Denoising Archival Films using a Learned Bayesian Model

Teodor Mihai Moldovan, Stefan Roth and Michael J. Black

Department of Computer Science
Brown University
Providence, Rhode Island 02912

CS-07-03
September 2007

Denoising Archival Films using a Learned Bayesian Model*

Teodor Mihai Moldovan Stefan Roth† Michael J. Black
Department of Computer Science
Brown University, Providence, RI, USA
teodor_moldovan@brown.edu; roth@cs.brown.edu; black@cs.brown.edu

December 31, 2007

Abstract

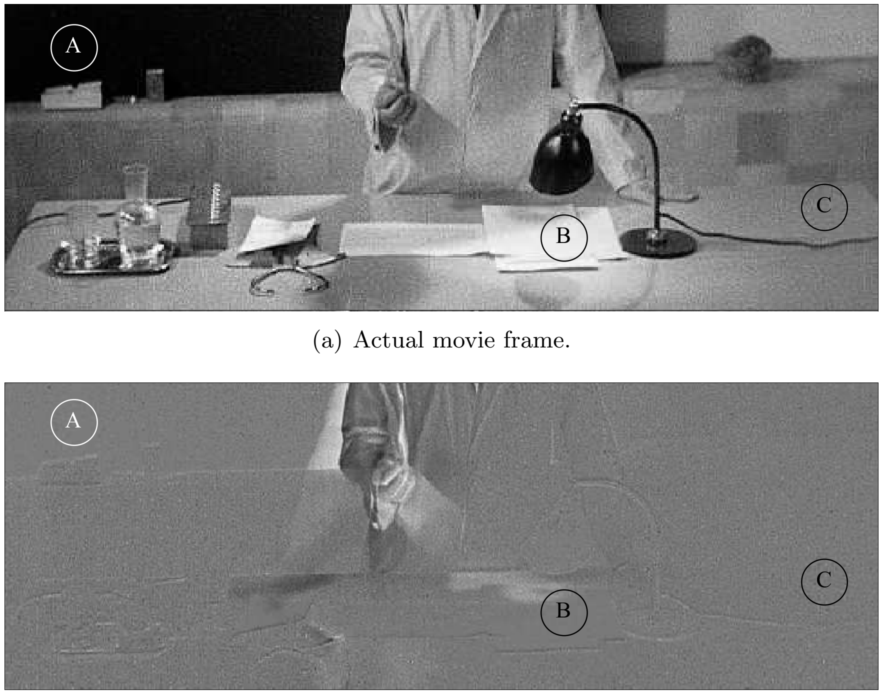
We develop a Bayesian model of digitized archival films and use this for denoising, or more specifically de-graining, individual frames. In contrast to previous approaches our model uses a learned spatial prior and a unique likelihood term that models the physics that generates the image grain. The spatial prior is represented by a high-order Markov random field based on the recently proposed Field-of-Experts framework. We propose a new model of the image grain in archival films based on an inhomogeneous beta distribution in which the variance is a function of image luminance. We train this noise model for a particular film and perform de-graining using a diffusion method. Quantitative results show improved signal-to-noise ratio relative to the standard *ad hoc* Gaussian noise model.

1 Introduction

The restoration of archival film footage requires solutions to a number of problems including the removal of scratches, dirt (e. g., hairs), and film grain resulting from the photographic process and digitization. In this paper we consider the problem of removing film grain and focus on the development of a Bayesian model that is learned from example images. In particular, we develop this model in the context of denoising individual film frames and leave the problem of temporal modeling for future work. The main contribution is the development of a physically-motivated noise model for film grain. We show that the noise in films can be modeled using an inhomogeneous beta distribution in which the variance of the noise is a function of image luminance. In particular, due to the physics of the film imaging process the variance of the noise decreases for very low and very high brightness values. We model the prior probability of natural images using the Field-of-Experts (FoE) model [Roth and Black, 2005a], a high-order Markov random field model for images and other dense scene representations that is based on extended cliques and is learned from a database of natural images. Roth and Black [2005a, 2007] showed that the FoE model gives state-of-the-art performance on various low-level vision tasks including denoising. As opposed to the simple Gaussian noise model used in [Roth and Black, 2005a], here a learned beta noise model is used to construct an image-dependent likelihood model, which is combined with the rich FoE spatial prior to give a fully learned Bayesian model of archival film frames.

*This is an extended version of [Moldovan et al., 2006].

†Stefan Roth is now with the Department of Computer Science, GRIS, TU Darmstadt, Germany. Email: stefan.roth@gris.informatik.tu-darmstadt.de



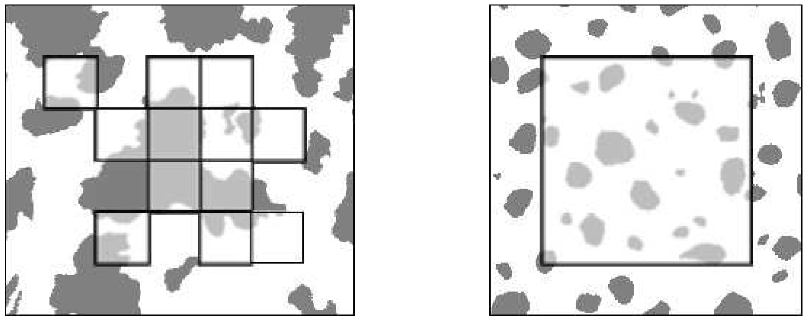
(a) Actual movie frame.

(b) Difference image revealing the grain (see text).

Figure 1: **Example frame (partially visible) from an archival film.**

1.1 Previous work

The removal of noise produced by film grain has a long history and previous authors have noted the dependence of the noise variance on image luminance [Naderi and Sawchuk, 1978; Tavildar et al., 1985; Sadhar and Rahagopalan, 2005]. Most recently, Liu et al. [2006] and Schallauer and Mörzinger [2006] presented methods for estimating the noise dependence on the luminance from a single frame. These previous methods, however, have used simplified models that assume Gaussian noise where the variance is a function of luminance. Typical denoising techniques exploit such a Gaussian noise model to perform Wiener filtering [Naderi and Sawchuk, 1978] or MAP estimation [Tavildar et al., 1985]. More recent work has used similar image models, but proposed better inference methods [Sadhar and Rahagopalan, 2005]. Work on removing film grain can be categorized into purely spatial [Naderi and Sawchuk, 1978; Tavildar et al., 1985; Sadhar and Rahagopalan, 2005] and spatio-temporal approaches [Kokaram and Godsill, 2002]. As already discussed, we are focusing on the spatial case here. In contrast, however, we model the non-Gaussian nature of the noise using an inhomogeneous beta distribution where the dependence between the noise and the luminance is learned from training images.



(a) Fine scale discretization. (b) Coarse discretization.

Figure 2: **Simplified model of film grain structure at two different discretization scales.** We assume film is composed of completely opaque or transparent regions. Fine scale discretization (a), such as in high-definition video, results in spatially correlated noise. More common to NTSC video is the coarse discretization scenario in (b), which produces uncorrelated noise.

2 Learned Likelihood Model of Film Grain

2.1 Physical motivation

Photographic film [Dainty and Shaw, 1974] is a strip of plastic covered by silver halide salts. The crystal sizes and structures determine the resolution and the sensitivity of the film. When exposed to radiation (in the visible or other spectrum), the salts release atomic silver that forms the latent image. After the process of film developing, the silver forms a metallic structure that blocks light. The density of the metallic silver depends on the amount of radiation absorbed by the film. As the intensity of the light increases, so does the density of the silver structure.

If the light is too intense, then the film achieves its minimum transparency. We call this the superior saturation point (SSP), which corresponds to the brightest tone possible (film regions in this condition are called overexposed). If the light is too dim, then the film achieves its maximum transparency. We call this the inferior saturation point (ISP), which corresponds to the darkest tone possible. The existence of these absolute minimum and maximum intensities implies that the pixel value distribution, and, thus, the noise distribution, has to have finite support. In particular, the noise distribution can not be Gaussian since the Gaussian distribution has infinite support.

Depending on the type of film we consider and the resolution of the digital scanner, the images could have spatially correlated or spatially uncorrelated grain structure. Figure 2 illustrates this for two different scanning resolutions. For NTSC resolution video and standard film, the discretized pixel size will always be significantly larger than the physical structure of the film grain, which results in spatially uncorrelated noise. The case of high-definition (HD) video is becoming increasingly important with the advent of commercially available HD-TV sets. Thus modeling film grain at HD resolution is an important application of advanced observation models that model the correlation structure of the noise. Previous work on modeling correlated noise for example includes [Kelly et al., 1988; Lee, 1998; Tsuzurugi and Okada, 2002; Portilla, 2004]. Nonetheless, we will not consider this case here, but point out that this is a valuable avenue of future work.

It is also important to note that film grain is temporally uncorrelated since each frame is produced by exposing an independent piece of film (this assumes that there is no pre-processing of the digitized film such as frame rate conversion).

2.2 Noise dependence on luminance

From a simple, approximate model of the physics of photographic film it is easy to see that there is a relationship between the variance from the grain (noise) and the brightness of a film patch. Assume that the film negative is covered with perfectly opaque regions of area B (ISP regions) and perfectly transparent regions of area W (SSP regions). Let the total area of the digitized patch of film be 1 so that $B + W = 1$. If we measure the intensity at a point in the patch we will measure 1 with probability W and 0 with probability B. Hence, the mean and the variance of the measurement are

$$\mu = 0 \cdot B + 1 \cdot W \qquad \sigma^2 = B(0-\mu)^2 + W(1-\mu)^2.$$

Combining these two equations, we get $\sigma^2 = \mu \cdot (1-\mu)$ without having to assume any particular distribution. Note that the maximum variance according to this equation is 1/4. To make the model more general we add the scale parameter $\sigma^2_{\max}$ representing an arbitrary maximum variance. We obtain the generic noise variance as a function of the mean luminance:

$$\sigma^2(\mu) = 4\sigma^2_{\max}\mu(1-\mu). \tag{1}$$

We thus find that the variance is a function of the average luminance over a patch, which is shown in Figure 4(a). In reality, film has complicated physical properties [Dainty and Shaw, 1974] and the process of digitization is also not an ideal sampling process. Hence, we expect real noise to deviate somewhat from this generic form.

2.3 Data exploration

Consider the example from a black-and-white film in Figure 1(a), which is taken from the movie "Das Testament des Dr. Mabuse" (Fritz Lang, director, 1933)[1]. This image is taken from a sequence of 400 frames in which the camera does not move (though the actor in the center does). Given the temporal independence of the grain we averaged the first 50 frames to obtain a mean image in which the grain in the static regions is effectively removed. Subtracting the top image from the mean reveals the pixel noise as shown in Figure 1(b). We use $\mathbf{m}$ to denote the mean image, $\mathbf{x}^{(t)}$ to denote the t-th frame, and $\mathbf{d}^{(t)}$ to denote the difference between the mean and frame t.

Ignoring the moving regions, the difference image reveals many of the properties of the grain. As predicted by the preceding derivation, there is almost no noise in the regions A and B marked in Figure 1, while region C is relatively noisy. This is because region A is close to the ISP while region B is close to the SSP.

Figure 3 shows histograms of the brightness values in each of these roughly uniform regions. In very bright and in very dark regions we find that the distribution of pixel values is skewed, because the range of admissible pixel values is limited. This motivates us to employ a model that allows for skewed distributions on a bounded interval.

2.4 A probabilistic model of film grain

Given the physical process and the observed noise statistics described above, we propose a model of image noise for film grain. For this model we assume that the film exhibits spatially independent

[1]The images used here are courtesy of Criterion Collection.

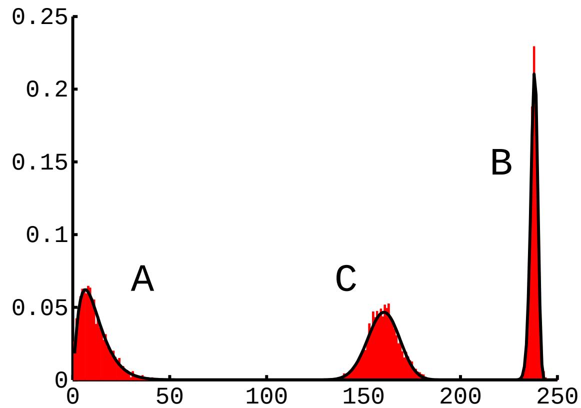


Figure 3: **Pixel value histograms for regions A, B, C represented in Fig. 1.** The black lines show the fit of the developed beta likelihood model to the pixels in each region.

noise due to the large relative size of the image pixels to the grain. For convenience of notation, we rescale and shift the image intensities to lie on the interval $[0, 1]$. We use $\mathbf{y}$ to represent a grainy image produced through a photographic process from the true (original) image denoted $\mathbf{x}$. Using the assumption of spatial independence we can write the likelihood of the noisy image given the underlying true image (in frame 0) as

$$p(\mathbf{y}^{(0)}|\mathbf{x}^{(0)}) = p(\mathbf{y}|\mathbf{x}) = \prod_{i=1}^{M} p(y_i|x_i). \tag{2}$$

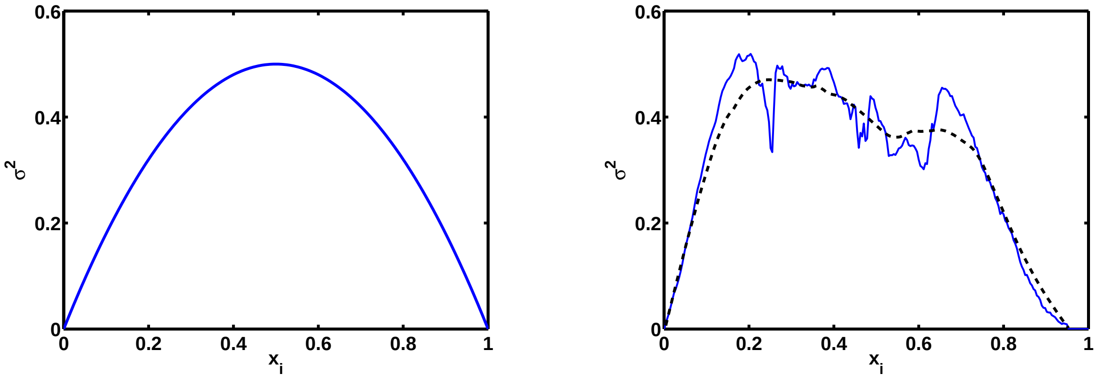
We claim that the pixels y_i are well modeled by an inhomogeneous beta distribution [Gelman et al., 2004, §A.1]

$$p(y_i|x_i) = \frac{\Gamma(\alpha(x_i) + \beta(x_i))}{\Gamma(\alpha(x_i)) \cdot \Gamma(\beta(x_i))} \cdot y_i^{\alpha(x_i)-1} \cdot (1 - y_i)^{\beta(x_i)-1}, \tag{3}$$

where Γ is the gamma function and $\alpha(x_i)$, $\beta(x_i)$ are the two parameters of the beta distribution, which here depend on the gray value of the true pixel x_i. We will discuss this dependence in more detail below, but we should already note that the noise it describes is not additive. We should also note here that the normalization term in this model cannot be ignored as in a homogeneous Gaussian likelihood, because the normalization factor depends on the true image $\mathbf{x}$.

The beta distribution is preferable over an inhomogeneous Gaussian distribution (used, e. g., in [Sadhar and Rahagopalan, 2005]), because the beta distribution has a finite support and models the skew that we empirically observe. In particular, for bright and dark tones it is skewed away from the respective end of the brightness range, and in the mid tones the distribution is mostly symmetric and has a shape similar to a standard Gaussian. The black lines in Fig. 3 show the fits of beta distributions to the samples from the respective image region. As we can see, the fit models the varying skew of the empirical data quite well.

To complete the model we assume that the mean value of y_i is equal to x_i (the true intensity value) and that the variance of y_i is a function of x_i. Using the relation between the parameters



(a) Generic noise variance function (scaled).

(b) Learned noise variance function.

Figure 4: **Generic and learned noise variance functions.** The learned noise variance function is trained on 50 frames of the sequence after correcting for global luminance changes. The dotted line represents a smoothed version (using Gaussian kernels).

α, β and the mean and variance (see Appendix A), we can express α and β as

$$\alpha(x_i) = x_i \cdot \left(\frac{x_i(1-x_i)}{\sigma^2(x_i)} - 1 \right) \tag{4}$$

$$\beta(x_i) = (1-x_i) \cdot \left(\frac{x_i(1-x_i)}{\sigma^2(x_i)} - 1 \right). \tag{5}$$

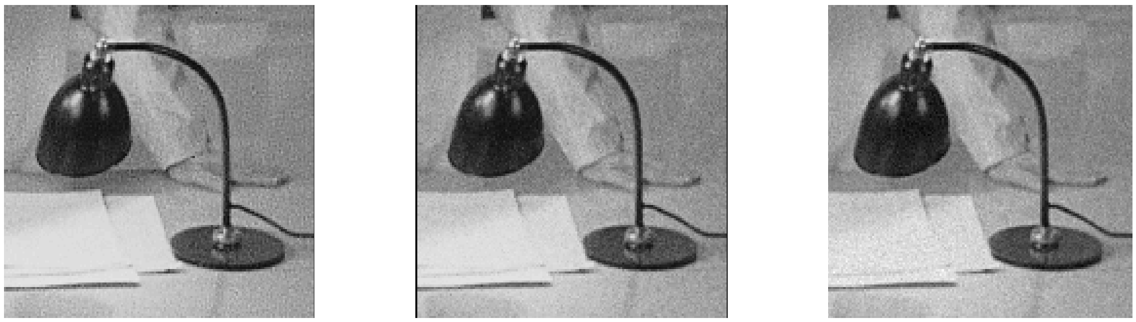
Our model has a functional parameter $\sigma^2(x_i)$ that we call the *noise variance function* (Liu et al. [2006] calls it the noise level function). We expect that the real noise variance function will look similar to the generic one derived above.

2.5 Learning the model parameters

We can train the model for a particular sequence by fitting the noise variance function to the data. In particular, we used 50 frames from our sequence, found regions with no motion, and computed the mean image $\mathbf{m}$. The regions with no motion are sufficiently large to cover the full range of gray values. We correct the sequence for global illumination changes, take the difference images $\mathbf{d}^t$ and compute the standard deviation of the noise for each gray value. The resulting variance function is plotted in Figure 4(b) together with a smoothed version that is obtained by putting a Gaussian kernel on each bin of the measured noise standard deviation:

$$\sigma^2(x_i) = (\sigma(x_i))^2 = \left(\sum_l \sigma_l \cdot G\left(\frac{x_i - l}{k} \right) \right)^2, \tag{6}$$

where l ranges over all intensities, σ_l is the standard deviation for intensity l, $G(\cdot)$ is the density of the unit-normal distribution, and k is the width of the Gaussian kernel (we used $k = \frac{60}{255}$). This smoothed variance function is continuous and differentiable, which is required for the denoising algorithm in Section 4.



(a) Original frame.

(b) Inhomogeneous beta noise applied to mean image.

(c) Gaussian noise added to mean image.

Figure 5: **Comparison between actual and synthetic noise.**

From our measurements we find that for this particular sequence the gray values for the ISP and SSP are $l_{\text{ISP}} = 0$ and $l_{\text{SSP}} = \frac{243}{255}$. As predicted by the generic noise variance function derived above, we empirically find that $\sigma(0) = \sigma(1) = 0$, i. e., there is no noise at the ISP or the SSP. Furthermore, near the ISP and SSP the noise is only moderate, but there is significantly more noise in the mid tones. On the other hand we find that the empirical noise variance function is somewhat skewed, which is in contrast to the symmetric generic variance function shown in Figure 4(a).

In order to determine if this is a realistic model of the image noise, we applied synthetic noise to the noise-free mean image by sampling from the conditional beta distribution (see Fig. 5(b)), which allows the comparison with an original frame shown in Figure 5(a). We can see that the learned, inhomogeneous beta model produces more realistic looking noise when compared to a simple Gaussian noise model (cf. Fig. 5(c)), particularly in bright or dark regions.

3 Image Prior

In our approach, we use the Fields-of-Experts model (FoE) for modeling the image prior, which was recently proposed by Roth and Black [2005a]. The FoE models the prior probability of natural images using a Markov random field model of high-order using cliques over spatially extended neighborhoods of pixels, for example 5×5 pixels. The clique potentials are modeled as a Product of Experts [Teh et al., 2003], in which the experts take the form of a Student-t distribution. Each expert models the response of a linear filter of the same size as the clique to the pixels of the clique. The overall model takes the following form:

$$p(\mathbf{x}) = \frac{1}{Z} \prod_{k=1}^{M} \prod_{i=1}^{N} \left(1 + \frac{1}{2}(\mathbf{J}_i^{\mathrm{T}} \mathbf{x}_{(k)})^2\right)^{-\alpha_i}. \tag{7}$$

The vector $\mathbf{x}_{(k)} \in \mathbb{R}^{n \cdot n}$ is the vector of the pixel intensities in the $n \times n$ patch corresponding to clique k, and $\mathbf{J}_i$ is a linear filter that is applied to the patch. The parameters α_i control the "weight" of each expert, and Z is an (unkown) normalization constant. In the FoE framework, the set of parameters $\{\mathbf{J}_i \in \mathbb{R}^{n \cdot n}, \alpha_i \in \mathbb{R}\}$ is learned from a large, generic set of images using contrastive divergence [Teh et al., 2003]. Later on, we will require the gradient of the log-prior with respect to the image, which in case of the FoE model can be written as

$$\nabla_{\mathbf{x}} \log p(\mathbf{x}) = \sum_{i=1}^{N} \mathbf{J}_{-}^{(i)} * \boldsymbol{\psi}'(\mathbf{J}^{(i)} * \mathbf{x};\, \alpha_i), \tag{8}$$

where $\mathbf{J}^{(i)}$ is a convolution filter corresponding to $\mathbf{J}_i$, the filter $\mathbf{J}_{-}^{(i)}$ is obtained by mirroring $\mathbf{J}^{(i)}$ around its center pixel, and

$$\psi'(y;\, \alpha) = -\frac{\alpha y}{1 + \frac{1}{2} y^2}. \tag{9}$$

More details on both the model and the learning procedure are given in [Roth and Black, 2005a]. In this work we treat the model largly as a black-box and use the model parameters available at [Roth and Black, 2005b].

4 Film Denoising

To remove grain from archival films we follow a Bayesian approach and combine the learned noise model with the Fields-of-Experts prior as just introduced. To that end we write the posterior probability of the true image given the noisy observation as

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x}) \cdot p(\mathbf{x}), \tag{10}$$

where $p(\mathbf{y}|\mathbf{x})$ is the learned likelihood model of film grain, and $p(\mathbf{x})$ is the FoE prior. Denoising (and here more specifically de-graining) proceeds by approximately maximizing the posterior using gradient ascent. If $\mathbf{x}^{(t)}$ denotes the image at iteration t, the gradient ascent updates the image according to

$$\mathbf{x}^{(t+1)} \leftarrow \mathbf{x}^{(t)} + \tau \cdot \left(\lambda \cdot \nabla_{\mathbf{x}^{(t)}} \log p(\mathbf{y}|\mathbf{x}^{(t)}) + (1 - \lambda) \cdot \nabla_{\mathbf{x}^{(t)}} \log p(\mathbf{x}^{(t)}) \right), \tag{11}$$

where τ is the step size of the descent and $\lambda \in (0, 1)$ is a weight term that specifies the importance of the data term with respect to the prior term. The derivative of the log-prior is given by Eq. (8). The derivative of the log-likelihood is calculated as outlined in Appendix A. We determined the λ parameter by maximizing the peak signal-to-noise ratio improvement ($\text{PSNR} = 20 \log_{10} 255/\sigma_e$, where σ_e is the standard deviation of the error) on a set of 5 training images with artificial noise from the beta noise model with the parameters learned from the sequence discussed above; we found $\lambda = 0.21$ as the optimal value. τ was selected from the range (0.1,1).

Next, we applied the de-graining algorithm based on the learned beta noise model to 150 frames from the sequence. Figure 6 shows some of the results. Qualitatively we find that the noise from the film grain has been suppressed well in all of the frames. In order to also make a quantitative statement, we applied beta noise to the mean image $\mathbf{m}$ using the learned variance function, and de-grained the artificial image. When comparing the result to that obtained with a standard homogeneous Gaussian likelihood term (with the appropriate standard deviation or equivalently an appropriate weight λ), we find that the beta model increases the PSNR from 36.25dB to 36.42dB.

Moreover, we quantitatively evaluated the performance on the image sequence from Figure 6, which is corrupted by real film grain. We used the mean image described in Section 2.3 as pseudo ground-truth, and measured the PSNR only in areas that were not moving. Averaged over 20

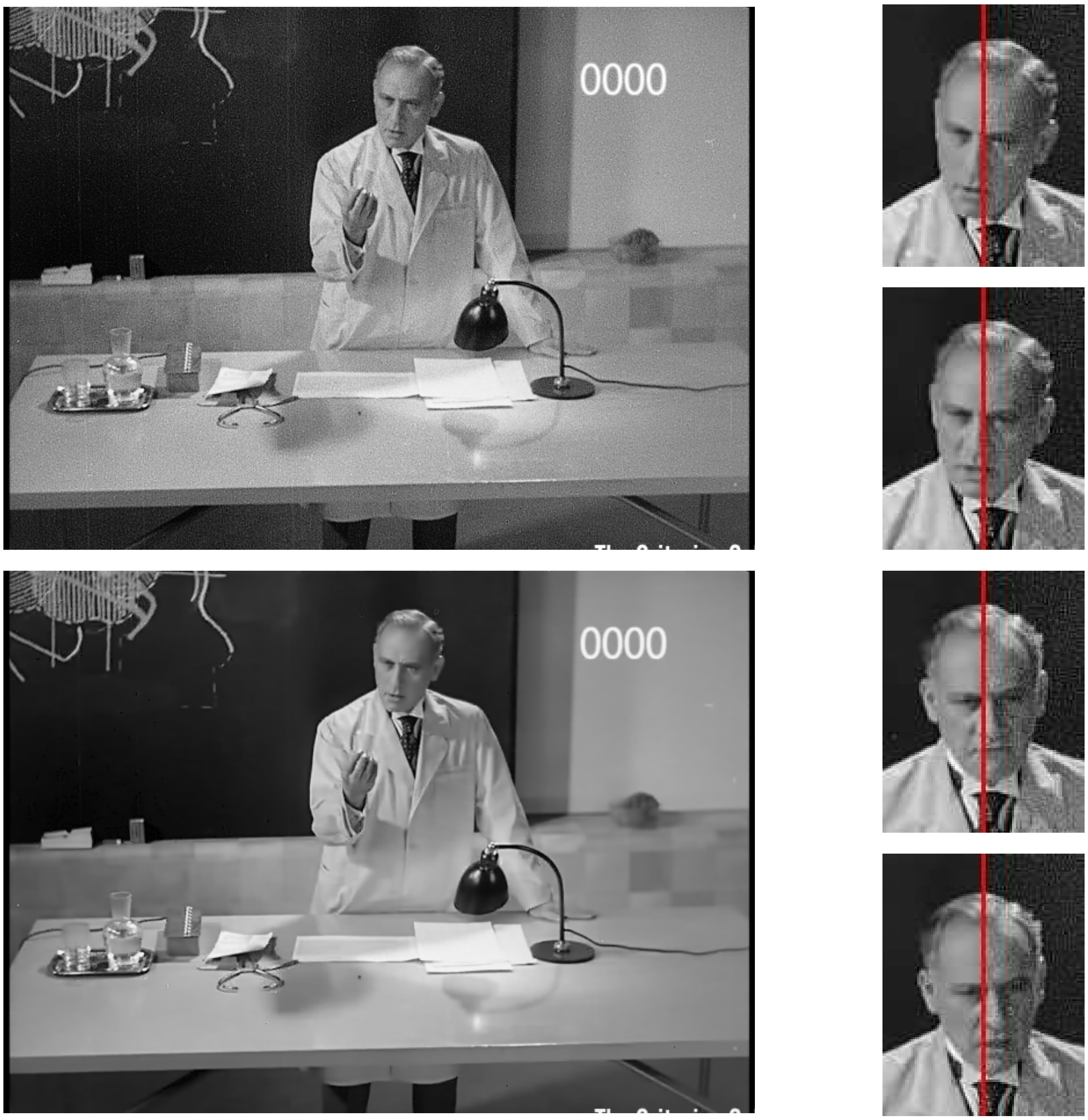


Figure 6: **Denoising using learned likelihood model.** *(left top)* Original sequence. *(left bottom)* De-grained sequence. *(right)* Detail results from various frames shown in "split screen" format with the left side being the restored version of the right side.

frames, we find using the learned inhomogeneous beta likelihood increases the PSNR from 34.44dB to 35.87dB when compared to a homogeneous Gaussian model. One reason why the beta noise model only leads to moderate PSNR improvements is that the data term only strongly deviates from a Gaussian data term near the ISP or the SSP. However, given that an inhomogeneous beta data term is only slightly more difficult to implement, we feel that the performance improvement outweighs the effort.

5 Summary

In this paper we studied the task of modeling application specific likelihood models in the concrete case of modeling image noise from film grain. Principled likelihood models are important in essentially all low-level vision applications, but have often been left aside. In the case of noise from film grain, we showed that the noise is non-Gaussian contrary to what is often assumed for generic image denoising techniques, and we also showed that the noise variance strongly depends on the intensity of the true underlying pixel. The latter fact had long been known in parts of the image processing literature, but most current denoising techniques do not model or take advantage of this fact. We proposed a noise model specific to photographic processes based on a spatially varying (i. e., inhomogeneous) beta distribution, whose parameters are learned from data. We showed that this learned likelihood model can be combined with the Fields-of-Experts prior and demonstrated the use in denoising individual film frames. The proposed model has potential applications in other fields as well, such as artificial noise synthesis, which is often needed in film post-processing.

Acknowledgments

We would like to thank Jay Cassidy, Kevin Manbeck, and Don McClure for interesting discussions. This material is based upon work supported by the National Science Foundation under Grant No. IIS-0535075, by Intel Research, and by NIH-NINDS R01 NS 50967-01 as part of the NSF/NIH Collaborative Research in Computational Neuroscience Program.

A Technical Details

A.0.1 Writing α and β in terms of mean and variance.

The mean of a beta distribution [Gelman et al., 2004, § A.1] is given in terms of the distribution parameters α and β as

$$\mu = \frac{\alpha}{\alpha + \beta}. \tag{12}$$

It immediately follows that

$$\beta = \alpha \frac{1 - \mu}{\mu} \qquad \text{and} \qquad \alpha + \beta = \frac{\alpha}{\mu}. \tag{13}$$

The variance of a beta distribution depends on the parameters as

$$\sigma^2 = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)}, \tag{14}$$

which we can simplify as follows:

$$\sigma^2 = \mu \frac{\beta}{(\alpha+\beta)(\alpha+\beta+1)} = \frac{\mu^2\beta}{\alpha(\frac{\alpha}{\mu}+1)} = \frac{\mu^3\beta}{\alpha(\alpha+\mu)} = \frac{\mu^3\alpha\frac{1-\mu}{\mu}}{\alpha(\alpha+\mu)} = \frac{\mu^2(1-\mu)}{\alpha+\mu}. \tag{15}$$

Solving for α, we obtain that

$$\alpha = \mu\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right), \tag{16}$$

and with above rule that

$$\beta = (1-\mu)\left(\frac{\mu(1-\mu)}{\sigma^2} - 1\right). \tag{17}$$

Assuming that $\mu = x_i$ and that the variance depends on x_i, we immediately obtain Eqs. (4) and (5).

A.0.2 Derivative of the log-likelihood

The log-likelihood of the developed model of film grain (cf. Eq. (3)) is written as follows:

$$\begin{aligned}\log p(y|x) =& \log\Gamma(\alpha(x)+\beta(x)) - \log\Gamma(\alpha(x)) - \log\Gamma(\beta(x)) \\ &+ (\alpha(x)-1)\cdot\log(y) + (\beta(x)-1)\cdot\log(1-y).\end{aligned} \tag{18}$$

By applying the chain rule, we obtain the derivative w. r. t. x as

$$\begin{aligned}\frac{\partial}{\partial x}\log p(y|x) =& \frac{\Gamma'(\alpha(x)+\beta(x))}{\Gamma(\alpha(x)+\beta(x))}\cdot(\alpha'(x)+\beta'(x)) - \frac{\Gamma'(\alpha(x))}{\Gamma(\alpha(x))}\cdot\alpha'(x) - \frac{\Gamma'(\beta(x))}{\Gamma(\beta(x))}\cdot\beta'(x) \\ &+ \alpha'(x)\cdot\log(y) + \beta'(x)\cdot\log(1-y),\end{aligned} \tag{19}$$

where $\frac{\Gamma'(\cdot)}{\Gamma(\cdot)}$ is the digamma function [Abramowitz and Stegun, 1964, §6.3]. If we rewrite

$$\alpha(x) = \frac{x^2 - x^3}{\sigma^2(x)} - x \qquad \text{and} \qquad \beta(x) = \frac{x^3 - 2x^2 + x}{\sigma^2(x)} + (x-1), \tag{20}$$

we obtain the following derivative expressions by applying the chain rule and simplifying:

$$\alpha'(x) = \frac{x(2-3x)}{\sigma^2(x)} - \frac{2x^2(1-x)}{\sigma^3(x)}\cdot\sigma'(x) - 1 \tag{21}$$

$$\beta'(x) = \frac{(1-x)(1-3x)}{\sigma^2(x)} - \frac{2x(1-x)^2}{\sigma^3(x)}\cdot\sigma'(x) + 1. \tag{22}$$

The derivative of the noise standard deviation (cf. Eq. (6)) is given as

$$\sigma'(x) = \sum_l h_l \cdot G'\left(\frac{x-l}{k}\right), \tag{23}$$

where $G'(\cdot)$ is the derivative of the Gaussian kernel. We should note once again that it is important to consider the normalization term in this derivation, because the normalization factor depends on the true image $\mathbf{x}$ that we want to solve for.

References

Milton Abramowitz and Irene A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* Dover, New York, New York, 1964.

J. C. Dainty and R. Shaw. *Image Science.* Academic Press, 1974.

Andrew Gelman, John B. Carlin, Hal S. Stern, and Donald B. Rubin. *Bayesian Data Analysis.* Chapman & Hall/CRC, 2004.

P. A. Kelly, H. Derin, and K. D. Hartt. Adaptive segmentation of speckled images using a hierarchical random field model. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 36 (10):1628–1641, October 1988. doi: 10.1109/29.7551.

Anil C. Kokaram and Simon J. Godsill. MCMC for joint noise reduction and missing data treatment in degraded video. *IEEE Trans. Signal Process.*, 50(2):189–205, February 2002. doi: 10.1109/78.978375.

Thomas C. M. Lee. Segmenting images corrupted by correlated noise. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(5):481–492, May 1998. doi: 10.1109/34.682178.

Ce Liu, William T. Freeman, Richard Szeliski, and Sing Bing Kang. Noise estimation from a single image. In *IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR)*, volume 1, pages 901–908, June 2006. doi: 10.1109/CVPR.2006.207.

Teodor Mihai Moldovan, Stefan Roth, and Michael J. Black. Denoising archival films using a learned Bayesian model. In *IEEE Int. Conf. on Image Proc. (ICIP)*, pages 2641–2644, October 2006. doi: 10.1109/ICIP.2006.313052.

F. Naderi and A. A. Sawchuk. Estimation of images degraded by film-grain noise. *Applied Optics*, 17(8):1228–1237, April 1978. URL `http://www.opticsinfobase.org/abstract.cfm?URI=ao-17-8-1228`.

Javier Portilla. Full blind denoising through noise covariance estimation using Gaussian scale mixtures in the wavelet domain. In *IEEE Int. Conf. on Image Proc. (ICIP)*, volume 3, pages 1217–1220, October 2004. doi: 10.1109/ICIP.2004.1419524.

Stefan Roth and Michael J. Black. Fields of experts: A framework for learning image priors. In *IEEE Conf. on Comp. Vis. and Pat. Recog. (CVPR)*, volume 2, pages 860–867, June 2005a. doi: 10.1109/CVPR.2005.160.

Stefan Roth and Michael J. Black. On the spatial statistics of optical flow. *Int. J. Comput. Vision*, 74(1):33–50, August 2007. doi: 10.1007/s11263-006-0016-x.

Stefan Roth and Michael J. Black. Learned FoE parameters. `http://www.cs.brown.edu/people/roth/foe/`, 2005b.

S. Ibrahim Sadhar and A. N. Rahagopalan. Image estimation in film-grain noise. *Signal Processing Letters*, 12(3):238–241, March 2005. doi: 10.1109/LSP.2004.840850.

Peter Schallauer and Roland Mörzinger. Rapid and reliable detection of film grain noise. In *IEEE Int. Conf. on Image Proc. (ICIP)*, pages 413–416, October 2006. doi: 10.1109/ICIP.2006.312481.

A. S. Tavildar, H. M. Gupta, and S. N. Gupta. Maximum a posteriori estimation in presence of film grain noise. *Signal Processing*, 8(3):363–368, June 1985. doi: 10.1016/0165-1684(85)90112-4.

Yeh Whye Teh, Max Welling, Simon Osindero, and Geoffrey E. Hinton. Energy-based models for sparse overcomplete representations. *J. Mach. Learn. Res.*, 4:1235–1260, December 2003. URL `http://www.jmlr.org/papers/volume4/teh03a/teh03a.pdf`.

Jun Tsuzurugi and Masato Okada. Statistical mechanics of the Bayesian image restoration under spatially correlated noise. *Physical Review E*, 66(6):066704, December 2002. doi: 10.1103/PhysRevE.66.066704.