantization errors in a given decoded subband frame are uncorrelated with quantization errors in another subband frame.

5. SCALE AND OFFSET ESTIMATION

For a mesh with uniform grid spacing of 2^ℓ, the total number of vertices $V = \lceil \frac{M}{2^\ell} \rceil \cdot \lceil \frac{N}{2^\ell} \rceil$, where M and N are the frame dimensions. We write $\mathbf{f}_1$ and $\mathbf{f}_0 \in \mathbb{R}^{MN\times 1}$ for the lexicographically ordered column vector representation of current and previous frames, respectively. Similarly, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta} \in \mathbb{R}^{V\times 1}$ represent column vectors of the illumination scale and offset parameters at vertex locations of the affine mesh. Applying the illumination compensation transform to the prior frame $\mathbf{f}_0$, we obtain an estimate of the current frame $\tilde{\mathbf{f}}_1$ as shown below

$$\tilde{\mathbf{f}}_1 = \mathbf{D_0}\, \mathbf{A}\boldsymbol{\alpha} + \mathbf{A}\boldsymbol{\beta}$$

where $\mathbf{D_0} \in \mathbb{R}^{MN\times MN}$ is a diagonal matrix with the diagonal entries containing the elements of $\mathbf{f}_0$ (i.e. $(\mathbf{D_0})_{ii} = (\mathbf{f}_0)_i$) and $\mathbf{A} \in \mathbb{R}^{MN\times V}$ represents the sparse affine interpolation matrix. We seek to estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, such that the distortion between $\mathbf{f}_1$ and $\tilde{\mathbf{f}}_1$ is minimized subject to smoothness constraints on the illumination parameters. In particular, we are interested in minimizing the Lagrangian cost C given by

$$C = \left\| \mathbf{f}_1 - \tilde{\mathbf{f}}_1 \right\|_2^2 + \gamma \left(\boldsymbol{\alpha}^T \mathbf{L} \boldsymbol{\alpha} + \boldsymbol{\beta}^T \mathbf{L} \boldsymbol{\beta} \right) \tag{6}$$

The smoothness condition is imposed by the Laplacian matrix $\mathbf{L}$. For the simple graph defined by the affine mesh shown in Fig. 1, the degree of a vertex v_i, denoted by $\deg(v_i)$, is equal to 6. The elements of $\mathbf{L}$ are given by

$$\mathbf{L}_{ij} = \begin{cases} \deg(v_i) = 6 & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i \text{ and } v_j \text{ are adjacent} \\ 0 & \text{otherwise} \end{cases}$$

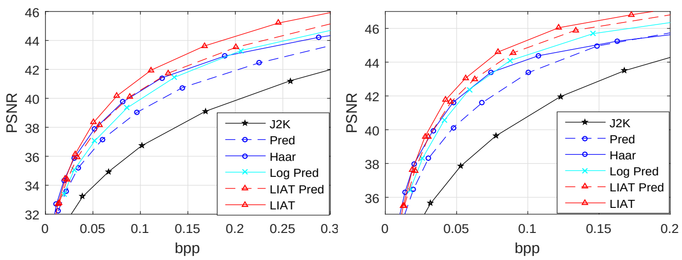
Although we employ the squared error regularization term in (6), alternative regularization schemes that encourage sparsity can be better suited to compression. We have investigated using the ℓ_1-norm for regularization, employing the iteratively reweighted least squares method to solve the optimization problem. For coarse mesh sizes (typically 32×32 or 64×64) we find that ℓ_1 regularization has little impact, so we choose to report results here only for the simplest formulation in (6).

As explained in Section 3, even though the β parameters are not included in the illumination compensation operator (4), we find it very beneficial to include β in the estimation phase as shown above. This is principally because the distortion measure that we employ, $\|\mathbf{f}_1 - \tilde{\mathbf{f}}_1\|_2^2$, is very sensitive to offsets. To discover meaningful α values, the freedom to introduce an offset to the illumination model is extremely useful.

6. EXPERIMENTAL RESULTS

Several image sequences that include various illumination conditions are tested to explore the performance of the proposed LIAT framework and to provide insight into the effectiveness of the update step. The sequences identified as *Scene 1* and *Scene 11* in [13–15] and shown in Fig. 3 correspond to indoor scenes with directional, non-uniform, illumination variations. The sequences *Building* and *National Mall-1* in Fig.3 correspond to outdoor scenes and are typical of low frame rate, surveillance video applications. The *Building*

Fig. 3. Sequences Scene 1, Scene 11, Building and National Mall-1.



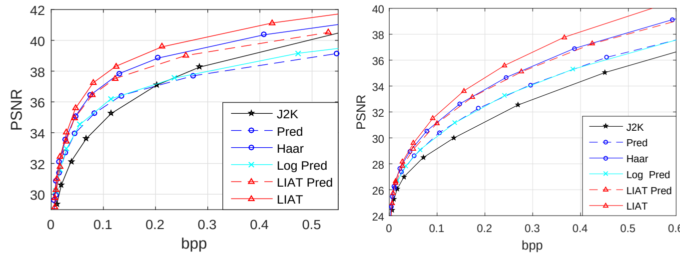
(a) Scene 1 seq. (b) Scene 11 seq.

Fig. 4. R-D results for the indoor images.

sequence, which we recorded ourselves, corresponds to one frame being captured every 15 minutes. The *National Mall-1* sequence [16, 17] (recorded on 08/10/2013) corresponds to a rate of one frame every 30 minutes. R-D performance results for coding the indoor and outdoor image sequences are shown in Fig. 4 and 5 respectively.

Curves labelled *"J2K"* relate to independent coding of each frame with a JPEG 2000 encoder that employs the 5/3 wavelet transform with 5 levels of spatial decomposition. All remaining curves correspond to schemes that employ either the predict step or both the predict and update steps in the temporal direction, to achieve 2 levels of temporal decomposition. In all cases, the resulting temporal subbands are coded with a JPEG 2000 encoder using the same configuration as that described earlier for the "*J2K*" case. The curves labelled "*Pred*" represent results in which the prior (even indexed) frame is used as a prediction for the current (odd indexed) frame. This corresponds to modelling illumination change as an offset (β) at every pixel, with the offset value given by the coded residue. The *"Pred"* scheme is then extended with an update step to implement the Haar transform; the corresponding curves are labelled *"Haar"* in each of the graphs. For further reference, we also show results of performing prediction in the log domain - curves labelled "*Log Pred*". Log domain processing, also known as "homomorphic" processing, is a well known strategy for separating slowly varying illumination effects from other aspects of a scene. Ideally, in the log domain the prediction residual between two frames with only illumination changes should be the log of the illumination change itself (i.e., $\log(\alpha_{0\to1}) = \log(f_1) - \log(f_0)$), which should be highly compressible. In practice, however, other effects such as tone mapping, reflections and flare make it difficult to reach this ideal situation. Results relating to the proposed framework are labelled *"LIAT"*; to help gauge the impact of the update step in the proposed scheme, results where the update step is omitted are also shown and these are labelled "*LIAT Pred*".

The superior performance of "*LIAT Pred*" in comparison to other *prediction only* schemes "*Pred*" and "*Log Pred*" demonstrates the effectiveness of our proposed mesh-based model and accompanying illumination adaptive prediction. By introducing an update step to "*Pred*", we obtain the results for *"Haar"*; in many cases an improvement of 1dB or more is observed. Similarly, when comparing "*LIAT Pred*" with *"LIAT"* the advantage of incorporating the update step is clearly evident. This is an important finding considering that prior work has only considered *predict only* transforms.



(a) Building seq. (b) National Mall-1 seq.

Fig. 5. R-D results for the outdoor images.

Finally, we note that the proposed LIAT framework remains superior to all other methods considered; achieving valuable gains for the two outdoor sequences that are typical of low frame rate video surveillance applications. Further results provided in Table 1, illustrate the advantage of the proposed LIAT framework. These results relate to two additional indoor sequences from [13–15] identified as *Scene 2* and *Scene 6;* and two outdoor sequences *National Mall-2* (record date of 09/2013) and *Oslo - Linpro AS* (record date of 03/2016) [16, 17].

It should be noted that all results relate to a mesh size of 64×64; we explored finer mesh sizes and in general found that coarser mesh sizes (e.g. 32×32 or 64×64) were better suited to modelling predominantly smooth illumination variations.

Table 1. R-D results for further image sequences

Sequence	bpp	J2K	Haar	LIAT Pred	LIAT
scene 2	0.05	34.50	37.35	36.68	**37.75**
	0.1	37.35	39.95	39.27	**40.45**
	0.2	40.81	42.55	42.30	**43.30**
scene 6	0.05	40.20	41.80	42.20	**42.50**
	0.1	42.43	43.48	44.10	**44.45**
	0.2	44.50	44.85	45.75	**46.25**
National Mall-2	0.1	30.15	32.40	31.95	**32.65**
	0.2	32.25	35.15	34.80	**35.60**
	0.4	35.45	38.45	38.25	**39.20**
Oslo - Linpro AS	0.1	33.50	35.75	35.60	**36.20**
	0.2	32.20	39.27	39.00	**39.85**
	0.4	35.50	42.65	42.50	**43.50**

7. CONCLUSIONS

Illumination compensation is beneficial for video compression. In this work we have shown that exploiting illumination change need not be restricted to prediction only schemes but can be performed in the context of richer transforms, specifically those incorporating both predict and update steps. We model illumination change using an affine mesh, thereby avoiding any artificial block boundaries from appearing in the compensated frame. We have also described a highly scalable framework for coding both transformed frame texture data and the accompanying mesh-based illumination model. Experimental results highlight the effectiveness of the mesh model and convey the benefits of the update step in the temporal transform. Future work will consider incorporating transforms that exploit scene motion.

8. REFERENCES

[1] K. Kamikura, H. Watanabe, H. Jozawa, H. Kotera, and S. Ichinose, "Global brightness-variation compensation for video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 8, no. 8, pp. 988–1000, Dec 1998.

[2] Yunfei Zheng, Peng Yin, O. Divorra Escoda, Xin Li, and C. Gomila, "Intra prediction using template matching with adaptive illumination compensation," in *2008 15th IEEE International Conference on Image Processing*, Oct 2008, pp. 125–128.

[3] Peng Yin, A. M. Tourapis, and J. Boyce, "Localized weighted prediction for video coding," in *2005 IEEE International Symposium on Circuits and Systems*, May 2005, pp. 4365–4368 Vol. 5.

[4] Joaquin Lopez, Jae Hoon Kim, Antonio Ortega, and George Chen, "Block-based illumination compensation and search techniques for multiview video coding," in *Proceedings of the Picture Coding Symposium*, 2004, pp. 1–6.

[5] S. H. Tsang and Y. L. Chan, "H.264 video coding with multiple weighted prediction models," in *2010 IEEE International Conference on Image Processing*, Sept 2010, pp. 2069–2072.

[6] T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, July 2003.

[7] J. M. Boyce, "Weighted prediction in the h.264/mpeg avc video coding standard," in *2004 IEEE International Symposium on Circuits and Systems (IEEE Cat. No.04CH37512)*, May 2004, vol. 3, pp. III–789–92 Vol.3.

[8] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[9] G. H. Park, M. W. Park, S. C. Lim, W. S. Shim, and Y. L. Lee, "Deblocking filtering for illumination compensation in multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 18, no. 10, pp. 1457–1461, Oct 2008.

[10] A. Secker and D. Taubman, "Lifting-based invertible motion adaptive transform (LIMAT) framework for highly scalable video compression," *IEEE Transactions on Image Processing*, vol. 12, no. 12, pp. 1530–1542, Dec 2003.

[11] B. Pesquet-Popescu and V. Bottreau, "Three-dimensional lifting schemes for motion compensated video compression," in *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.01CH37221)*, 2001, vol. 3, pp. 1793–1796 vol.3.

[12] B. Girod and S. Han, "Optimum update for motion-compensated lifting," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 150–153, Feb 2005.

[13] "Phos image database," `http://robotics.pme.duth.gr/phos2.html`.

[14] V. Vonikakis, D. Chrysostomou, R. Kouskouridas, and A. Gasteratos, "Improving the robustness in feature detection by local contrast enhancement," in *2012 IEEE International Conference on Imaging Systems and Techniques Proceedings*, July 2012, pp. 158–163.

[15] Vasillios Vonikakis, Dimitrios Chrysostomou, Rigas Kouskouridas, and Antonios Gasteratos, "A biologically inspired scale-space for illumination invariant feature detection," *Measurement Science and Technology*, vol. 24, no. 7, pp. 074024, 2013.

[16] "AMOS dataset," `http://amos.cse.wustl.edu/dataset`.

[17] N. Jacobs, N. Roman, and R. Pless, "Consistent temporal variations in many outdoor scenes," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–6.