The Multilinear Compound Gaussian Distribution

Raghu G. Raj
U.S. Naval Research Laboratory, Radar Division
Washington D.C. 20375

Alan C. Bovik
Dept. of ECE, The University of Texas at Austin
Austin, TX 78712

***Abstract*— We introduce a novel generalization of the compound Gaussian (CG) (or Gaussian Scale Mixture [1]) distribution which extends the Gaussian component of the CG model to a multilinear distribution. The resulting model, which we call the Multilinear Compound Gaussian (MCG) distribution, subsumes both GSM [1] and the previously developed MICA [3-4] distributions as complementary special cases; thereby allowing us to model a richer class of stochastic phenomena. First we derive the structural characterization of the MCG distribution and develop some of its important theoretical properties. Thereafter we describe a parameter estimation algorithm for learning this model from sample data, and then deploy this for modeling textures, including natural (i.e. optical) and SAR images. Our simulation results demonstrate how, for each case, we obtain improved performance over the CG model; thus indicating the versatility of the MCG model in accurately modeling various natural phenomena of interest.**

***Index Terms*— GSM, MICA, MCG, Bayesian, Nonlinear**

I. INTRODUCTION

The compound Gaussian (CG) model—also known as the Gaussian Scale Mixture or GSM [1] in the image processing community—is a well known statistical distribution that has been employed as a powerful low-level prior model for important applications such as the denoising [2] of natural images. In the radar community too, CG is known to be an accurate statistical model for the wavelet distribution of SAR images [5] with applications such as image reconstruction [6], and also forms the basis of the K-distribution [7-8] that is frequently used for modeling RCS returns from sea clutter data. Indeed CG is a very versatile statistical model that specializes to many well known distributions such as the α-stable and symmetrized Gamma distributions [1] and which serve as useful prototypes of heavy-tailed processes—thus providing the probabilistic underpinnings for the generation of various sparse-structured stochastic phenomena such as described above.

Consider a random variable X that can be decomposed into the following product form (pointwise product of 2 vectors):

$$X = S \cdot B \tag{1}$$

The probability density function of X is thus given by:

$$P_X(x) = \int P_X(x|B=\beta)P_B(\beta)\cdot d\beta \tag{2}$$

$$= \int \frac{1}{Z} P_S(x/\beta)\cdot P_B(\beta)\cdot d\beta \tag{3}$$

where Z is a normalizing constant for $P_S(x/\beta)$.

In the special case where $S \sim N(\mu, \Sigma)$ is a Gaussian random vector with mean μ and covariance matrix Σ, and B is a non-Gaussian random vector consisting of *i.i.d.* components, X in (1) is said to CG distributed random variable:

$$P_X(x) = \int \frac{1}{\sqrt{2\pi|\tilde{\Sigma}|}} exp\left(-((x-\mu)/\beta)^H\tilde{\Sigma}^{-1}((x-\mu)/\beta)\right) P_B(\beta)\cdot d\beta \tag{4}$$

The problem that we are concerned with in this paper is the structure and properties of (1) when the conditional density $P_X(x|B=\beta)$ in (2) is replaced by *a more general multilinear distribution* of the following form:

$$P_X(x|B=\beta) = \frac{1}{Z} g(X) \prod_i p_i(x_i) \tag{5}$$

where, $X = [x_1, \ldots, x_d]$. The probability densities p_i generalize Gaussian random variables in a manner specified below, while the multilinear term $g(X)$ captures the higher-order statistical interactions between the components of X.

In Section 2, we introduce a non-linear system model that serves as the basis of our theoretical development of the MCG model and its properties. We will see how the MCG model developed subsumes both CG and the previously developed multilinear ICA (MICA) distribution [3-4] as complementary special cases. Section 3 describes an algorithm for estimating the parameters of the MCG model. We conclude by providing simulation results demonstrating the performance of the MCG distribution in Section 4 for Texture, Optical and SAR images.

II. MULTILINEAR COMPOUND GAUSSIAN MODEL

We first describe the construction of the MCG distribution in terms of a novel non-linear system model. Thereafter in section II-B we present some theoretical properties of the MCG model.

A. Structure of the MCG Distribution

Let $x = [x_1, x_2, \ldots, x_d]^T \in \mathbb{R}^d$ be an observed random vector distributed according to the MCG model (5). We now demonstrate how (5) can be synthesized via the generative model shown in Figure 1. The system F in Figure 1 consists of a core non-linearity φ preceded by a linear system $y = As +$

Supported by NRL Base Program

Γ, where $y = [y_1, y_2, \dots, y_d]^T \in \mathbb{R}^d$ and $s = [s_1, s_2, \dots, s_d]^T \in \mathbb{R}^d$ is a Gaussian random vector consisting of *i.i.d.* components $s_i \sim N(0,1)$ whose density is denoted by $q(s_i)$. Vectors $\Gamma = [\Gamma_1, \Gamma_2, \dots, \Gamma_d]^T$ and $\sigma = [\sigma_1, \sigma_2, \dots, \sigma_d]^T$ respectively determine the mean and variances for the various Gaussian channels. Matrix $C = [\mathrm{C}_1, \mathrm{C}_2, \dots, \mathrm{C}_\mathrm{d}]^\mathrm{T} = \mathrm{A}^{-1}$ is assumed to be invertible and thus determines the interaction between the various Gaussian sources. Vector $\beta = [\beta_1, \beta_2, \dots, \beta_d]^T \in \mathbb{R}^d$ is an instantiation of random variable B of the MCG model corresponding to (1). The constants γ_1 and γ_2 play an important role in determining the properties of the non-linear transformation φ, as we show in the next section. All the operations in system F, other than the action of matrix A, are pointwise in the components of the vector.

Given this, an important observation is that for a given $B = \beta$, pointwise invertibity of the non-linearity φ results in a multilinear distribution of the following form (a full derivation is given in the extended version of this paper):

$$P_X(x|B=\beta) = \frac{K}{|J(F)|}\, g(J) \prod_{k=1}^{d} p_i\left(\frac{z_k-\mu_k}{\beta_k}\right) \qquad (6)$$

where $p_k\left(\frac{z_k-\mu_k}{\beta_k}\right) = K_k exp\left(-a_k\left[\tilde{\varphi}\left(\frac{z_k-\mu_k}{\beta_k}\right) - c_k\right]^2\right)$

K_k is a normalizing constant, $a_k = (1/2)\sum_{i=1}^{d}\left(C_{i,k}^2/\sigma_k^2\right)$

and $c_k = \Gamma_k + \frac{\sum_{j\neq k}\sum_{i=1}^{d}\frac{C_{i,j}C_{i,k}}{\sigma_k^2}\Gamma_j}{\sum_{k=1}^{d}\frac{C_{i,k}^2}{\sigma_k^2}}$, $\tilde{\varphi} \equiv \varphi^{-1}$.

Also, $g(J) = exp\left[-\sum_{i\neq j} G_{i,j}\,\tilde{\varphi}\left(\frac{z_i-\mu_i}{\beta_i}\right)\tilde{\varphi}\left(\frac{z_j-\mu_j}{\beta_j}\right)\right]$

where $G_{i,j} = \sum_{k=1}^{d}\frac{C_{k,i}C_{k,j}}{\sigma_k^2}$ and $K = \frac{exp\left(-\sum_{k=1}^{d}\frac{\left(C_k^T\Gamma\right)^2}{2\sigma_k^2}\right)}{(2\pi)^{d/2}\prod_{k=1}^{d}\sigma_k K_k}$

We have found the following linear-quadratic non-linearity to effectively capture the statistics of real data:

$$\varphi(s) = \left[\gamma_1 s\sigma u_1(\sigma s) + \gamma_2 \varphi_q(s\sigma) u_2(\sigma s)\right]\beta \qquad (7)$$

where, functions u_1 and u_2 are defined in the next section, $\varphi_q(s) = s^2 sgn(s)$, $\beta \geq 0$, $\sigma \geq 0$, and $s \in \mathbb{R}$.

This non-linearity furnishes a natural generalization of the Gaussian distribution—in particular it consists of both a linear and quadratic channel *and reduces to a multidimensional Gaussian distribution when only the linear channel is active*.

B. Theoretical Properties of MCG

As described in the previous section, an important consideration is the characterization of the conditions under which equation (7) is invertible. This is because invertibility makes it possible to generate a multilinear distributions whereas in general one will obtain a mixture distributions which, though more general, are computationally less tractable. The following Lemma gives a characterization the invertibility of φ in (7) for the case where d=1 (which therefore applies to the invertibility of the general case since φ is a component-wise transform).

Lemma 1: Let $\gamma_1 u_1(s) = \lambda_1$ and $\gamma_2 u_2(s) = \lambda_2$ such that λ_1, λ_2 are real constants. Then φ in Equation (6) is invertible if and only if $\lambda_1\lambda_2 \geq 0$.

Proof:

We consider first the converse case i.e. assume $\lambda_1\lambda_2 \geq 0$:

Case 1: $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. Then equation (6) becomes:

$|x| = [\gamma_1|\sigma||s|u_1(\sigma s) + \gamma_2\sigma^2|s|^2 u_2(\sigma s)]\beta$ (since $sgn(x) = sgn(\sigma s)$)

$\Rightarrow |s|^2\sigma^2\beta\lambda_2 + \sigma\beta\lambda_1|s| - |x| = 0$

$\Rightarrow |s| = \frac{-\sigma\beta\lambda_1 \pm \sqrt{(\sigma\beta\lambda_1)^2 + 4|x|\sigma^2\beta\lambda_2}}{2\sigma^2\beta\lambda_2(s)}$

Thus the only feasible solution is:

$$|s| = \frac{-\sigma\beta\lambda_1 + \sqrt{(\sigma\beta\lambda_1)^2 + 4|x|\sigma^2\beta\lambda_2}}{2\sigma^2\beta\lambda_2}$$

Case 2: $\lambda_1 \leq 0$ and $\lambda_2 \leq 0$

In this case, $|x|sgn(x) = -sgn(s)[\gamma_1|\sigma||s|u_1(\sigma s) + \gamma_2\sigma^2|s|^2 u_2(\sigma s)]\beta$

We once again solve:

$$|x| = [\gamma_1|\sigma||s|u_1(\sigma s) + \gamma_2\sigma^2|s|^2 u_2(\sigma s)]\beta$$

to get the unique solution above together with the relationship: $sgn(x) = -sgn(\sigma s)$.

Thus in both cases φ is invertible thus proving sufficiency.

Now assume that φ is invertible. We then have 4 cases:

Case 1: $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$. See above.

Case 2: $\lambda_1 \leq 0$ and $\lambda_2 \leq 0$. See above.

Case 3: $\lambda_1 \geq 0$ and $\lambda_2 < 0$

In this case we are forced to solve the original quadratic problem wherein:

$$\lambda_2\beta\sigma^2 s^2 sgn(s) + \lambda_1\beta\sigma s - x = 0$$

To solve this, we consider 2 sub-cases:

Case 3.1: $s \geq 0$. Then:

$$\lambda_2\beta\sigma^2 s^2 + \lambda_1\sigma s - x = 0$$

$\Rightarrow s = \frac{-\sigma\beta\lambda_1 \pm \sqrt{(\sigma\beta\lambda_1)^2 + 4x\sigma^2\beta\lambda_2}}{2\sigma^2\beta\lambda_2}$

Since $s \geq 0$ by assumption, the feasible solutions are:

$$s = \frac{-\sigma\beta\lambda_1 - \sqrt{(\sigma\beta\lambda_1)^2 + 4x\sigma^2\beta\lambda_2}}{2\sigma^2\beta\lambda_2}$$

(provided: $4x\sigma^2\beta\lambda_2 > -(\sigma\beta\lambda_1)^2$) (8)

and, $s = \frac{-\sigma\beta\lambda_1 + \sqrt{(\sigma\beta\lambda_1)^2 + 4x\sigma^2\beta\lambda_2}}{2\sigma^2\beta\lambda_2}$

(provided: $4x\sigma^2\beta\lambda_2 < 0$) (9)

Both (7-8) can be satisfied if:

$$-(\sigma\beta\lambda_1)^2 < 4x\sigma^2\beta\lambda_2 < 0 \qquad (10)$$

Since $\lambda_2 < 0$ by assumption, (10) is satisfied for any $x < 0$. Thus in general there can be 2 solutions to equations (6) and so φ is not invertible for case 3.1.

Case 3.2: $s \leq 0$. Here we solve:

$$-\lambda_2\beta\sigma^2 s^2 + \lambda_1\beta\sigma s - x = 0$$

$\Rightarrow s = \frac{-\sigma\beta\lambda_1 \pm \sqrt{(\sigma\beta\lambda_1)^2 - 4x\sigma^2\beta\lambda_2}}{-2\sigma^2\beta\lambda_2} = \frac{-\sigma\beta\lambda_1 \pm \sqrt{(\sigma\beta\lambda_1)^2 + 4x\sigma^2\beta|\lambda_2|}}{2\sigma^2\beta|\lambda_2|}$

Here the only feasible solution is:

$$s = \frac{-\sigma\beta\lambda_1 - \sqrt{(\sigma\beta\lambda_1)^2 + 4x\sigma^2\beta|\lambda_2|}}{2\sigma^2\beta|\lambda_2|} \quad (11)$$

provided that $4x\sigma^2\beta|\lambda_2| > -(\sigma\beta\lambda_1)^2$ (12)

In general (11) can be violated by choosing *x* accordingly and thus φ is not invertible.

Case 4: $\lambda_1 < 0$ and $\lambda_2 \geq 0$. By a similar reasoning as applied to Case 3.1, φ is not invertible in general for this case also.

Thus φ is generally invertible only for cases (1) and (2). This establishes the lemma. □

From Lemma 1 we note that by setting $\gamma_1 = 1$ and $\gamma_2 = 0$, MCG reduces to the CG distribution as a special case. The following Lemma details a complementary direction in which the above model can be extended.

Lemma 2: Let $\lambda_1(s) = \gamma_1(s)u_1(s)$ and $\lambda_2(s) = \gamma_2(s)u_2(s)$. Then φ in (4) is invertible if $\lambda_1(s)\lambda_2(s) = 0$.

Proof: The proof follows by noting that the condition $\lambda_1(s)\lambda_2(s) = 0$ ensures that one and only one of the two channels (i.e. the linear and quadratic channels as described in II-A) are active for a given value of s. Thus invertibility of φ trivially follows. This establishes the lemma. □

We note that under Lemma 2, MCG reduces to the MICA distribution [3-4] which thus formally complements the CG model. *Thus the MCG model subsumes both the CG and MICA distributions as special cases.*

Importantly the following lemma, whose proof is similar to that of the central theorem in [3], furnishes a closed form expression for the Jacobian $|J(X)|$:

Lemma 3: The Jacobian for the MCG model is:

$$|J(X)| = |A| \prod_{k=1}^{d} \beta_k \psi(y_k) \quad (13)$$

where $\psi(y_k) = \lambda_1 + 2\lambda_2|y_k|$ □

We can easily show that Lemma 3 reduces to the central theorem in [3] for the special case of MICA distributions. The existence of the above closed form expression for $|J(X)|$ obviates the need for Monte-Carlo simulation for the estimation of system parameters. The following section delves into more details on the algorithmic aspects of parameter estimation for the MCG model.

III. MCG Parameter Estimation Algorithm

Similar to the general strategy in [1], we take a two-stage approach to estimating the parameters of the MCG model. Given a set of *N* neighborhoods (of wavelet coefficients) of size *MxM* (where N>>M; for e.g. we choose *N*=2000, *M*=3 in our simulations below) which are organized in a matrix X, we estimate the optimum parameter β (for a given set of multilinear parameters) corresponding to each such sampled neighborhood. Each neighborhood is then normalized by the corresponding β parameter. Thereafter the optimum multilinear parameters are determined in the manner described below. This process is then iterated several times until convergence.

In order to simply matters, we assume as in [1] that B is a scalar random variable. Let $x = [x_1, \ldots, x_d]^T$ denote a neighborhood sampled from the image. To determine optimum β parameter for each neighborhood we solve the following equation w.r.t. β:

$$\frac{dP_X(x|B=\beta)}{d\beta} = \sum_{k=1}^{d} \frac{a_k|x_k|}{\lambda_2\beta} f\left(\frac{x_k}{\beta}\right) + \sum_{i\neq j} G_{i,j}\left[f\left(\frac{x_i}{\beta}\right) sgn(x_j) + f(x_j/\beta) sgn(x_i)\right] + \sum_{k=1}^{d} \frac{2\lambda_2 sgn(\tilde{\varphi}(x_k/\beta))}{\lambda_1 + 2\lambda_2|\tilde{\varphi}(x_k/\beta)|} \frac{1}{\sqrt{\lambda_1^2 + 4\lambda_2|x_k/\beta|}} (x_k/\beta) - d = 0 \quad (14)$$

where, $f(x) = \frac{-\lambda_1 + \sqrt{\lambda_1^2 + 4\lambda_2|x|}}{\sqrt{\lambda_1^2 + 4\lambda_2|x|}}$. Since β is a scalar parameter, we can easily solve for β in a computationally tractable manner by employing a brute force approach of evaluating (14) for every value of β in a interval [0, max(X)] and choosing the β that renders the R.H.S. of (14) closest to zero. Furthermore it is easy to verify—both analytically and numerically—that in the *CG limit* (i.e. $\lambda_1 = 1, \lambda_2 \to 0$) the optimum solution of (14) is exactly the closed form solution of β for the GSM case [1] as one would expect.

Having obtained the set of wavelet coefficient neighborhoods normalized by random variable B, we now detail how to update the remaining parameters of the MCG model. Firstly in our simulations, consistent with the CG model of the wavelet structure of natural images, we set $\mu = 0$ and $\Gamma = 0$. Thus the parameters that remain to be estimated are matrix C, vector σ, and the scalars λ_1 and λ_2. For estimating C and σ, we employ a similar trick as used in [4]: let $y^i = \tilde{\varphi}_\lambda(x^i/\beta)$ where x^i denotes the i^{th} vector of neighborhood coefficients, and where the subscript denotes that the non-linear function $\tilde{\varphi}$ depends on $\lambda = [\lambda_1, \lambda_2]$; then set $C = \sqrt{Q}$ (where $Q = \frac{1}{N}\sum_{k=1}^{d} y^k (y^k)'$), and set σ_i to be the standard deviation of the ith component of the normalized and inverted wavelet coefficient neighborhoods $\{y^k\}_k$. Note that the computation of C and σ assumed knowledge of λ, but did not involve any iterative procedure for estimation.

To compute the optimum λ (given knowledge of the other multilinear parameters) we resort to an unconstrained iterative maximum-likelihood estimation approach based on the following descent equations:

$$\frac{dP_X(x|B=\beta)}{d\lambda_1} = -\sum_{k=1}^{d} \frac{1}{\psi(\tilde{\varphi}(x_k/\beta))}\left[1 + 2\lambda_2 sgn\left(\tilde{\varphi}(x_k/\beta)\right)\tilde{\varphi}'(x_k/\beta)\right] - \sum_{k=1}^{d} 2a_k\tilde{\varphi}(x_k/\beta)\,\tilde{\varphi}'(x_k/\beta) - \sum_{i\neq j} G_{i,j}\left[\tilde{\varphi}(x_i/\beta)\tilde{\varphi}'(x_j/\beta) + \tilde{\varphi}(x_j/\beta)\tilde{\varphi}'(x_i/\beta)\right] \quad (15)$$

A similar gradient equation can also be derived with respect to λ_2. Thus the following algorithm summarizes our MCG parameter estimation procedure:

0) Initialize β according to the CG model. From this initialize C and σ.
1) Re-estimate β by solving (14)
2) Normalize *X* according to β computed above and re-estimate C and σ accordingly
3) Estimate the optimum λ via (15). Re-estimate C and σ accordingly
4) Iterate (1)-(3) until convergence.

IV. Simulation Results

The above MCG parameter estimation algorithm was applied to a set of N=2000 neighborhoods of wavelet coefficients of size 3x3. We chose the bi-orthogonal 1.1 wavelet and chose the horizontal sub-band in our simulations (although the results hold generally for any sub-band and choice of wavelet filters). Having learnt the optimal MCG parameters (including λ) from the training set, we subsequently sample multiple set of N wavelet neighborhoods and, for each case, compute the KLD (Kullback Leibler Divergence) between the normalized empirical histograms of the β-normalized version of each wavelet coefficient and the histogram as predicted by the CG and MCG models respectively.

Tables 1-3 show respectively the average performance of CG and MCG models for the Herringbone texture [9], the hangers SAR image [10], and the Baboon Optical image. The results recorded here are statistically significant. The optimum value of λ for each case is also displayed. Since $\lambda_1\lambda_2 < 0$ for the case of Baboon image, by Lemma 1 a unique inverse cannot be guaranteed; nevertheless the generative algorithm furnishes above one possible solution that is consistent with the MCG model. From the results it is clear that the MCG distribution outperforms CG in all cases.

Thus MCG is a general model that can be used a building block for more complex probabilistic models of scene structure; and more generally opens up exciting new avenues for accurately modeling diverse stochastic phenomena.

References

[1] M. J. Wainwright and E. P. Simoncelli, "Scale mixtures of Gaussians and the statistics of natural images", in *Neural Information Processing Systems 12*, Vol. 12, pp. 855–861, 1999.

[2] J. Portilla, V. Strela, M.J. Wainwright, and E.P. Simoncelli, "Image denoising using scale mixtures of Gaussians in the wavelet domain," *IEEE Trans Image Processing*, vol.12(11), pp. 1338--1351, Nov 2003.

[3] R.G. Raj and A.C. Bovik, "MICA: A multilinear ICA decomposition for natural image modeling," *IEEE Trans. on Image Processing*, vol: 17, no: 3, March 2008, Page(s): 259-271.

[4] R.G. Raj and A.C. Bovik, "A fast multilinear ICA Algorithm," *IEEE International Conference on Image Processing*, Hong Kong 2010.

[5] Z. Chance, R.G. Raj, and D.J. Love, "Information-Theoretic Structure of Multistatic Radar Imaging," *IEEE Radar Conference*, Kansas City, May 2011.

[6] R.G. Raj, Z. Chance, and D.J. Love, "A sparse bayesian approach to multistatic radar imaging," *IEEE Asilomar Conf. on Signals, Systems and Computers*, Pacific Grove, CA, November 2011.

[7] M.W. Long, *Radar Reflectivity of Land and Sea*, Artech House, 3rd edition, 2001.

[8] J. Wang, A. Dogandzic, and A. Nehorai,"Maximum likelihood estimation of compound-gaussian clutter and target parameters," *IEEE Trans. Signal Processing*, vol. 54, no. 10, October 2006.

[9] http://sipi.usc.edu/database/database.cgi?volume=textures

[10] Sandia National Laboratories, "X-band synthetic aperture radar imagery," http://www.sandia.gov/radar/imageryx.html.

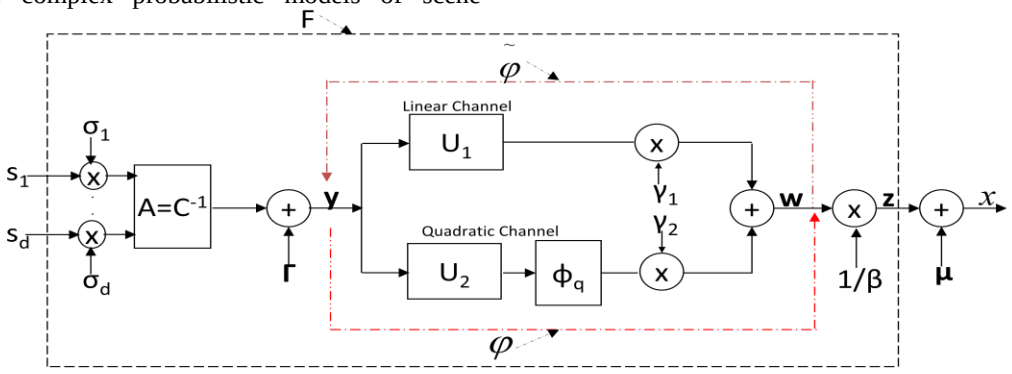

Figure 1

Avg.KLD	Channel1	Channel2	Channel3	Channel4	Channel5	Channel6	Channel7	Channel8	Channel9
CG (GSM)	0.5163	0.5706	0.4407	0.4365	0.4119	0.4377	0.4312	0.5248	0.4782
MCG	0.0266	0.0335	0.0226	0.0471	0.0611	0.0455	0.0320	0.0340	0.0326

Table 1:Herringbone Texture. Optimum $\lambda = [\lambda_1 = 0.6368, \lambda_2 = 0.0051]$. Each channel is a different node in the neighborhood structure.

Avg.KLD	Channel1	Channel2	Channel3	Channel4	Channel5	Channel6	Channel7	Channel8	Channel9
CG (GSM)	0.6231	0.3568	0.4989	0.4783	0.2797	0.4303	0.4676	0.4311	0.5609
MCG	0.0842	0.0353	0.0464	0.0784	0.0550	0.0556	0.0601	0.0587	0.0809

Table 2: Hangers SAR Image. Optimum $\lambda = [\lambda_1 = 1.4369, \lambda_2 = 0.0020]$. Each channel is a different node in the neighborhood structure.

Avg.KLD	Channel1	Channel2	Channel3	Channel4	Channel5	Channel6	Channel7	Channel8	Channel9
CG (GSM)	0.4036	0.3087	0.2989	0.2514	0.2756	0.2872	0.3322	0.2966	0.3174
MCG	0.0353	0.0254	0.0256	0.0237	0.0231	0.0251	0.0233	0.0200	0.0274

Table 3: Baboon Image. Optimum $\lambda = [\lambda_1 = 0.6018, \lambda_2 = -0.0092]$. Each channel is a different node in the neighborhood structure.