DENOISING ARCHIVAL FILMS USING A LEARNED BAYESIAN MODEL

Teodor Mihai Moldovan, Stefan Roth, Michael J. Black

Brown University, Department of Computer Science, Providence, RI, USA
teodor_moldovan@brown.edu; roth@cs.brown.edu; black@cs.brown.edu

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. Here we consider the problem of removing film grain and focus on the development of a fully 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. Our 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 also model the prior probability of natural images using the recently proposed *Field-of-Experts* (FoE) model [1] that captures the spatial statistics using a high-order Markov random field (MRF) and is learned from a database of natural images. On images with artificial Gaussian noise, this FoE model has previously been shown to give state-of-the-art denoising results. Here the 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 material is based upon work supported by the National Science Foundation under Grant No IIS-0535075.

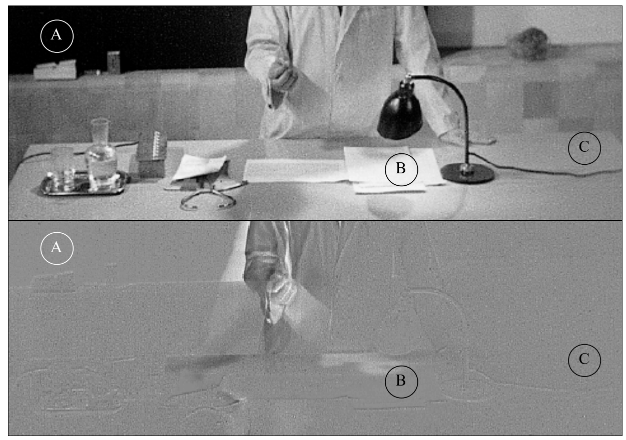

Fig. 1. Example frame (partially visible) from an archival film: *(top)* Actual movie frame. *(bottom)* Difference image revealing the grain (see text).

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 [2, 3, 4]. Previous methods, however, have used simplified models that assume Gaussian noise where the variance is a simple function of luminance. Typical denoising techniques exploit a Gaussian noise model to perform Wiener filtering [2] or MAP estimation [3]. Like our method many of these previous approaches focus on denoising individual frames rather than sequences [5]. 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.

Physical motivation

Photographic film [6] 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. Note that the absolute maximum and minimum transparency will mean that the noise process cannot be Gaussian as the support of the distribution is finite.

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. For NTSC resolution video and standard film, the discretized pixel size will always be significantly larger than the physical structure of the film grain resulting in spatially uncorrelated noise. Of course, film grain is also 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).

Noise depends on luminance

From a simple 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 B (ISP regions) and perfectly transparent regions 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. By adding a scale parameter to the above result, we obtain the generic noise variance

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

We thus find that the variance is a function of the average luminance over a patch. In reality, film has complicated physical properties [6] and the process of digitization is also not an ideal sampling process. Hence, we expect the noise variance function to deviate somewhat from this generic form.

2. DATA EXPLORATION

Consider the example from a black-and-white film in Figure 1 (top). 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 (bottom). We use M to denote the mean image, F^t to denote the t-th frame, and D^t to denote the difference between the mean and frame t.

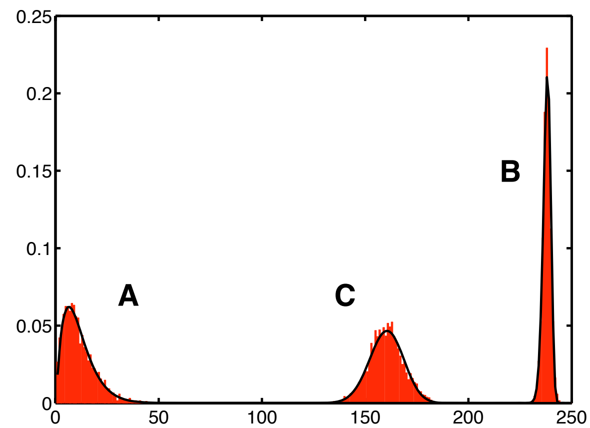

Fig. 2. Pixel value histograms for regions A, B, C represented in Figure 1.

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 2 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.

3. THE BETA NOISE MODEL

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 noise due to the large relative size of the image pixels to the grain. Let Y^0 be the observed image and X^0 be the true image. For convenience of notation, we rescale and shift the image intensities to lie on the interval $[0, 1]$, and denote those images X and Y. Using the assumption of spatial independence we can write the likelihood as

$$p(Y^0|X^0) = p(Y|X) = \prod_{ij} p(y_{ij}|x_{ij}). \tag{2}$$

We claim that the pixels y_{ij} are well modeled by an inho-

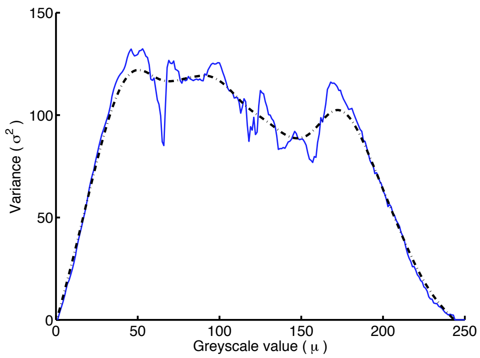

Fig. 3. The noise variance function trained on 50 frames of the sequence, corrected for global luminance changes. The dotted line represents a smoothed version (using Gaussian kernels).

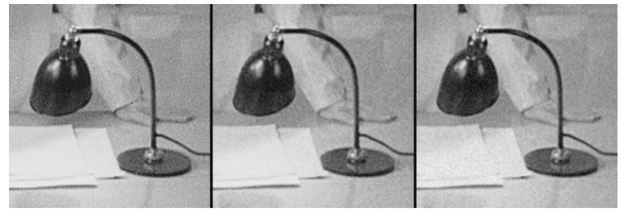

Fig. 4. Comparison between actual and synthetic noise: *(left)* Original frame. *(middle)* Inhomogeneous beta noise added to mean image. *(right)* Gaussian noise added to mean image.

mogeneous beta distribution

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

where Γ is the gamma function and $\alpha(x_{ij})$, $\beta(x_{ij})$ are two parameters of the beta distribution that depend on the grayvalue of the true pixel x_{ij}. The beta distribution is preferable over an inhomogeneous Gaussian distribution (used e. g. in [4]), because the beta distribution has a finite support and models the skew that we empirically observe. We should also note that this likelihood term does not assume additive noise.

To complete the model we assume that the mean value of y_{ij} is equal to x_{ij} (the true intensity value) and that the variance of y_{ij} is a function of x_{ij}. Using the relation between the parameters α, β and the mean and variance, we can express α and β as

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

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

Our model has a functional parameter $\sigma^2(x_{ij})$ that we call the noise variance function. We expect that the real noise variance function will look similar to the generic one derived above.

Fitting the model to data.

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 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 D^t and compute the variance of the noise for each grey value. The resulting variance function is plotted in Figure 3 together with a smoothed version that is obtained by putting a small Gaussian kernel on each bin of the discrete noise variance. The smoothed variance function is continuous and differentiable, which is required for the denoising algorithm in Section 4.

From our measurements we find that for this particular sequence the grayvalues for the ISP and SSP are $G_{\text{ISP}} = 0$ and $G_{\text{SSP}} = 243$. 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 midtones. We also find that the empirical noise variance function is somewhat skewed, which is in contrast to the symmetric generic variance function.

In order to determine if this is a realistic model of the image noise, we added noise to the noise-free mean image, which allows the comparison with an original frame in Figure 4. We can see that the learned, inhomogeneous beta model produces more realistic looking noise when compared to a simple Gaussian noise model, particularly in bright or dark regions.

4. FILM DENOISING

To remove grain from archival films we follow a Bayesian approach and combine the noise model developed above with a recent prior model of images called Fields of Experts (FoE) [1]. The FoE is a high-order Markov random field model that captures rich structural properties of natural images using large cliques in the random field. The model is trained on a database of generic natural images. For brevity we will omit any detailed discussion of this image model and refer the reader to [1]; the main point to note here is that the model gives us an unnormalized prior probability $p(X)$ of the true image X. Combining this with the likelihood of the noisy image given the true image $p(Y|X)$, we can write the posterior probability of the true image as

$$p(X|Y) \propto p(Y|X) \cdot p(X). \tag{6}$$

De-graining proceeds by approximately maximizing the posterior using a gradient ascent procedure related to nonlinear

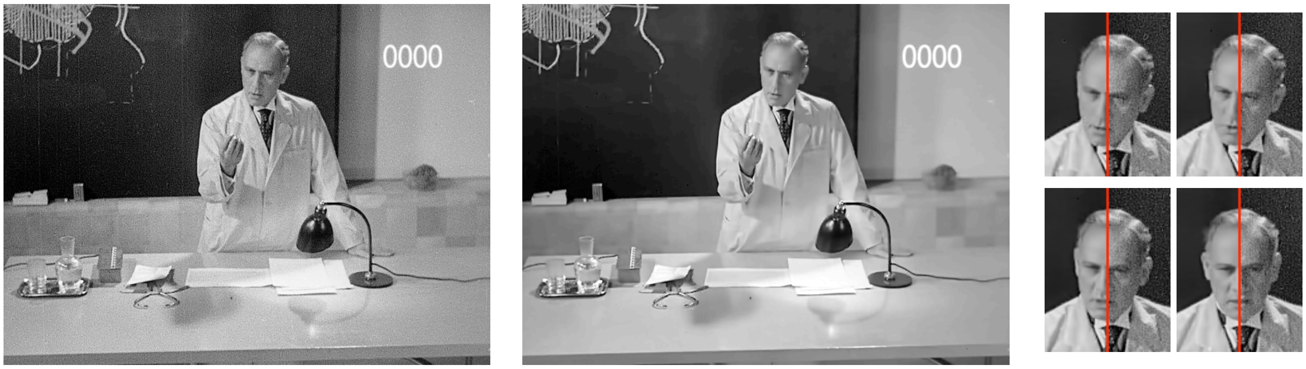

Fig. 5. Denoising using learned model: *(left)* Original sequence. *(middle)* De-grained sequence. *(right)* Detail results from various frames shown in "split screen" format with the left side being the restored version of the right side.

diffusion techniques (see [1]). If $X^{(t)}$ denotes the image at iteration t, the gradient ascent updates the image according to

$$X^{(t+1)} \leftarrow X^{(t)} + \tau \cdot \left(\lambda \cdot \frac{\partial}{\partial X^{(t)}} \log p(Y | X^{(t)}) + (1-\lambda) \cdot \frac{\partial}{\partial X^{(t)}} \log p(X^{(t)}) \right), \tag{7}$$

where τ is the step size and $\lambda \in (0, 1)$ is a weight term that specifies the importance of the data term with respect to the prior term. We set the λ parameter such that we obtain the maximum peak signal-to-noise ratio (PSNR) improvement on a set of 5 training images; we concluded that the optimum value is $\lambda = .21$. For τ we used values in the range (0.1,1). The derivative expressions in this de-graining procedure can be expressed in closed form but are omitted here for brevity.

We applied the de-graining algorithm based on the learned beta noise model to 150 frames from the sequence discussed above. Figure 5 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 added beta noise to the mean image 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, 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 5, which is corrupted by real film grain. We used the mean image described in Section 2 as pseudo ground-truth, and measured the PSNR only in areas that were not moving. Averaged over 20 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. CONCLUSIONS

To the best of our knowledge, the presented inhomogeneous beta noise model is the first realistic learned model of the noise specific to photographic processes. We believe that the proposed model has broad applications in many fields, including artificial noise synthesis, as well as image and film de-graining.

Acknowledgments: We thank Stu Geman, Jay Cassidy, Kevin Manbeck, and John Hughes for their suggestions regarding image denoising. Images used here are from the film "Das Testament des Dr. Mabuse" (1933, Courtesy Criterion Collection).

6. REFERENCES

[1] S. Roth and M. J. Black, "Fields of experts: A framework for learning image priors," in *CVPR 2005*, vol. 2, pp. 860–867.

[2] F. Naderi and A. A. Sawchuk, "Estimation of images degraded by film-grain noise," *Applied Optics*, vol. 17, no. 8, pp. 1228–1237, 1978.

[3] A. S. Tavildar, H. M. Gupta, and S. N. Gupta, "Maximum a posteriori estimation in presence of film grain noise," *Signal Processing*, vol. 8, pp. 363–368, 1985.

[4] S. I. Sadhar and A. N. Rahagopalan, "Image estimation in film-grain noise," *Signal Processing Letters*, vol. 12, no. 3, pp. 238–241, 2005.

[5] A. C. Kokaram and S. J. Godsill, "MCMC for joint noise reduction and missing data treatment in degraded video," *IEEE T. Image Proc.*, vol. 50, no. 2, pp. 189–205, 2002.

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