Statistical Modeling of Multi-camera Images

Rajiv Soundararajan, Alan C. Bovik and Sriram Vishwanath

Department of Electrical and Computer Engineering, The University of Texas at Austin

1 University Station C0803, Austin, TX 78712, USA

Email: rajiv.s@mail.utexas.edu, bovik@ece.utexas.edu and sriram@ece.utexas.edu

Abstract—**This paper aims at developing mutual statistical descriptions of images in a visual sensor network. A two-dimensional Gaussian scale mixture (GSM) model is proposed to describe the joint distribution of wavelet coefficients in the same subband of two different images. This model is tested on two image databases, and it is shown that the proposed model always achieves a smaller divergence from the empirical distribution than a (jointly) Gaussian model. The GSM model is then extended to blocks of wavelet coefficients in the same subband and scale of non-overlapping images. A new denoising algorithm is devised that uses this model and achieves better performance than denoising based on marginal models.**

I. INTRODUCTION

The study of visual sensor networks (VSN) has gained significant importance recently, with increased deployment in applications such as video surveillance, traffic control etc. A *visual sensor network* is a network of isolated cameras that are imaging different scenes or possibly different views of the same scene. To better understand the fundamental limits of compression and communication in such systems, it is essential that a good model of the data sources (images) be developed. We adopt a statistical approach to this problem and propose Gaussian scale mixtures to succinctly and accurately model a VSN.

There has been significant prior work in describing the statistics of wavelet coefficients for a single image [1] [2]. A Gaussian scale mixture (GSM) distribution has been shown to be a good model for describing the statistics of wavelet coefficients for single images [3]. The problem of the joint distribution of a pair of wavelet coefficients in a single image is addressed in [4] and [5]. On the other hand, the problem of distributed video coding with multi-view cameras is considered in [6] and [7].

We consider a pair of potentially non-overlapping images in a VSN and describe the joint distribution of the wavelet coefficients of these images. In particular, we perform a multi-scale wavelet decomposition of the two images and study the joint distribution of a pair of wavelet coefficients belonging to the same subband at the same scale. We require that the model for the joint distribution be such that the marginal distribution is a Gaussian Scale Mixture (GSM) distribution [5]. Therefore, our model for the joint statistics of two images is a two-dimensional GSM. Note that both the one-dimensional GSM for a single image and the two-dimensional GSM for a pair of images are good models for the wavelet coefficients in all the sub-bands other than the subband obtained by successive low pass filtering. Thus, throughout the paper, a subband

refers to any subband except for the subband obtained by successive low pass filtering at different scales. We show that the proposed model has a smaller Kullback-Leibler (KL) divergence with the empirical joint distribution than a jointly Gaussian model for the wavelet coefficients.

We also extend this model to blocks of wavelet coefficients in the same subband at the same scale. We demonstrate the utility of this model in VSNs through a denoising application. We show the potential benefits that can be attained by using two correlated noisy images while denoising one of the images. In effect, we show that we can improve the quality of a denoised image by using both noisy images as against simply using the given noisy image. The denoising algorithm proposed in this paper is particularly useful when the images are highly correlated although non-overlapping.

The rest of the paper is organized as follows. We recall the marginal distribution of wavelet coefficients in Section II and propose models for the joint distribution of wavelet coefficients in Section III. We discuss the denoising algorithm in Section IV and present results demonstrating the performance of these models in Section V. Finally, we conclude the paper in Section VI.

II. MARGINAL DISTRIBUTION OF WAVELET COEFFICIENTS

We now recall the GSM model for the marginal distribution of the wavelet decomposition coefficients of a single image introduced in [5]. The inherent assumption with this model is that the coefficients are independent and identically distributed within each subband. Let S be a wavelet coefficient in a given subband and scale. The random variable S is distributed as

$$S = \sqrt{e^X} U,$$

where X and U are independent and Gaussian distributed. U is assumed to be a normal random variable with mean zero and unit variance. Thus the premultiplier e^X corresponds to the variance of the wavelet coefficient while U corresponds to the detail information. Note that the model proposed here assumes that the non-linearity in the GSM model proposed in [5] is an exponential function. We make this assumption as it simplifies the foregoing formulation without negatively impacting our ability to match the empirical distribution. The parameters of the model can be estimated either through the method of moments as in [8] or by maximum a posteriori estimation [5]. The GSM model is shown to be better than a Gaussian model for the marginal distribution of wavelet

coefficients in an image through a denoising application in both these papers.

III. JOINT DISTRIBUTION OF WAVELET COEFFICIENTS OF TWO IMAGES

In this section, we present two models for the joint distribution of wavelet coefficients in two non-overlapping images. The first model considers a pair of wavelet coefficients from the two images, while the second model considers a block of wavelet coefficients from each image.

A. Wavelet coefficients in the same subband - Model 1

We now consider the joint distribution of a pair of wavelet coefficients (one from each image) in the same subband and scale. As before, the coefficients are assumed to be i.i.d within each subband of both images. Let S_1 and S_2 be wavelet coefficients (scalars) belonging to the same subband at the same scale of two different images. The model proposed below is specific to corresponding subbands of the two images. Therefore, the statistical parameters may be different for a different pair of subbands. The motivation behind the proposed model is that variance fields between images are correlated while the details are independent. This idea is drawn from the joint distribution of wavelet coefficients across scales [5]. Moreover, empirical observations also reveal that the wavelet coefficients of two different images are decorrelated while the variances of the wavelet coefficients are correlated. Thus, S_1 and S_2 are distributed as

$$S_1 = \sqrt{e^{X_1}} U_1 \qquad S_2 = \sqrt{e^{X_2}} U_2,$$

where X_i and U_i are independent of each other and each is Gaussian distributed for $i = 1, 2$. Further, the random variables U_1 and U_2 have zero mean and unit variance and are independent of each other. The joint distribution between S_1 and S_2 is on account of the associated joint between X_1 and X_2. X_1 and X_2 are jointly Gaussian distributed with mean μ_1 and μ_2 and covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

Let $\{s_{1k}\}_{k=1}^n$ and $\{s_{2k}\}_{k=1}^n$ be the observed wavelet coefficients of the two images in the same subband at the same scale, where n is the total number of wavelet coefficients in the subband. We depart from the procedure of estimating the parameters using the method of moments [8] since this method could lead to $\rho \notin [-1, 1]$. Also, we desist from obtaining the maximum a posteriori estimate of the parameters as detailed in [5] for simplicity. Instead, we adopt a simple suboptimal procedure to estimate μ_1, μ_2, σ_1^2, σ_2^2 and ρ as follows. We construct the maximum likelihood estimate of x_{1k} and x_{2k} given by $\hat{x}_{1k} = \log s_{1k}^2$ and $\hat{x}_{2k} = \log s_{2k}^2$ respectively. These estimates are treated as realizations of X_1 and X_2 to construct

maximum likelihood estimates of the parameters. Thus,

$$\mu_1 = \frac{1}{n} \sum_{k=1}^n \log(s_{1k}^2) \qquad \sigma_1^2 = \frac{1}{n} \sum_{k=1}^n (\log(s_{1k}^2) - \mu_1)^2$$

$$\mu_2 = \frac{1}{n} \sum_{k=1}^n \log(s_{2k}^2) \qquad \sigma_2^2 = \frac{1}{n} \sum_{k=1}^n (\log(s_{2k}^2) - \mu_2)^2$$

$$\rho = \frac{1}{n\sigma_1\sigma_2} \sum_{k=1}^n (\log(s_{1k}^2) - \mu_1)(\log(s_{2k}^2) - \mu_2).$$

This model achieves a smaller KL divergence with the empirical distribution than a jointly Gaussian model. The results of this experiment are presented in Section V. However, it turns out that the correlation coefficient ρ, is not large enough for natural scenes. This limits the utility of the model for applications that look to exploit the correlation between the images. We thus consider an extended model in the following section.

B. Blocks of wavelet coefficients in the same subband - Model 2

We now propose a statistical model for blocks of wavelet coefficients in the two images. The wavelet coefficients in each image in every subband and scale are partitioned into non-overlapping blocks, each consisting of a total of N coefficients. We assume that non-overlapping blocks of wavelet coefficients within an image are independent. This assumption is made since the objective of this paper is in studying the joint statistics between two images. Let $\overline{S}_i = (S_{i1}, S_{i2}, \ldots, S_{iN})$, $i \in \{1, 2\}$, denote a block of wavelet coefficients (in the same subband and scale) in the two images. We propose a GSM vector model as

$$\overline{S}_1 = \sqrt{e^{X_1}} \overline{U}_1 \qquad \overline{S}_2 = \sqrt{e^{X_2}} \overline{U}_2,$$

where X_i and $\overline{U}_i$ are independent of each other for $i = 1, 2$. Further, $\overline{U}_1$ and $\overline{U}_2$ are Gaussian random vectors with zero mean and identity covariance matrix and are independent of each other. X_1 and X_2 are Gaussian distributed with mean μ_1 and μ_2 with covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}.$$

We now describe the estimation of the parameters of the model. Let M be the total number of blocks in each image. Note that M denotes the total number of blocks across subbands at different scales. Observed wavelet coefficients of Image i, $i \in \{1, 2\}$ in block j, $j \in \{1, \ldots, M\}$ are denoted by $\{s_{ijk}\}_{k=1}^N$. The estimates of the parameters are obtained in the same fashion as before and shown below. Let

$$\eta_{ij} = \frac{1}{N} \sum_{k=1}^N s_{ijk}^2.$$

Then,

$$\mu_i = \frac{1}{M}\sum_{j=1}^{M}\log\eta_{ij}$$

$$\sigma_i^2 = \frac{1}{M}\sum_{j=1}^{M}(\log\eta_{ij}-\mu_i)^2$$

$$\rho = \frac{1}{M\sigma_1\sigma_2}\sum_{j=1}^{M}(\log\eta_{ij}-\mu_1)(\log\eta_{ij}-\mu_2)$$

for $i \in \{1,2\}$.

In this model, the correlation between the two images is modeled through the variance of a block of coefficients rather than the variance of individual coefficients. Also, we impose that a block of wavelet coefficients has the same premultiplier. We find that this yields higher correlations between the premultipliers for certain natural scenes, which allows us to exploit the model in a denoising application described in the next section.

IV. Denoising

We demonstrate the utility of the model for the block of wavelet coefficients through a denoising application. Let Y_{ij}^N denote the wavelet coefficients in block j of the noisy Image i, where $i \in \{1,2\}$ and j, $j \in \{1,\ldots,M\}$. The received signal is given by

$$Y_{ij}^N = S_{ij}^N + Z_{ij}^N,$$

where Z_{ij}^N is Gaussian distributed with mean zero and co-variance matrix $\sigma_{Z_i}^2\mathbf{I}_{N\times N}$. We assume that the noise and the source statistics i.e. $\sigma_{Z_1}^2$, $\sigma_{Z_2}^2$, μ_1, μ_2, σ_1^2, σ_2^2 and ρ are known exactly for every subband and every scale at the receiver while constructing the estimate.

The estimation procedure is similar to Bayesian least squares estimate in [3], the difference being that we assume knowledge of the source and channel statistics at the receiver. Also, the proposed algorithm uses both images while denoising one of the images. The minimum mean squared error (MMSE) estimate of the denoised wavelet coefficient in a given subband at a scale is given by

$$\hat{s}_{ijk} = \int \frac{e^{x_i}}{e^{x_i}+\sigma_{Z_i}^2}y_{ijk}p(x_i|y_{1j}^N,y_{2j}^N)dx_i,$$

where $k \in \{1,2,\ldots,N\}$. Here,

$$p(x_1|y_{1j}^N,y_{2j}^N) = \int p(x_1,x_2|y_{1j}^N,y_{2j}^N)dx_2$$

$$= \int \frac{p(y_{1j}^N,y_{2j}^N|x_1,x_2)p(x_1,x_2)}{p(y_{1j}^N,y_{2j}^N)}dx_2$$

$$= \frac{p(y_{1j}^N|x_1)}{p(y_{1j}^N,y_{2j}^N)}\int p(y_{2j}^N|x_2)p(x_1,x_2)dx_2,$$

where

$$p(y_{1j}^N,y_{2j}^N) = \int\int p(y_{1j}^N,y_{2j}^N|x_1,x_2)p(x_1,x_2)dx_1dx_2$$

$$= \int\int p(y_{1j}^N|x_1)p(y_{2j}^N|x_2)p(x_1,x_2)dx_1dx_2.$$

Similarly,

$$p(x_2|y_{1j}^N,y_{2j}^N) = \int p(x_1,x_2|y_{1j}^N,y_{2j}^N)dx_1$$

$$= \frac{p(y_{2j}^N|x_2)}{p(y_{1j}^N,y_{2j}^N)}\int p(y_{1j}^N|x_1)p(x_1,x_2)dx_1.$$

We will refer to the algorithm presented above as the VSN algorithm.

We will compare this algorithm with a denoising algorithm that just uses the given noisy image for denoising. This algorithm will be referred to as the point to point (PTP) algorithm and is similar to the algorithm in [3] with differences in the estimation procedure. Even in this PTP algorithm, we assume that the source and channel statistics are known at the receiver. In contrast to the VSN algorithm, the PTP algorithm only requires knowledge of the marginal statistics of the source. The MMSE estimate of the wavelet coefficient according to the PTP algorithm is given by

$$\hat{s}_{ijk} = \int \frac{e^{x_i}}{e^{x_i}+\sigma_{Z_i}^2}y_{ijk}p(x_i|y_{ij}^N)dx_i$$

$$= \int \frac{e^{x_i}}{e^{x_i}+\sigma_{Z_i}^2}y_{ijk}\frac{p(y_{ij}^N|x_i)p(x_i)}{p(y_{ij}^N)}dx_i,$$

where

$$p(y_{ij}^N) = \int p(y_{ij}^N|x_i)p(x_i)dx_i.$$

The results of comparison are presented in the next section.

V. Results

We present numerical results for the superior performance of both the models proposed in this paper. We begin by considering the model for the joint distribution of wavelet coefficients in two images (Model 1). We test this model on two different image databases, the van Hateren database [9] and the Kodak database [10]. A wavelet decomposition on the images is performed using the BiorSplines wavelet at a single level. We observe that the KL divergence between the GSM model and empirical distribution is much smaller than the relative entropy between the Gaussian model and the empirical distribution. The Gaussian model is chosen to have the same first and second order statistics as the empirical distribution. The KL divergences for the wavelet coefficients in one of the subbands are shown in Fig. 1 and Fig. 2 respectively. The X-axis contains the different image pairs for which the joint distribution is computed. The KL divergences obtained for different image pairs are sorted in increasing order and shown in the plots.
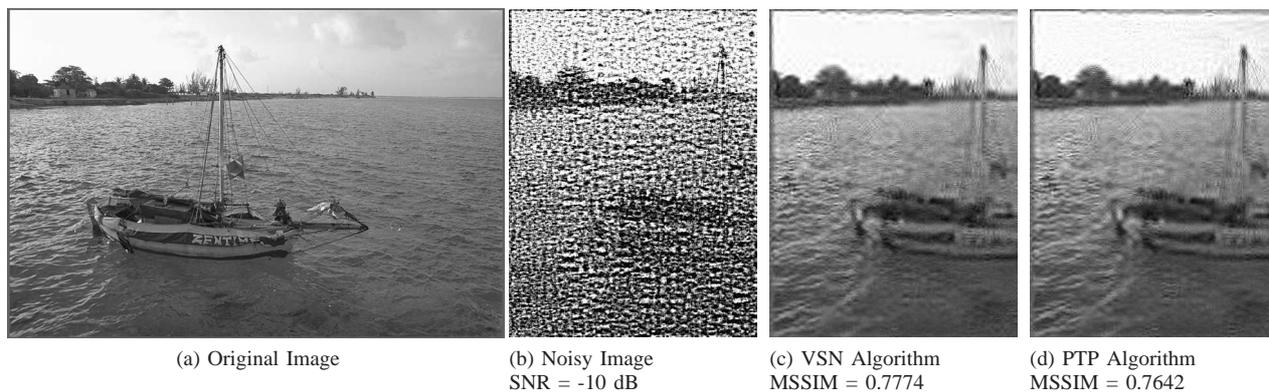
(a) Original Image | (b) Noisy Image SNR = -10 dB | (c) VSN Algorithm MSSIM = 0.7774 | (d) PTP Algorithm MSSIM = 0.7642

Fig. 3: Results of denoising sailing1.bmp

(a) Original Image | (b) Noisy Image SNR = -10 dB | (c) VSN Algorithm MSSIM = 0.8886 | (d) PTP Algorithm MSSIM = 0.8760
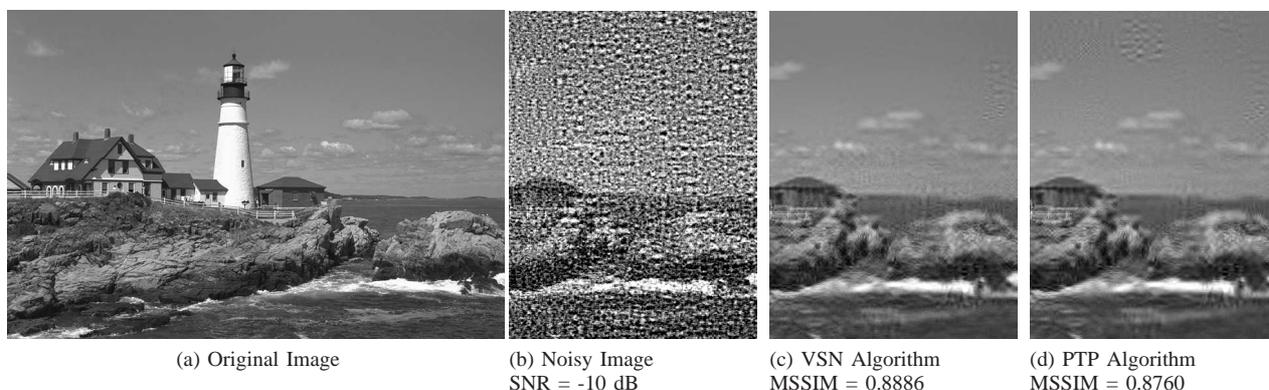
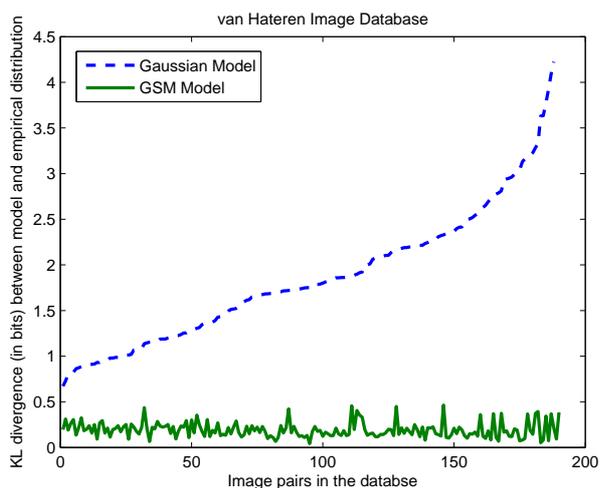Fig. 4: Results of denoising lighthouse2.bmp

Fig. 1: Superior performance of GSM model over Gaussian model on van Hateren database
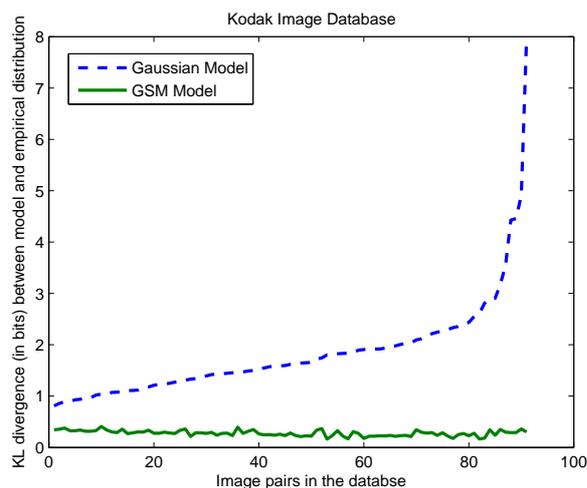
Fig. 2: Superior performance of GSM model over Gaussian model on Kodak database

We now present results of the denoising experiments for a couple of images in the Kodak database [10]. Each image is split into two halves and the wavelet coefficients in both halves are corrupted by additive white Gaussian noise. The wavelet coefficients are obtained by computing a wavelet transform using the BiorSplines wavelet at three scale levels. The two halves are received at different signal to noise ratios (SNR). In Fig. 3, the second half of 'sailing1.bmp' is received at 10 dB

while the first half is received at -10 dB. Note that the noise variances in each subband are adjusted to obtain a uniform SNR for demonstration purposes. It is clear that the VSN algorithm achieves a better multiscale structural similarity index (MSSIM) [11] value than the PTP algorithm. Recall that the VSN algorithm is based on Model 2. In Fig. 4, the first half of 'lighthouse2.bmp' is received at 10dB while the second half is received at -10 dB. We wish to remark that both 'sailing1.bmp' and 'lighthouse2.bmp' are images which have a high value of the correlation coefficient ρ in different subbands. The superior performance of the VSN algorithm is under conditions of high correlation between the two halves. Moreover, the VSN algorithm performs better when the other half is obtained at a much better SNR than the given half. This is demonstrated through numerical results in Table I where the MSSIM values of the first half of the denoised 'sailing1.bmp' are tabulated.

TABLE I: Performance variation of MSSIM with SNR for VSN and PTP algorithms (first half of 'sailing1.bmp' image)

SNR1 = SNR of first half SNR2 = SNR of second half	VSN MSSIM	PTP MSSIM
SNR1 = 1 SNR2 = 10	0.8550	0.8521
SNR1 = 0.1 SNR2 = 10	0.7774	0.7642
SNR1 = 0.01 SNR2 = 10	0.7527	0.7149

VI. CONCLUSION

We present models for the joint distribution of wavelet coefficients in the same subband of two different images. We show that a GSM model (Model 1) is better than a Gaussian model in terms of achieving a smaller KL divergence with the empirical distribution. We devise a denoising algorithm based on another GSM model (Model 2) that exploits the inherent correlation between the images. The algorithm is particularly useful, when the images are highly correlated in the VSN. Moreover, the algorithm also provides an approach to improve the quality of a denoised image given a less noisy correlated image as well. The model could also be potentially useful in distributed compression and other communication problems in visual sensor networks.

REFERENCES

[1] E. P. Simoncelli, "Statistical modeling of photographic images," in *Handbook of Image and Video Processing*, 2005, pp. 961–974.

[2] J. A. Guerrero-Colon, E. P. Simoncelli, and J. Portilla, "Image denoising using mixtures of Gaussian scale mixtures," in *Proc. IEEE Int Conf on Image Processing*, San Diego, CA 2008.

[3] J. Portilla, V. Strela, M. J. Wainwright, and E. P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 12, pp. 1338–1351, Nov. 2003.

[4] R. W. Buccigrossi and E. P. Simoncelli, "Image compression via joint statistical characterization in the wavelet domain," *IEEE Transactions on Image Processing*, vol. 8, pp. 1688–1701, Dec. 1999.

[5] M. J. Wainwright, E. P. Simoncelli, and A. S. Willsky, "Random cascades on wavelet trees and their use in modeling and analyzing natural imagery," *Applied and Computational Harmonic Analysis*, vol. 11, pp. 89–123, Jul. 2001.

[6] A. M. R. Puri and K. Ramchandran, "PRISM: A video coding paradigm with motion estimation at the decoder," *IEEE Transactions on Image Processing*, vol. 16, Oct. 2007.

[7] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for multi-view video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 17, pp. 1474–1484, Nov. 2007.

[8] J. Portilla, V. Strela, M. Wainwright, and E. Simoncelli, "Adaptive Wiener denoising using a Gaussian scale mixture model in the wavelet domain," in *Proc. IEEE Int Conf on Image Processing*, Thessaloniki, Greece 2001.

[9] J. H. van Hateren and A. van der Schaaf, "Independent component filters of natural images compared with simple cells in primary visual cortex," in *Proc. R. Soc. Lond.*, 1998, pp. 359–366.

[10] "Kodak lossless true color image suite," available at http://r0k.us/graphics/kodak/.

[11] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *Proc. 37th Asilomar Conf on Signals, Systems and Computers*, Pacific Grove, CA 2003.