RATE-DISTORTION OPTIMIZED ILLUMINATION ESTIMATION FOR WAVELET-BASED VIDEO CODING

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

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

ABSTRACT

We propose a rate-distortion optimized framework for estimating illumination changes (lighting variations, fade in/out effects) in a highly scalable coding system. Illumination variations are realized using multiplicative factors in the image domain and are estimated considering the coding cost of the illumination field and input frames which are first subject to a temporal Lifting-based Illumination Adaptive Transform (LIAT). The coding cost is modelled by an ℓ_1 norm optimization problem which is derived to approximate a quadratic-log function which emerges from rate-distortion considerations. The optimization problem is solved using ADMM. The proposed solution works the same or better than a mesh-based approach proposed in prior work, where sparsity was controlled by explicitly choosing mesh parameters. In the compression-inspired formulation presented here, sparsity is discovered automatically through the solution of a convex program that depends only on a target rate-distortion operating point.

Index Terms— Wavelet-based coding, Illumination estimation, R-D optimization, Scalable video coding.

1. INTRODUCTION

Video compression techniques achieve compression efficiency by compensating for the temporal variations that occur between frames. In such schemes the current frame is predicted using a reference frame and the difference between the current frame and the prediction is coded as the residual. For sequences with illumination variations due to lighting changes or fade in/out effects, illumination compensation is essential in reducing the residual signal.

Weighted prediction methods are used in H.264 [1] and HEVC [2] standards to compensate temporal illumination changes in order to enhance the coding efficiency. Illumination compensation in the context of compression has been mostly studied using a scale and offset applied to a reference frame f_0 to predict the target frame f_1 by $\tilde{f_1} = \alpha f_0 + \beta$ [1–4]. The scale (α) and offset (β) parameters are typically spatially slow varying and estimated using block-based models providing a piecewise constant representation of the illumination variation. Although the β term forms part of the illumination model, it need not be separately coded, since it is an additive term in the image domain which can simply be folded into the coded residual information.

Block-based strategies can produce abrupt and visually-annoying changes in illumination across block boundaries [5]. Moreover,

block-based coding approaches are less amenable to highly scalable coding schemes [6].

In our previous work [4] we proposed a scalable coding framework that utilizes a lifting-based illumination adaptive transform (LIAT) to exploit inter-frame redundancy in the presence of illumination variations. A spatially affine mesh-based model replaced the block-based models, enabling a spatially smooth description of illumination change. Importantly, we showed that the inclusion of an update step in addition to the predict step in the temporal transform schemes for illumination compensation can markedly improve the coding efficiency of conventional predict only schemes.

In our prior work [4] the illumination field was defined using a fixed size mesh, limiting the illumination representation to the same sparsity over the whole frame. Furthermore, the optimal mesh size had to be discovered by searching through a range of plausible sizes and noting the corresponding rate-distortion (R-D) performance. In this work we address these limitations while retaining the highly scalable nature of the coding framework.

We also note that alternative coding schemes [7–9] that use blocks to perform spatially varying illumination compensation, estimate the block-based illumination parameters in a manner that is greedy from a R-D optimization perspective. That is illumination parameters are determined by taking into consideration the R-D impact on the current block and its causal neighbours. The impact on future blocks that are yet to be coded is not considered and therefore only a sub-optimal solution can be reached. In contrast, in this work we estimate the illumination model parameters by solving a convex program that finds a globally optimum R-D operating point.

In a compression framework, both the illumination field and the texture information are subject to a spatial transform prior to coding; therefore we formulate a model for coding cost in the transform domain. Modelling the problem in the transform domain also resolves the need to explicitly deal with the offset value β for illumination compensation since the offset term becomes part of the coding cost of the residue component. While we provide R-D coding results, our emphasis is to illustrate that the proposed compression inspired convex formulation can effectively distribute information in a sequence between coded multiplicative illumination terms and texture data.

Since our proposed framework uses temporal and spatial wavelet transform, it naturally endows spatio-temporal scalability. In addition, the wavelet subband coefficients are subject to embedded block coding as defined by the EBCOT algorithm of JPEG 2000 [10], thereby ensuring rate or quality scalability.

It should be noted that this work primarily focuses on applications with considerable illumination variation among frames. For our study here, global motion is compensated as a preprocessing step to illumination compensation. Apart from this, we do not consider motion at all in this work, so that any local scene motion remains un-



Fig. 1. The lifting-based illumination adaptive temporal transform

compensated in our experiments. This is not a necessary restriction of our approach, but allows us to focus exclusively on the problem of illumination compensation. The proposed approach is suitable as-is for applications in low frame rate surveillance videos.

2. LIFTING-BASED ILLUMINATION ADAPTIVE TRANSFORM

The Lifting-based Illumination Adaptive Transform (LIAT) framework proposed in our previous work [4] applies a t+2D approach where an illumination adaptive temporal subband transform is first applied to the input frames, followed by a 2D spatial wavelet transform on the resulting temporal subbands and the illumination field. Consider a pair of frames f_0 and f_1 , and an estimated illumination variation parameter α ; the predict and update steps in the LIAT framework [4] can be written as

$$h(\alpha) = f_1 - \alpha f_0, \quad l(\alpha) = f_0 + b h(\alpha) \tag{1}$$

Here, the update factor b is chosen such that the temporal synthesis gain is minimized [11] and as explained in [4] the distortion minimizing choice for this case is $b = \frac{\alpha}{1+\alpha^2}$. The LIAT framework for two levels of temporal decomposition is illustrated in Fig. 1. Illumination fields $(\alpha_{0\to 1}, \alpha_{0\to 2}, \text{ and } \alpha_{2\to 3})$ and temporal subband frames $(l_0^{(2)}, h_0^{(1)}, h_0^{(2)}, h_1^{(1)})$ undergo spatial wavelet decomposition and embedded block coding to generate highly scalable codestreams for both texture and illumination information.

3. R-D OPTIMIZED ILLUMINATION ESTIMATION IN LIAT FRAMEWORK

According to (1) we have three components to code after applying the LIAT; the illumination field (α), the temporal high pass (h) and low pass (l) subband frames. For convenience we refer to them as LIAT subband frames and denote all of them by s. We also use \mathcal{A} to indicate the spatial analysis linear operator (i.e. wavelet transform) which is used to transform subband frame s into a collection of spatial subbands for coding such that $y_s = \mathcal{A}(s)$. The coding of these spatial subbands contributes to the total distortion (D) of reconstructed frames in the image domain and overall coded length (L). The R-D optimality condition in the encoder is achieved when the total Lagrangian cost functional $J = D + \lambda L$ is minimized. It is well known that for $\lambda > 0$ the cost functional J in the encoder is minimized when the distortion length slope $\partial D/\partial L$ for each subband frame is equal to $-\lambda$. Let $y_{s,n}$ be the transformed coefficient of the LIAT subband s at position n. Using the high rate coding model, we formulate the total distortion and total coded length as [12, 13]

$$D = \sum_{s,n} D_{s,n} = \sum_{s,n} |y_{s,n}|^2 g_{s,n} e^{-aL_{s,n}}$$
(2)

$$L = \sum_{s,n} L_{s,n} = \sum_{s,n} \left(L_{s,n} + L_{s,n}^{\sigma} \right) \tag{3}$$

where $a = 2 \ln 2$, $g_{s,n}$ is the corresponding spatio-temporal synthesis gain, and $L_{s,n}^{\sigma}$ and $L_{s,n}$ refer to the coding length required to signal the significance and magnitude bits respectively. The parameter $g_{s,n}$ is essentially the product of the spatial and temporal synthesis gains.

In the EBCOT coding strategy a significance state is first coded which indicates the presence of a zero or non-zero coefficient at a particular location. This is then followed by a refinement stage which communicates the sign and magnitude bits of the coefficient if it has been signaled as being significant (i.e. non-zero) [10]. We use $L_{s,n}^{\sigma}$ to indicate the required bits to communicate the significance state and $L_{s,n}$ to represent the bits required for coding the magnitude and sign of the transformed subband coefficient at position *n*. The EBCOT algorithm employs arithmetic coding to signal the significance state; therefore for our analytical work we model the corresponding number of bits required as $L_{s,n}^{\sigma} = -\log_2 p_{s,n}$, where $p_{s,n}$ refers to the probability of the subband coefficient being non-zero and is estimated a priori based on the observation of similar subband coded frames [13].

Given the distortion model in (2) and under R-D optimality conditions $\partial D/\partial L = -\lambda$, it is not beneficial to expend any bits if $|y_{s,n}|^2 \leq \lambda/(ag_{s,n})$. This means that distortion $D_{s,n} = g_{s,n}|y_{s,n}|^2$ and the rate $L_{s,n}$ is equal to the number of bits required to signal a non-significant state which in reality should be very small and for our analytical work is assumed to be zero $(L_{s,n} \approx 0)$. For values of $|y_{s,n}|^2$ exceeding the threshold $\lambda/(ag_{s,n})$ it is necessary to expend bits to communicate the coefficient magnitude; the optimal number of bits required and the resulting distortion after coding can be readily derived and is further explained in [13]. The corresponding Lagrangian cost functional $J(y_{s,n})$ at R-D optimal operation is given in equation (4).

$$J_{y_{s,n}} = \begin{cases} g_{s,n} |y_{s,n}|^2 & \text{if } |y_{s,n}|^2 \leq \frac{\lambda}{ag_{s,n}} \\ \frac{\lambda}{a} + \frac{\lambda}{a} \ln\left(\frac{g_{s,n} |y_{s,n}|^2}{\lambda/a}\right) + \lambda L_{s,n}^{\sigma} & \text{otherwise} \end{cases}$$
(4)

We refer to $J_{y_{s,n}}$ as a "quadratic-log" function of $y_{s,n}$ which is depicted in Fig. 2. Therefore, total coding cost of the LIAT subband frame *s* can be written as

$$J_s = \sum_n J_{y_{s,n}} \tag{5}$$

Our aim is to estimate α such that the total Lagrangian cost functional

$$J = J_{\alpha} + J_{h(\alpha)} + J_{l(\alpha)} \tag{6}$$

is minimized. The quadratic-log function in (4) is neither convex or sub-quadratic. Therefore in this paper, as an initial step to solve (6), we use an ℓ_1 -norm upper bound function as illustrated in Fig. 2. This ℓ_1 -norm upper bound is a surrogate function for the quadratic-log function and is used as a precursor to better approximations of the problem in (6) that we intend to pursue in the future. The gradient of the ℓ_1 function for practical values of $L_{s,n}^{\sigma}$ can be obtained from (4) as

$$m_{s,n} = \frac{\lambda/a + \lambda L_{s,n}^{s}}{\sqrt{\frac{\lambda}{ag_{s,n}}}}$$
(7)



Fig. 2. Quadratic-log cost function and the corresponding ℓ_1 norm as its upper bound.

Using the ℓ_1 -norm function $||m_s y_s||_1 = \sum_n |m_{s,n} y_{s,n}|$ as an upper bound for J_s , the cost objective functional in (6) can be approximated by ℓ_1 -norm functions as

$$\underset{\alpha}{\operatorname{argmin}} C(\alpha) = \|m_{\alpha}y_{\alpha}\|_{1} + \|m_{h}y_{h(\alpha)}\|_{1} + \|m_{l}y_{l(\alpha)}\|_{1}$$
(8)

4. OPTIMIZATION USING ADMM

In this section, we introduce new notation and make certain approximations so as to rewrite the cost objective of equation (8) in a form that can be readily solved by the well known convex optimization method ADMM (alternating direction of multipliers) [14]. Let $z = y_{\alpha} = \mathcal{A}(\alpha)$ and $\alpha = \mathcal{S}(z)$, where \mathcal{S} is the spatial synthesis operator (inverse operator for \mathcal{A}). As the cost objective function in (8) is a function of α , it can be also written as a function of z using (1).

$$\underset{z}{\operatorname{argmin}} C(z) = \|m_{\alpha} z\|_{1} + \|m_{h} \mathcal{A}(f_{1}) - m_{h} \mathcal{A}(f_{0} \mathcal{S}(z))\|_{1}$$

$$+ \|m_{l} \mathcal{A}(f_{0} + bf_{1}) - m_{l} \mathcal{A}(bf_{0} \mathcal{S}(z))\|_{1}$$
(9)

If b' is approximated such that $\mathcal{A}(bf_0\mathcal{S}(z)) \simeq b'\mathcal{A}(f_0\mathcal{S}(z))$, the optimization problem in (9) can be rewritten as:

$$\underset{z}{\operatorname{argmin}} C(z) = \|m_{\alpha} z\|_{1} + \|m_{h} \mathcal{A}(f_{1}) - m_{h} \mathcal{D} z\|_{1}$$

$$+ \|m_{l} \mathcal{A}(f_{0} + bf_{1}) - m_{l} b' \mathcal{D} z\|_{1}$$
(10)

where \mathcal{D} is a linear operator such that $\mathcal{D}z = \mathcal{A}(f_0\mathcal{S}(z))$. We will later discuss how b' is approximated. The optimization problem of equation (10) can now be expressed in a form that can be readily solved by ADMM as shown below.

minimize
$$f(z) + g(\tilde{z})$$

subject to $\tilde{z} - Dz = 0$ (11)

Here, $f(z) = ||m_{\alpha}z||_1$ and $g(\tilde{z}) = ||m_h\mathcal{A}(f_1) - m_h\tilde{z}||_1 + ||m_l\mathcal{A}(f_0 + bf_1) - m_lb'\tilde{z}||_1$. Note that f(z) refers to the coding cost of the α field while $g(\tilde{z})$ signifies the coding cost of the temporal high pass and low pass LIAT subband frames. The variables z and \tilde{z} enable each of these cost terms to be expressed in a most natural and straightforward manner while the constraint of equation (11) defines the relationship or dependency between the two variables. The augmented Lagrangian of (11) can be written in the scaled format of ADMM as (see [14] section 3.1.1)

$$L\rho = f(z) + g(\tilde{z}) + \frac{\rho}{2} \|\mathcal{D}z - \tilde{z} + u\|_{2}^{2} - \frac{\rho}{2} \|u\|_{2}^{2}$$
(12)

where u is the scaled dual variable for the constraint $\tilde{z} - Dz = 0$. Using the linearized ADMM algorithm (see [15] sec.4.4.2), the solution can be iteratively found by the following update steps [15].

$$z^{k+1} = \underset{z}{\operatorname{argmin}} \left(f(z) + \frac{1}{2\mu} \| z - v^k \|_2^2 \right)$$
(13)
$$\tilde{z}^{k+1} = \underset{\tilde{z}}{\operatorname{argmin}} \left(g(\tilde{z}) + \frac{1}{2\gamma} \| \tilde{z} - (u^k + \mathcal{D}z^{k+1}) \|_2^2 \right)$$
$$u^{k+1} = u^k + \mathcal{D}z^{k+1} - \tilde{z}^{k+1}$$



Fig. 3. Sample images from indoor and outdoor image sequences.

where $v^k = z^k - \frac{\mu}{\gamma} \mathcal{D}^T (\mathcal{D} z^k - \tilde{z}^k + u^k)$ and $0 < \mu \le \gamma / ||\mathcal{D}||_2^2$. The proof of ADMM convergence can be found in [14]. The reports of convergence using the linearized algorithm are also discussed in [15] and references therein.

To approximate b' in (10), we assume that the update factors can be well modelled by a smoothly varying 2D field. From our observations, this remains predominantly true for the vast majority of image regions. To determine b', first, we find α_l as a smooth version of α field by removing its high frequency components in the wavelet domain and b_l is found as $b_l = \frac{\alpha_l}{1+\alpha_l^2}$. Then, the low pass filter coefficients for the 5/3 wavelet are used to filter the b_l . We take the absolute values of the coefficients [-1/8, 2/8, 6/8, 2/8, -1/8] and normalize them to obtain a smooth "local averaging" filter as in (14).

$$\mathcal{F} = [1/12, 2/12, 6/12, 2/12, 1/12] \tag{14}$$

Finally, b' is estimated by filtering b_l with the local averaging filter as in (15).

$$b' = \mathcal{F} * b_l \tag{15}$$

5. DISCUSSION AND EXPERIMENTAL RESULTS

Fig. 3 shows some samples of image sequences which are used in HD resolution to test the proposed algorithm. The indoor sequences correspond to scenes with directional, non-uniform, illumination variations from [16–18]. The outdoor sequences are from low frame rate surveillance cameras from the AMOS dataset [19] and the "building sequence" is one which we captured at a frame rate of one frame every 15 minutes. We compare the performance of the proposed R-D optimized illumination estimation scheme with the mesh-based illumination model from our previous work. We also compare with conventional highly scalable coding schemes, these relate to (i) replacing the temporal transform with the Haar and (ii) independent coding of each frame with JPEG 2000.

R-D results and estimated α fields for a pair of images (second and third frames, second row of Fig. 3 for both indoor and outdoor sequences) which are from Scene 11 and the Oslo-Linpro AS dataset are illustrated in Fig. 4 and 5 respectively. Results of the proposed R-D optimized method are labelled "LIAT-RDO" while "LIAT-Mesh" indicates the R-D results of illumination estimation by employing the affine mesh model. Curves labelled "J2K" relate to independent coding of each frame with a JPEG 2000 encoder that employs the 5/3 wavelet transform. Results labelled "Haar" relate to temporal Haar transform being applied on image sequences prior to the spatial transform. 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.

As seen in Fig. 4(a) the proposed method outperforms other approaches. For illumination estimation that relies upon the affine mesh model, we need to consider a range of different mesh sizes to identify a mesh model that produces the best R-D performance. For Fig. 4 using mesh size N = 16 shows the best result among other mesh sizes. We have shown results for N = 64 for comparison.



Fig. 4. R-D results for the second and third frame of Scene 11 sequence in Fig.3 using illumination estimation with (b) proposed R-D optimized method (c) affine mesh size 16 and (d) mesh size 64.



Fig. 5. R-D results for the second and third frame of Oslo-Linpro AS sequence in Fig.3 using illumination estimation with (b) proposed R-D optimized method (c) affine mesh size 16 and (d) mesh size 64.

Figure 4(c) and (d) illustrate that the sparsity of the illumination field estimated by a predefined mesh size does not change spatially over the frame. However, the illumination field estimated by the proposed R-D optimized method in Fig. 4(b) is sharp in some areas while being smooth in other parts. This can be understood as the proposed framework allowing the α field to adapt its spatial sparsity based on a target R-D operating point. The procedure considers placing high frequency information either in texture or α field taking heed of their final coding cost.

The differences between the mesh-based and the proposed illumination estimation are most evident at boundaries of objects where there is an abrupt change in illumination, for example the top edge of the book in the frame of Fig. 4. Such sudden changes in illumination do not conform to any predefined smooth models and hence our prior scheme is inferior in handling these regions. Similar observations can be made for the case shown in Fig. 5. We note that this outdoor data set corresponds to much larger variations in illumination. In such cases that there is not enough temporal correlation between frames, it is preferred to send them separately without applying any temporal transform. As seen in 5(a), "J2K" is working better than "Haar". However, the superior performance of our proposed scheme illustrates its ability to efficiently adapt to sharp tran-



Fig. 6. Indoor sequences R-D results for two temporal levels of LIAT framework.



Fig. 7. Outdoor sequences R-D results for two temporal levels of LIAT framework.

sitions in illumination in both the temporal and spatial dimensions. R-D results of the image sequences for two temporal levels of the LIAT framework are shown in Fig. 6 and Fig. 7. The mesh size N for mesh-based illumination estimation is chosen such that the best R-D result is achieved for the "LIAT-Mesh" curves. Although we are using an approximation of the cost function (i.e. ℓ_1 norm upper bound), we are able to achieve the same or better R-D performance than the mesh based scheme which requires a parameter search to identify the best model. The experimental results therefore validate our proposed optimization strategies for the non-convex quadratic-log cost function.

6. CONCLUSIONS

In this paper we propose theoretical foundations for a R-D cost functional to estimate multiplicative illumination terms in the context of a lifting-based illumination adaptive coding framework. We note that this paper is not just about coding; we show that a compression inspired convex formulation can be used to effectively distribute the information in a sequence between multiplicative illumination terms and non-multiplicative residual terms. In a fair comparison we find that the optimized result consistently outperforms parametric meshbased models, in which the mesh parameters are tuned by trial and error, which is a very promising result.

7. REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] M. Haghighat, R. Mathew, A. Naman, and D. Taubman, "Lifting-based illumination adaptive transform (LIAT) framework using mesh-based illumination modelling," in 2014 IEEE International Conference on Image Processing (ICIP), 2017.
- [5] 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.
- [6] 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.
- [7] J. H. Kim, P. Lai, J. Lopez, A. Ortega, Y. Su, P. Yin, and C. Gomila, "New coding tools for illumination and focus mismatch compensation in multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1519–1535, Nov 2007.
- [8] J. H. Hur, S. Cho, and Y. L. Lee, "Adaptive local illumination change compensation method for H.264/AVC-based multiview video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, no. 11, pp. 1496–1505, Nov 2007.
- [9] 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.
- [10] D. Taubman, "High performance scalable image compression with ebcot," *IEEE Transactions on Image Processing*, vol. 9, no. 7, pp. 1158–1170, Jul 2000.
- [11] B. Girod and S. Han, "Optimum update for motioncompensated lifting," *IEEE Signal Processing Letters*, vol. 12, no. 2, pp. 150–153, Feb 2005.
- [12] David S. Taubman and Michael W. Marcellin, JPEG 2000: Image Compression Fundamentals, Standards and Practice, Kluwer Academic Publishers, Norwell, MA, USA, 2001.
- [13] R. Mathew, D. Taubman, and P. Zanuttigh, "Scalable coding of depth maps with r-d optimized embedding," *IEEE Transactions on Image Processing*, vol. 22, no. 5, pp. 1982–1995, May 2013.
- [14] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends* (R) in Machine Learning, vol. 3, no. 1, pp. 1–122, 2011.

- [15] Neal Parikh, Stephen Boyd, et al., "Proximal algorithms," *Foundations and Trends* (*in Optimization*, vol. 1, no. 3, pp. 127–239, 2014.
- [16] "Phos image database," http://robotics.pme.duth. gr/phos2.html.
- [17] 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.
- [18] 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.
- [19] "AMOS dataset," http://amos.cse.wustl.edu/ dataset.