Modeling Multiscale Differential Pixel Statistics

David Odom[a] and Peyman Milanfar[a]

[a]Electrical Engineering Department, University of California,
Santa Cruz CA. 95064 USA

ABSTRACT

The statistics of natural images play an important role in many image processing tasks. In particular, statistical assumptions about differences between neighboring pixel values are used extensively in the form of prior information for many diverse applications. The most common assumption is that these pixel difference values can be described be either a Laplace or Generalized Gaussian distribution. The statistical validity of these two assumptions is investigated formally in this paper by means of Chi-squared goodness of fit tests. The Laplace and Generalized Gaussian distributions are seen to deviate from real images, with the main source of error being the large number of zero and close to zero nearby pixel difference values. These values correspond to the relatively uniform areas of the image. A mixture distribution is proposed to retain the edge modeling ability of the Laplace or Generalized Gaussian distribution, and to improve the modeling of the effects introduced by smooth image regions. The Chi-squared tests of fit indicate that the mixture distribution offers a significant improvement in fit.

1. INTRODUCTION

Natural image statistics provide a useful tool for many image processing tasks. The statistics of natural images have been studied extensively, and a survey of this work can be found in.[1] This paper focuses on parametric statistical models for the differences between neighboring pixels. These models find frequent use as prior information for many different applications. This spans a wide range of uses including denoising, superresolution, compression, and many others. Various distributions have been proposed to model the outputs of bandpass filters applied to images. In this work, derivative statistics are the particular aspect of image statistics under study. The derivative statistics have been modelled by the Generalized Gaussian distribution in[2] and utilized in many other works, such as in[3] and in.[4] The Laplace assumption has been proposed explicitly,[5] and implicitly by applications such as total variation denoising.[6] The statistical validity of these assumptions has never been thoroughly investigated. This paper provides an analysis of how well these proposed distributions fit over a wide variety of images by means of Chi-squared tests of fit. A new distribution, consisting of the mixture of a Gaussian and a Laplacian is proposed and analyzed under the same framework. The calculated Chi-squared figures are used to make a comparison of fit between the distributions, similarly to the distribution tests performed for DCT coefficient statistics in.[7] Section 2 contains a description of the pixel differences being studied. Section 2.1 describes the distributions used to describe the pixel differences. The behavior across scale is discussed in Section 2.2, and Section 3 describes the statistical goodness of fit tests of these distributions. Section 4 contains a summary and future work.

2. PIXEL DIFFERENCES AND DISTRIBUTION STATISTICS

To obtain the data for which the tests of goodness of fit are to be performed, a sample set of 300 grayscale images with pixel values in the range [0,1] was used. Each image was shifted vertically and horizontally by a maximum of three pixels, and then subtracted from the original. No reflection was used at border regions; the areas with no overlap were simply eliminated from the pixel difference data set. A sampling of 20 images from the test set is shown in Figure 1.

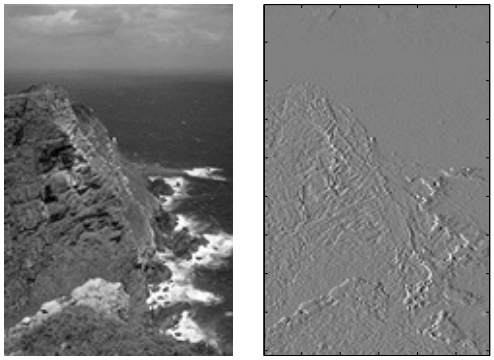
Following the notation of,[8] these pixel differences for a given image, $\mathbf{X}$, can be represented as follows

Further author information:
D.O. : E-mail: davyodom@soe.ucsc.edu, Phone: (831)459-4141, Fax: (831)-459-4829
P.M. : E-mail: milanfar@soe.ucsc.edu, Phone: (831)-459-4929, Fax: (831)-459-4829

Figure 1. Sample of images tested

$$\underline{X}_{(l,m)} = \underline{X} - S_x^l S_y^m \underline{X}, \tag{1}$$

where $\underline{X}$ is a column-stacked vector of the original image. S_x^l and S_y^m are matrices that shift $\mathbf{X}$ by l and m pixels in horizontal and vertical directions respectively. $\underline{X}_{(l,m)}$ is the vector form of image $\mathbf{X}$ shifted horizontally by l pixels and vertically by m pixels and subtracted from the unshifted $\mathbf{X}$. For instance, an image from the test database, along with the difference image $\underline{X}_{(1,0)}$, are shown in Figure 2.



a b

Figure 2. a: Original Image, b: $\underline{X}_{(1,0)}$

2.1. Distributions

The pixel differences defined above have been modeled by several different distributions in the past. We present these distributions along with a proposed alternate distribution below. All distributions are assumed to be zero-mean*. In what follows, we model the pixels of the difference images under study by the random variable

*This assumption was confirmed as valid by allowing the mean to vary for each distribution and comparing to the zero-mean case. The estimted means were indeed found to be esentially zero in all cases.

x, with n samples denoted by $\{x_i\}$, $0 < i < n$.

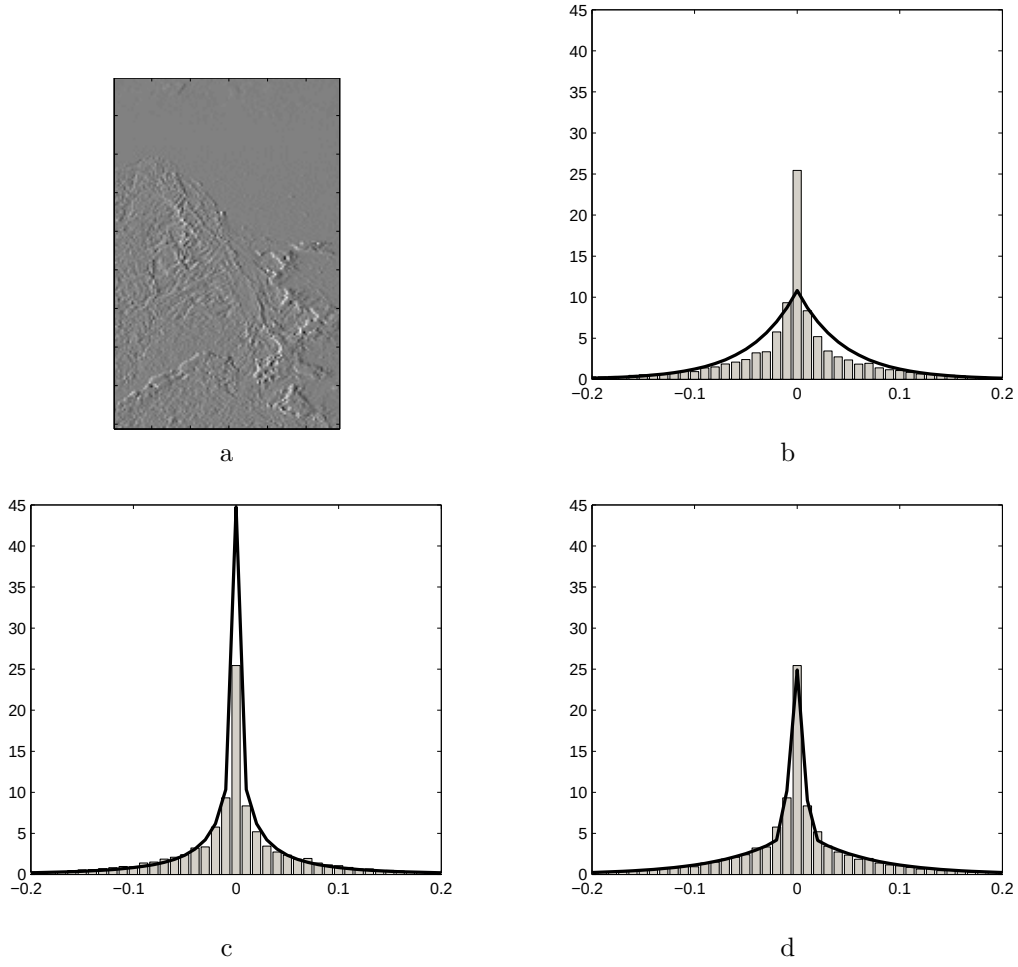


Figure 3. a: $\underline{X}_{(1,0)}$, b: Laplace ML estimate, c: Generalized Gaussian ML estimate, and d: Mixture model ML estimate

2.1.1. Laplace Distribution

The zero-mean classical Laplace distribution is given by the density function

$$p_L(x; s) = \frac{1}{2s} e^{-|x|/s}, \qquad -\infty < x < \infty \tag{2}$$

where $s > 0$ is the scale parameter.

The maximum likelihood[†] estimate for the scale parameter is given as follows

$$\hat{s} = \sum_{i=1}^{n} \frac{|x_i|}{n} \tag{3}$$

In Figure 3b, the histogram of pixel difference values for the pixel difference image of Figure 3a, along with the superimposed Laplace ML fit, are shown.

[†]Maximum likelihood (ML) estimators are used due to their desirable properties of consistency, asymptotic normality, and efficiency.[9]

2.1.2. Generalized Gaussian Distribution

The Generalized Gaussian (GG) Distribution is given by the following density function

$$p_{GG}(x;\alpha,\beta) = \frac{\beta}{2\alpha\Gamma(\frac{1}{\beta})} \exp\left(-\left|\frac{x}{\alpha}\right|^{\beta}\right) \tag{4}$$

For the special case of $\beta = 1$ the GG distribution becomes a Laplace distribution, and for $\beta = 2$, it becomes a Gaussian distribution.

The ML estimate $\hat{\beta}$ is the solution of the following transcendental equation

$$1 + \frac{\psi(1/\hat{\beta})}{\hat{\beta}} - \frac{\sum_{i=1}^{n} |x_i|^{\hat{\beta}} \log|x_i|}{\sum |x_i|^{\hat{\beta}}} + \frac{\log\left(\frac{\hat{\beta}}{n}\sum_{i=1}^{n} |x_i|^{\hat{\beta}}\right)}{\hat{\beta}} = 0 \tag{5}$$

The Newton-Raphson iterative procedure as implemented by Do and Vetterli[10] can be used to solve this equation. With $\hat{\beta}$ determined, $\hat{\alpha}$ is given by

$$\hat{\alpha} = \left(\frac{\hat{\beta}}{n}\sum_{i=1}^{n} |x_i|^{\hat{\beta}}\right)^{1/\hat{\beta}} \tag{6}$$

In Figure 3c, the histogram of pixel difference values for the pixel difference image of Figure 3a along with the superimposed Generalized Gaussian ML fit are shown.

2.1.3. Laplace-Gauss Mixture

Several observations can be made about the pixel difference data and the ML fits for the Laplace and Generalized Gaussian distributions shown above. It appears that the tails of the distributions describe the data well, but the center region is not described well by either distribution. In other words, the detail or edge regions of the image are well described, but the relatively uniform areas are not. Motivated by these observations, we suggest a new distribution which consists of a Laplace distribution to model the edge behavior, and a Gaussian distribution to model the more uniform regions.

The mixture distribution is defined as follows,

$$p(x;\underline{\theta}) = Ap_L(x;s) + (1-A)p_G(x;\sigma^2), \tag{7}$$

where $\underline{\theta} = [s, \sigma^2, A]$, p_L is the Laplace distribution defined in Equation 2, and p_G is a zero mean Gaussian distribution defined by

$$p_G(x_i;\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{-x^2}{2\sigma^2}}. \tag{8}$$

The parameters for the mixture distribution are estimated by using the EM algorithm.[11] Namely, the mixing parameter, A, is fixed, and the ML estimates for the two distributions are calculated. Next, the mixing parameter is estimated with the other distribution parameters fixed at the new estimate. This process is repeated until convergence.[11]

In Figure 3d, the histogram of pixel difference values for the pixel difference image of Figure 3a, along with the superimposed Laplace/Gauss mixture ML fit, are shown.

2.2. Parameter Behavior Across Scale

As the scale of the pixel difference varies from fine to coarse (increasing l and m from Equation 1) the parameters of each distribution vary as well. In order to understand the overall behavior of these distributions, it is desirable to describe each parameter's behavior as some simple function of scale. Empirically, each parameter was observed to exhibit exponential, logarithmic, or linear behavior.

In particular, the s parameter of the Laplace distribution and the Laplace component of the mixture model was observed to grow according to the exponential relationship

$$s(l,m) = s_0 \exp\left(\frac{\gamma(|l|+|m|-1)}{|l|+|m|}\right), \qquad |l|+|m| > 0 \tag{9}$$

where s_0 is the s value at the finest scale ($|l|+|m|=1$).

The α parameter of the Generalized Gaussian distribution exhibited logarithmic growth of the form

$$\alpha(l,m) = \alpha_0 \frac{\ln(\zeta(|l|+|m|))}{\ln(\zeta)}, \qquad |l|+|m| > 0 \tag{10}$$

where α_0 is the α parameter value at the finest scale.

The β parameter of the Generalized Gaussian distribution grows linearly,

$$\beta(l,m) = \beta_0 + \eta(|l|+|m|-1), \qquad |l|+|m| > 0 \tag{11}$$

where β_0 is the β parameter value at the finest scale.

The variance of the Gaussian component of the mixture model behaves similarly to α of the Generalized Gaussian distribution.

$$\sigma^2(l,m) = \sigma_0^2 \frac{\ln(\zeta(|l|+|m|))}{\ln(\zeta)}, \qquad |l|+|m| > 0 \tag{12}$$

where σ_0^2 is the variance at the finest scale.

The mixing parameter, A, exhibits linear growth,

$$A(l,m) = A_0 + \eta(|l|+|m|-1), \qquad |l|+|m| > 0 \tag{13}$$

where A_0 is the mixing parameter value at the finest scale.

It is informative to observe the specific case of parameter growth for the average ML parameter estimates. These averages are taken for each parameter of each distribution across the entire set of test images. In order to fit the models of Equations 9-13 to the average parameter behavior, γ, ζ, or η was determined for each case by a least-squares criterion.

For the Laplace distribution, $s_0 = 0.0457$ and the exponential growth parameter γ is 0.6731. The normalized[‡] average s values for each scale along with the approximated exponential fit are shown in Figure 4a.

For the Generalized Gaussian distribution, $\alpha_0 = 0.0219$ and the parameter dictating its growth is given by $\zeta = 4.327$. $\beta_0 = 0.6022$ and the associated linear slope is $\eta = 0.0434$. The normalized average parameters for the Generalized Gaussian are shown with their approximations in Figure 4b.

For the mixture model, $s_0 = 0.0620$, and the corresponding γ parameter is 0.6534. It is interesting to note that the rate of growth of the s parameter of the Laplace component of the mixture model is very close to the

[‡]Each parameter and its estimate are divided by the corresponding minimum parameter value for clarity in Figures 4a,b, and c.

rate of growth of the s component of the Laplace distribution alone. The variance of the Gaussian component is dictated by $\sigma_0^2 = 0.00075$ and $\zeta = 2.0157$. The mixing parameter, A, is seen to have an η value of 0, i.e. constant,with $A_0 = 0.6795$. The normalized parameters for the mixture model along with their least-squares fits are shown in Figure 4c.

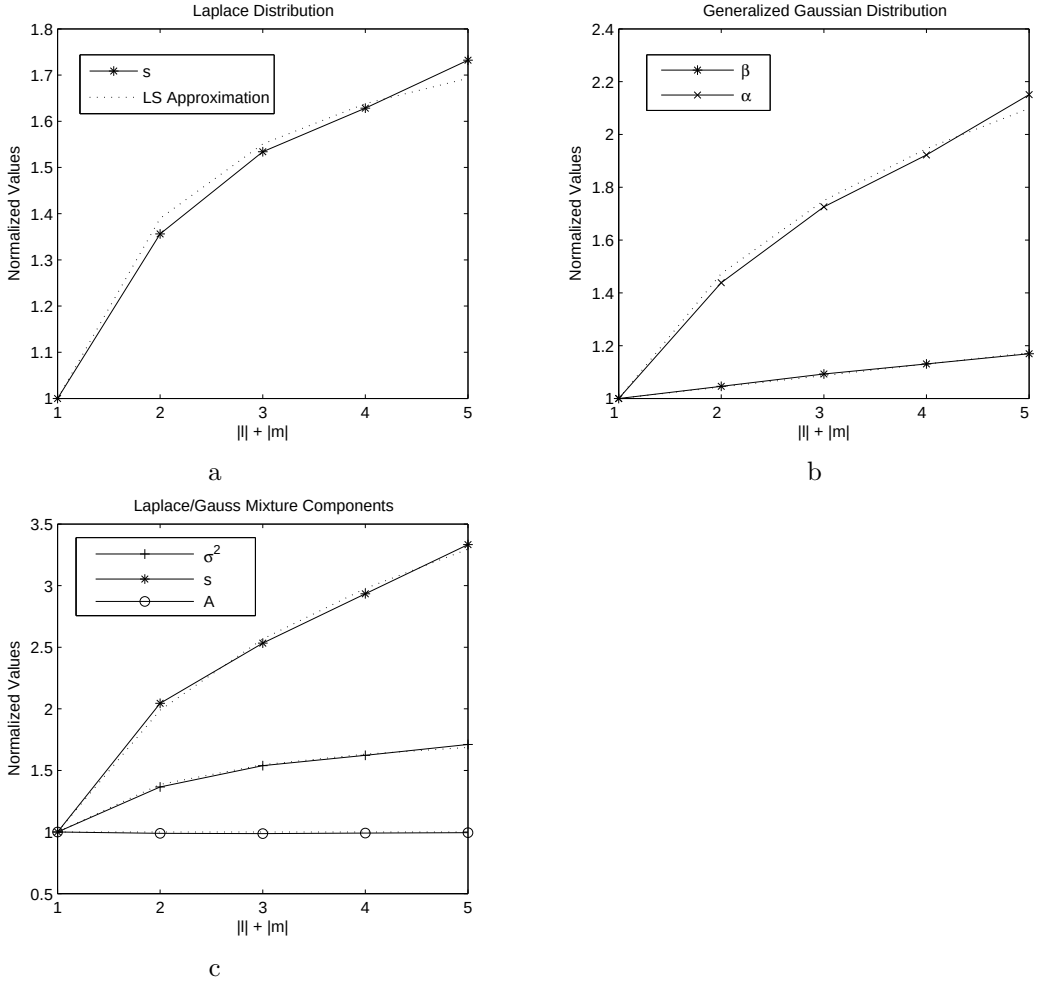


Figure 4. Parameter Behavior Across Derivative Scale

The results presented above give insight into the behavior of the overall model. Several observations can be made about the average behavior of the mixture model. The mixing parameter can be seen to describe what percentage of the image is described by the Laplacian component, and what percentage by the Gaussian component. On average, $A \approx 0.64$ with a slope of 0 across scale, thus the Laplacian component describes approximately 64% of the image at *every* scale. The Laplace component has heavier tails at coarser scales, by an amount dictated by the relationship expressed in Equation 9. This reflects a loss of detail as scale becomes more coarse. Similarly, the Gaussian component has a larger variance at coarser scales, by an amount dictated by the relationship of Equation 12, reflecting an increase in local variance of the relatively uniform image regions.

3. GOODNESS OF FIT TESTS

Visually observing the "quality" of fit of probabilistic distributions to pixel difference data is not sufficient to evaluate the use of each distribution as a prior. A statistical test is needed to quantitatively evaluate the goodness of fit, and to give a means of comparison between distributions. The Chi-squared test of fit[7,9] was employed

to test the goodness of fit for all distributions. Each specified distribution, $F_0(x)$, was divided into k classes, which were taken to be successive intervals in the range of x. The probability of an observation coming from each class can be easily calculated from the assumed distribution function, and is denoted by $p_{0i}, i = 1, 2, \ldots, k$. The observed frequency in each class is given by n_i, with n total observations. The Chi-squared test statistic follows, approximately, a Chi-squared distribution with $(k - c - 1)$ degrees of freedom where c is the number of estimated parameters. The Chi-squared statistic is then given by the following formula

$$\chi^2 = \sum_{i=1}^{k} \frac{(n_i - np_{0i})^2}{np_{0i}}. \tag{14}$$

The Chi-squared figures were calculated for every image in the test set and averaged for each scale of pixel difference. k was set to 75 for all calculations to facilitate comparison. The size of k is proportional to the size of the data sets and a minimum expected frequency per bin is assured. The results are shown in Figure 5.

Lower Chi-squared figures indicate a better fit. The Chi-squared values confirm what Figures 3b, c, and d suggest. The Laplace/Gauss mixture model offers the best fit at all scales tested, even with the additional recovery of one degree of freedom over the Generalized Gaussian distribution taken into account. Tables 1, 2, and 3 contain the average Chi-squared figures for each distribution across the set of test images. All Chi-squared figures indicate that on average, for the degrees of freedom used in the test, the hypothesis that the data is derived from any of the proposed distribution is rejected, even at a significance level of 0.01. This lack of statistical significance does not imply that these distributions are not useful as priors, only that they are not perfect statistical descriptors of the data. However, the relative fit of each distribution is apparent in Figure 5.

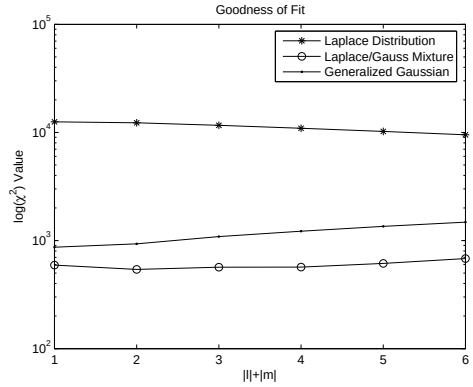


Figure 5. Average Chi-squared figures.

4. CONCLUSION

Several distributions were presented that have been used to describe pixel difference statistics. A new distribution that consists of the mixture of a Laplace component and a Gauss component was proposed as an alternative distribution. The behavior of each distribution was analyzed across scale, and simple models were presented to explain this behavior. A Chi-squared test of fit was performed for each distribution across scale. The results of these tests indicate that the proposed mixture model presents the best fit on average for each scale tested. The logical extension of this work would be to take the model presented and use it as a prior in image processing tasks. As mentioned in the introduction, a wide range of applications utilize assumptions on the pixel difference distributions. A prior that is a better fit should provide better results for any of these applications, if a computationally efficient framework can be found.

ACKNOWLEDGMENTS

This work was supported in part by the US Air Force Office of Scientific Research Grant F49620-03-1-0387

	l=0	l=1	l=2	l=3
m=0		s=0.0445 χ^2=13058	s=0.0630 χ^2=13230	s=0.0705 χ^2=12282
m=1	s=0.0469 χ^2=12019	s=0.0567 χ^2=11626	s=0.0671 χ^2=11691	s=0.0738 χ^2=11160
m=2	s=0.0670 χ^2=11991	s=0.0694 χ^2=11363	s=0.0742 χ^2=11083	s=0.0792 χ^2=10307
m=3	s=0.0758 χ^2=11178	s=0.0779 χ^2=10526	s=0.0813 χ^2=10122	s=0.0849 χ^2=9511

Table 1. Average Laplace Parameters

	l=0	l=1	l=2	l=3
m=0		α=0.0205 β=0.5790 χ^2=786.1	α=0.0303 β=0.6052 χ^2=888.8	α=0.0353 β=0.6315 χ^2=1005.6
m=1	α=0.0234 β=0.6254 χ^2=947.1	α=0.0286 β=0.6318 χ^2= 868.6	α=0.0341 β=0.6377 χ^2= 953.2	α=0.0390 β=0.6510 χ^2= 1053.0
m=2	α=0.0355 β=0.6529 χ^2=1037.1	α=0.0376 β=0.6660 χ^2=1066.2	α=0.0410 β=0.6758 χ^2=1125.2	α=0.0450 β=0.6897 χ^2=1232.1
m=3	α=0.0440 β=0.6973 χ^2=1324.0	α=0.0463 β=0.7099 χ^2=1478.4	α=0.0491 β=0.7186 χ^2=1473.9	α=0.0524 β=0.7296 χ^2=1478.9

Table 2. Average Generalized Gaussian Parameters

	l=0	l=1	l=2	l=3
m=0		s=0.0599 σ^2=7e-4 A=0.6908 χ^2=609.74	s=0.0854 σ^2=1.6e-3 A=0.6774 χ^2=515.07	s=0.0949 σ^2=1.9e-3 A=0.6756 χ^2=496.46
m=1	s=0.0642 σ^2=8e-4 A=0.6683 χ^2=575.75	s=0.0770 σ^2=1.3e-3 A=0.6743 χ^2=508.56	s=0.0911 σ^2=1.7e-3 A=0.6718 χ^2=486.67	s=0.0987 σ^2=2.1e-3 A=0.6763 χ^2=482.75
m=2	s=0.0919 σ^2=1.7e-3 A=0.6641 χ^2=597.73	s=0.0944 σ^2=1.8e-3 A=0.6656 χ^2=548.23	s=0.0997 σ^2=2.0e-3 A=0.6734 χ^2=537.97	s=0.1050 σ^2=2.4e-3 A=0.6758 χ^2=541.86
m=3	s=0.1006 σ^2=2.2e-3 A=0.6710 χ^2=734.86	s=0.1037 σ^2=2.5e-3 A=0.6702 χ^2=683.79	s=0.1073 σ^2=2.9e-3 A=0.6765 χ^2=685.94	s=0.1110 σ^2=2.5e-3 A=0.6780 χ^2=680.6

Table 3. Average Mixture Model Parameters

REFERENCES

1. A. Srivastava, A. Lee, E. Simoncelli, and S.-C. Zhu, "On advances in statistical modeling of natural images," *Journal of Mathematical Imaging and Vision* **18**, pp. 17–33, 2003.
2. S. Mallat, "A theory for multiresolution signal decomposition: The wavelet representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **11**, pp. 674–693, July 1989.
3. J. Huang and D. Mumford, "Statistics of natural images and models," in *CVPR*, **1**, pp. 541–547, June 1999.
4. E. Simoncelli and E. Adelson, "Noise removal via Bayesian wavelet coring," in *Proceedings of 3rd IEEE International Conference on Image Processing*, **1**, pp. 379–382, 1996.
5. M. Green, "Statistics of images, the TV algorithm of Rudin-Osher-Fatemi for image denoising and an improved denoising algorithm," report, UCLA CAM 02-55, Oct. 2002.
6. L. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D* **60**, pp. 259–268, Feb. 1992.
7. J. Chang, J. Shin, N. Kim, and S. Mitra, "Image probability distribution based on generalized gamma function," *IEEE Signal Processing Letters* **12**, pp. 325–328, Apr. 2005.
8. S. Farsiu, D. Robison, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing* **13**, pp. 1327–1344, Oct. 2004.
9. A. Stuart, J. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics*, vol. 2A, Arnold, sixth ed., 1999.
10. M. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized gaussian density and kullback-leibler distance," *IEEE Transactions on Image Processing* **11**, pp. 146–158, Feb. 2002.
11. A. Dempster, N. Laird, and D. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the Royal Statistical Society, Series B* **39**(1), pp. 1–38, 1977.