A Frequency Domain Performance Analysis of Horn and Schunck's Optical Flow Algorithm for Deformable Motion

Thomas S. Denney, Jr. and Jerry L. Prince

Abstract— **A frequency domain performance analysis of Horn and Schunck's optical flow (HSOF) algorithm for estimation of deformable motion is presented. Noise sources in the algorithm are modeled using the discrete Fourier transform of the brightness pattern. This noise model along with the estimation error covariance function derived in previous work is used to derive an expression for the expected performance of the optical flow estimate that is valid for an arbitrary discrete brightness pattern. Simulation results are presented that demonstrate the validity of our methods and show that HSOF is more accurate that the optical flow estimate of Anandan for certain low-frequency patterns.**

I. Introduction

An important application in both image processing and computer vision is estimating the motion of deforming biological tissues from image sequences obtained with modalities such as ultrasound, computed tomography, and magnetic resonance (MR) imaging. The motion estimation problem in these applications differs from those typically considered in the computer vision literature in three ways: 1) occlusion does not occur because the image is a 2-D slice of a 3-D object rather than a 2-D projection of a 3-D object; 2) the motion must often be estimated from only two image frames because the temporal sampling is often coarse; and 3) it is possible to modify the spatial brightness pattern of the deforming object itself—for example, by changing the acoustic frequency in ultrasound, using contrast agents in CT, and applying tags in MR [1].

In the case of two-frame motion estimation in the presence of noise, differential methods such as Horn and Schunck's optical flow algorithm (HSOF) [2] are thought to perform poorly relative to other methods because of the difficulty of approximating spatial and temporal derivatives of a discrete image sequence [3]. Region-matching techniques such as Anandan's optical flow algorithm (AOF) [4] are generally considered to perform better in this case because no derivatives are required [3]. Recent research, however, has shown that for deformable motion with no occlusion, the performance of HSOF is highly dependent on the brightness pattern of the object undergoing motion [5], [6], and in applications such as MR tagging [1] where the brightness pattern of the object can be controlled, use of the optimal brightness pattern can result in accurate HSOF estimates of motion. Denney and Prince proposed a brightness pattern optimization procedure in [6] for a parametric class of patterns that was based on the estimation error covariance function derived by modeling HSOF as a stochastic linear smoother. This procedure computed the optimal tradeoff between low-frequency patterns that yield good derivative approximations and high-frequency patterns that provide good information about the underlying velocity field.

In this correspondence, we extend the results of [6] by developing an HSOF performance measure that is valid for an arbitrary discrete pattern. Noise sources in the optical flow algorithm are modeled using the discrete Fourier transform (DFT) of the brightness pattern. This noise model along with the estimation error covariance function derived in [6] is used to derive an expression for the expected performance of the optical flow estimate that we call the frequency domain expected performance (FDEP). We use the FDEP to analyze the HSOF performance of 16 patterns. Our results show that the HSOF performance is primarily a function of the pattern bandwidth. We then demonstrate that for certain low bandwidth patterns, HSOF consistently outperforms AOF.

II. Derivation of the FDEP

We begin with the stochastic linear smoother formulation of HSOF presented in [6]. The HSOF estimation error covariance is shown in [6] to be Σ^{-1}, where Σ is a matrix obtained by finite-differencing the velocity estimate equation in [2] on an $N \times N$ lattice. The expected performance of the HSOF estimate is given by the average error covariance

$$p = \frac{\sigma_u^2}{2N^2}\mathrm{tr}[\Sigma^{-1}] \tag{1}$$

where the variance σ_u^2 is a parameter that models the smoothness of the true velocity field. The matrix Σ is a function of the brightness pattern gradient $\nabla\varphi$ and a measurement noise variance given by

$$\sigma_w^2(\varphi) = \frac{2}{\Delta t^2}\sigma_a^2 + \frac{\Delta t^2}{4}\sigma_v^2(\varphi) \tag{2}$$

where σ_a^2 is the variance of the additive noise in the image sequence, Δt is the temporal sampling period, and

$$\sigma_v^2(\varphi) = \mathcal{E}\{(v_{\max}^T H(\varphi) v_{\max})^2\}. \tag{3}$$

Here, $\mathcal{E}$ is the expectation operator, $H(\varphi)$ is the Hessian of the brightness pattern φ, and $v_{\max}$ is the maximum magnitude velocity vector in the true velocity field. An expectation is required because $H(\varphi)$ is modeled as a random matrix to account for random placement of φ relative to the image frame. The variance $\sigma_w^2(\varphi)$ models errors in computing φ_t from a pair of images with a two-point forward difference. A similar expression can be derived for any difference formula for which the truncation error is known (see [6] for details). If an analytical expression for φ is available, $\sigma_w^2(\varphi)$ can be computed in closed form (cf. [6]). Equations (1)–(3) form the basis for the pattern optimization results in [6].

In [6], p was computed for a class of brightness patterns related by a single parameter. In this correspondence, we derive a p for a broader class of brightness patterns by expanding φ in a Fourier series. We

Manuscript received November 4, 1993; revised December 5, 1994. This work was supported by Whitaker Foundation Biomedical Engineering Research Grant 91-0108 and National Science Foundation Presidential Faculty Fellow Award MIP-9350336. The associate editor coordinating the review of this paper and approving it for publication was Dr. Homer H. Chen.

T. S. Denney is with the Department of Electrical Engineering, Auburn University, Auburn, AL 36849 USA.

J. L. Prince is with the Department of Electrical and Computer Engineering, The Johns Hopkins University, Baltimore, MD 21218 USA.

IEEE Log Number 9413718.

assume that φ is twice differentiable, bandlimited, has finite energy, and is periodic $[\varphi(x+T, y+T) = \varphi(x,y)]$. In this case, φ can be expressed as the truncated Fourier series

$$\begin{aligned}\varphi(x,y) = a_{00} &+ \sum_{m=1}^{N/2} a_{m0}\cos\omega_m x + c_{m0}\sin\omega_m x \\ &+ \sum_{n=1}^{N/2} a_{0n}\cos\omega_n y + b_{0n}\sin\omega_n y \\ &+ \sum_{m,n=1}^{N/2} a_{mn}\cos\omega_m x\cos\omega_n y \\ &\quad + b_{mn}\cos\omega_m x\sin\omega_n y \\ &\quad + c_{mn}\sin\omega_m x\cos\omega_n y \\ &\quad + d_{mn}\sin\omega_m x\sin\omega_n y \end{aligned} \tag{4}$$

where $\omega_m = 2\pi m/T$ and $\omega_n = 2\pi n/T$. We now allow φ to have a random shift (q_x, q_y) relative to the image frame and a random orientation ϕ relative to the velocity field. If we assume that q_x and q_y are uniformly distributed on $[0,T)$ and ϕ is uniformly distributed on $[0, 2\pi)$, (3) becomes (see [7] for details)

$$\begin{aligned}\sigma_v^2(\varphi) &= \mathcal{E}_\phi \mathcal{E}_{q_x,q_y}\left\{\left\{\left[v_{\max}^T R^T(\phi)H(\varphi)R(\phi)v_{\max}\right]^2\right\}\right\} \\ &= \frac{3\pi^4\|v_{\max}\|^4}{T^4}\sum_{m=1}^{N/2}\left[a_{m0}^2 + c_{m0}^2 + a_{0m}^2 + b_{0m}^2\right]m^4 \\ &\quad + \frac{3\pi^4\|v_{\max}\|^4}{2T^4}\sum_{m,n=1}^{N/2}\left[a_{mn}^2 + b_{mn}^2 + c_{mn}^2 + d_{mn}^2\right] \\ &\quad \times [m^2+n^2]^2 \end{aligned} \tag{5}$$

where $H(\varphi)$ is the Hessian of $\varphi(x+q_x, y+q_y)$ and

$$R(\phi) = \begin{bmatrix}\cos\phi & -\sin\phi \\ \sin\phi & \cos\phi\end{bmatrix}.$$

If φ is available as a discrete image defined on an $N \times N$ lattice with uniform spacing h, we define $\varphi[m,n] = \varphi(mh, nh)$, where $0 \le mn \le N-1$. In this case, $\sigma_v^2(\varphi)$ can be expressed in terms of its discrete Fourier transform (DFT) coefficients as follows:

$$\begin{aligned}\sigma_v^2(\varphi) &= \frac{3\pi^4\|v_{\max}\|^4}{16T^4}\left[|H_{\frac{N}{2}0}|^2 + |H_{0\frac{N}{2}}|^2 + \frac{16}{3}|H_{\frac{N}{2}\frac{N}{2}}|^2\right] \\ &\quad + \frac{12\pi^4\|v_{\max}\|^4}{N^4T^4}\sum_{k=1}^{N/2-1}\left[|H_{k0}|^2 + |H_{0k}|^2\right]k^4 \\ &\quad + \frac{6\pi^4\|v_{\max}\|^4}{N^4T^4}\sum_{k,l\in\Gamma}\left[|H_{kl}|^2 + |H_{k(N-l)}|^2\right][k^2+l^2]^2 \end{aligned} \tag{6}$$

where $T = hN$, H_{kl} is the DFT of $\varphi[m,n]$, and Γ is the index set $\{kl \mid k,l = 1,\cdots,\frac{N}{2} \quad kl \ne \frac{N}{2}\frac{N}{2}\}$. Equations (2) and (6) relate the measurement noise variance to the discrete frequency components in the pattern $\varphi[m,n]$. Each frequency component makes an incremental contribution to $\sigma_w^2(\varphi)$ that depends on its magnitude and distance from the frequency origin. We formulate the FDEP by incorporating (2) and (6) into the average error covariance in (1). Our procedure for computing the FDEP is summarized in Fig. 1.

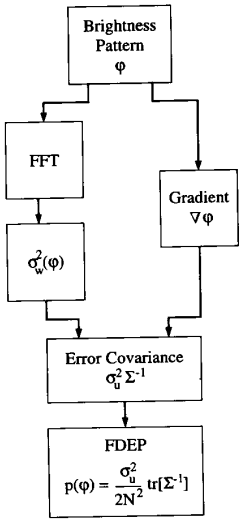

Fig. 1. Block diagram of FDEP algorithm.

III. Simulation Results

The FDEP for each of the sixteen brightness patterns in Fig. 2 was calculated, assuming that each pattern has gray levels in the range $[0,1]$, $\sigma_u^2 = 0.05$ s^{-1}, $\sigma_a^2 = 0.001$, $\Delta t = 1.0$, and $\|v_{\max}\| = 2.54$ cm/s. The resulting FDEP is shown beneath the corresponding pattern. It turns out that Patterns 1–16 are ordered according to their FDEP, with the best performance expected by Pattern 1. Interestingly, all of the selected simulated patterns are expected to perform better than all of the natural scenes. This phenomenon is explained below.

A deformed image was created for each pattern for the velocity field shown in Fig. 3, and zero-mean white Gaussian noise with variance 0.001 was added to each image pair. HSOF was then used to compute a velocity estimate. We used the spatial difference operators defined in [2], a two-point forward difference for the temporal derivative, and no presmoothing. The mean-square-error (MSE) between each estimated velocity $\hat{v}_{ij}$ and the true velocity v_{ij} was computed according to the formula

$$\text{MSE} = \frac{1}{2N^2}\sum_{i,j=1}^{N}\|v_{ij} - \hat{v}_{ij}\|^2. \tag{7}$$

The MSE is plotted versus FDEP in Fig. 4, where for each image we plot a circle at the location (x,y) = (FDEP, actual MSE) with the image number indicated near the circle. The dotted line represents a linear regression fit to the data points; the solid line is the line $y = x$. The linear regression calculation has a strong correlation (correlation coefficient ≥ 0.96) between the actual MSE and the FDEP. This means that the FDEP is a good predictor of the *relative* performance of different images—i.e., that we can reliably order their performance. Since most of the circles lie far away from the $y = x$ line, however, the FDEP is a relatively poor predictor of the actual MSE. We note, however, that the synthetic images tend to lie near the $y = x$ line while the natural scenes tend to lie far above it. It follows that the actual performance of the natural scenes is far worse than we predicted. Since the main difference between these synthetic and natural images is their bandwidth, we conclude that the FDEP is more accurate at predicting the quantitative value of the actual MSE in low bandwidth images than in higher bandwidth images.

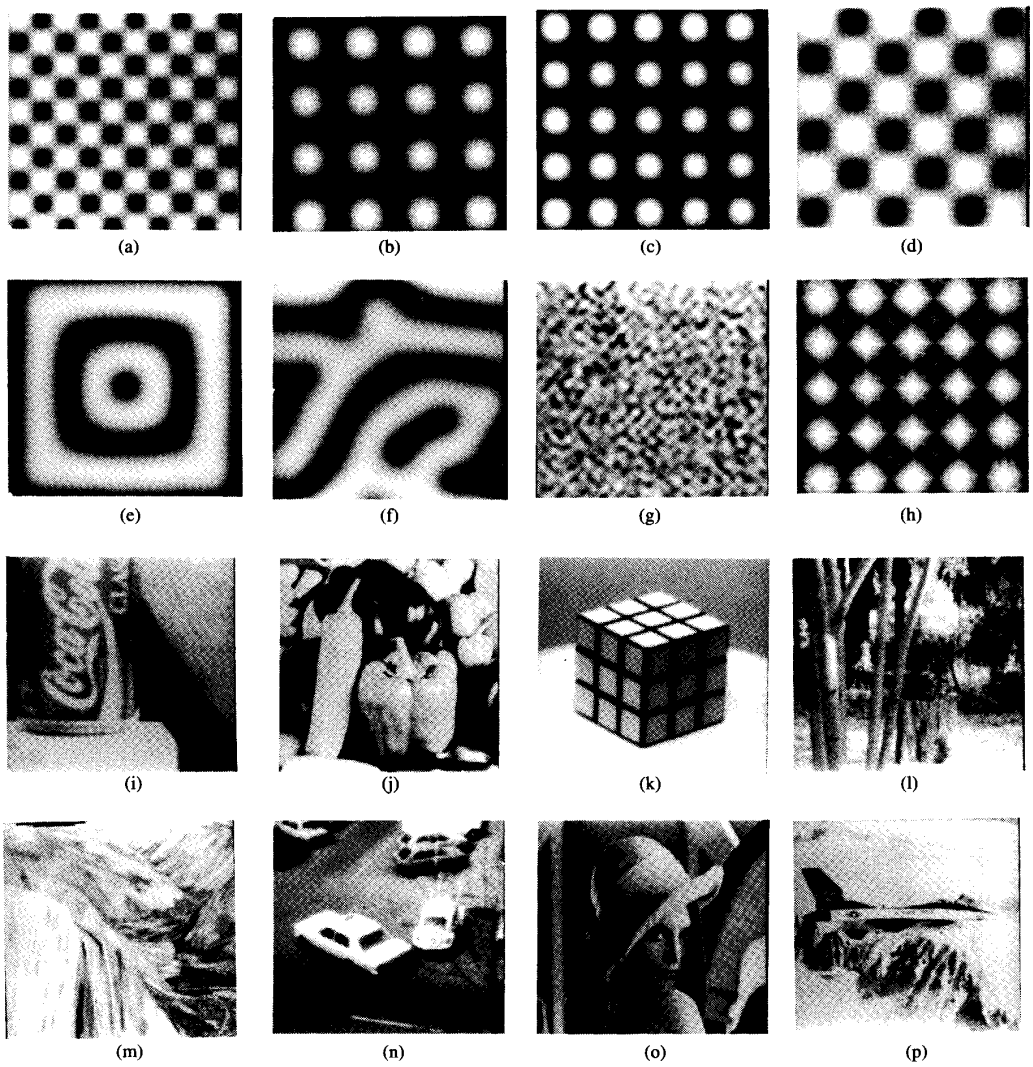

Fig. 2. Brightness patterns [(pattern number)—FDEP]. (a)—0.0561; (b)—0.0591; (c)—0.0591; (d)—0.0593; (e)—0.0679; (f)—0.0681; (g)—0.0711; (h)—0.0810; (i)—0.0841; (j)—0.0881; (k)—0.0882; (l)—0.0919; (m)—0.0935; (n)—0.0948; (o)—0.0961; (p)—0.0976.

To demonstrate the effect of high frequency components on both the expected performance of HSOF and the actual MSE, Pattern 10 (peppers) was filtered by a series of lowpass filters with different cutoff frequencies. For each cutoff frequency the FDEP of the filtered pattern was computed and an image pair was simulated for the velocity field in Fig. 3. A plot of the FDEP versus cutoff frequency is shown by the solid line in Fig. 5; the actual MSE is shown by the dotted line. Both the FDEP and the actual MSE improve as the high-frequency components are removed; the best performance is predicted at a cutoff frequency of about 0.06. This result suggests that for a given set of imaging parameters, there is a critical frequency f_c in which frequency components in the pattern above f_c degrade the performance of HSOF. The derivation of an expression for f_c is an open problem.

The above results and those of [6] show that the performance of HSOF is strongly influenced by the brightness pattern, and we now have a fairly good grasp of the theoretical reasons behind this phenomenon. But one might ask what is the performance of HSOF relative to other well-established (and mostly favored) optical flow methods? In our final simulation, we begin to explore this issue by comparing the performance of HSOF with the performance of the region-matching optical flow algorithm of Anandan (AOF) for the 16 image pairs described above. Our implementation of AOF is the one used by Barron *et al.* [3]. The MSE for both HSOF and AOF for each pattern is plotted in Fig. 6. Since the patterns are ordered by increasing expected performance, the MSE of the HSOF estimates tend to get worse with increasing pattern number; HSOF, however, performs much better than AOF on the low-frequency patterns while AOF performs better on the high frequency patterns.

IV. Discussion

We have described an approach to analyze the performance of Horn and Schunck's optical flow algorithm based on the frequency content of the brightness pattern. Our experimental results show that

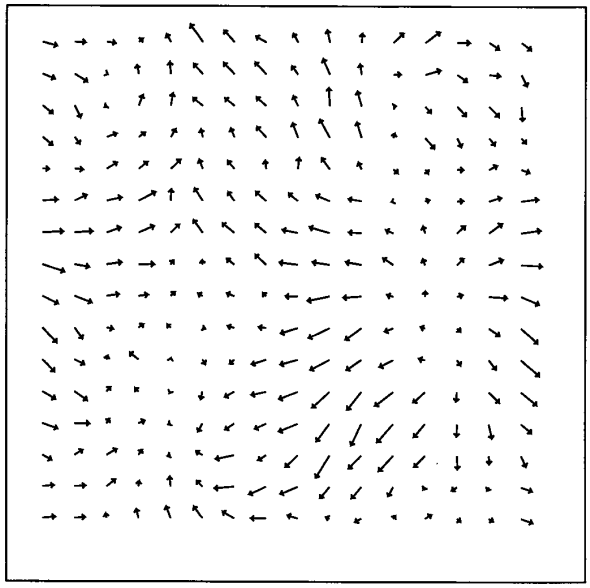

Fig. 3. Velocity field.

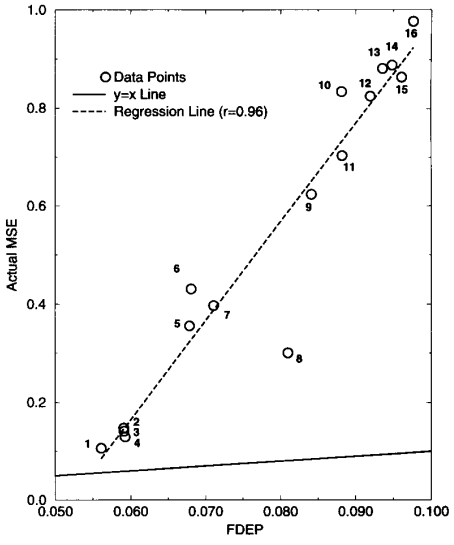

Fig. 4. Actual MSE versus FDEP.

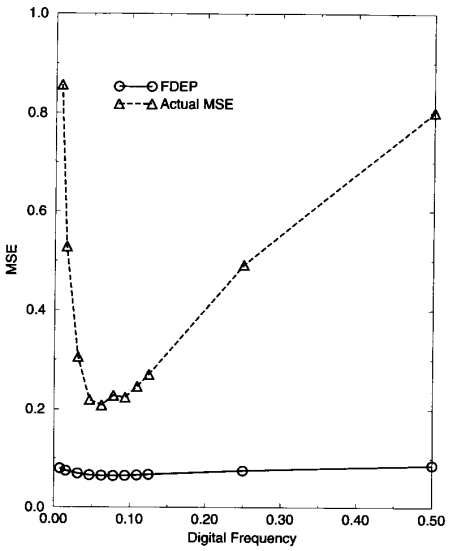

Fig. 5. HSOF performance versus cutoff frequency.

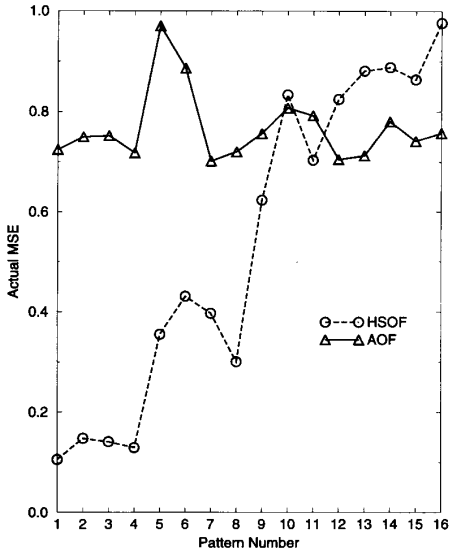

Fig. 6. HSOF and AOF performance.

our method can accurately predict the relative expected performance of HSOF for a given set of imaging and velocity field parameters but does not correctly predict the quantitative error, especially for images with high frequency content. Our results also suggest that for a given set of imaging and velocity parameters, images having bandwidth greater than or less than a critical frequency f_c tend to have poorer performance than those having bandwidth f_c. This idea may be useful in determining the bandwidth of spatio-temporal prefilters for conventional motion estimation applications. Developing an analytic formula for f_c is a subject of future research.

Finally, our results suggest that one reason why HSOF is generally considered to be a poor algorithm for motion estimation by the computer vision community is that it has been typically applied in situations where the brightness patterns have significant frequency components above f_c. Our simulation results suggest that when the brightness patterns have a lower frequency content, the quantitative accuracy of HSOF is at least competitive with—and in some cases much better than—region-matching algorithms. An interesting subject of future research is to derive optimal brightness patterns for other algorithms such as AOF and compare the performance of the different algorithms with each algorithm using its optimal pattern.

References

[1] E. A. Zerhouni, D. M. Parish, W. J. Rogers, A. Yangand, and E . P. Shapiro, "Human heart: Tagging with MR imaging—A method for noninvasive assessment of myocardial motion," *Radiology*, vol. 169, pp. 59–63, 1988.

[2] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artificial Intell.*, vol. 17, pp. 185–203, 1981.

[3] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of optical flow techniques," *Int. J. Comput. Vision*, vol. 12, no. 1, pp. 43–77, Feb. 1994.
[4] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *Int. J. Comput. Vision*, vol. 2, pp. 283–310, 1989.
[5] J. L. Prince and E. R. McVeigh, "Motion estimation from tagged MR image sequences," *IEEE Trans. Med. Imag.*, vol. 11, no. 2, pp. 238–249, June 1992.
[6] T. S. Denney, Jr. and J. L. Prince, "Optimal brightness functions for optical flow estimation of deformable motion," *IEEE Trans. Image Processing*, vol. 3, no. 2, pp. 178–191, Mar. 1994.
[7] T. S. Denney, Jr., "Stochastic estimation of deformable motion from magnetic resonance tagged cardiac images," Ph.D. thesis, The Johns Hopkins Univ., Baltimore, MD, 1994.

Linear-Quadratic Noise-Smoothing Filters for Quantum-Limited Images

Cheuk L. Chan, Aggelos K. Katsaggelos, and A. V. Sahakian

***Abstract*— In this correspondence, we consider the use of nonlinear estimators for the noise smoothing of images obtained under quantum-limited imaging conditions. A Volterra expansion is investigated from which a set of linear-quadratic filters is derived using higher order statistics. The filters are applicable for single frame and multiple frames of a single scene imaged under low-light levels.**

I. Introduction

Quantum noise is a signal-dependent, Poisson distributed noise source that can arise in a variety of image processing applications. For example, in medical imaging, the intentional reduction of radiation dosage in clinical exams for safer fluoroscopy procedures is accompanied by an ensuing degradation in image quality. This undesirable artifact is termed quantum mottle and is created by the depletion of necessary X-ray photons for imaging [1], [2]. In remote-sensing applications and low-light-level acquisitions such as astronomical imaging, the unavailability of light photons causes a similar loss in image quality [3].

Various approaches for postprocessing of images degraded by quantum noise have been proposed in the past. Because this type of noise is typically placed into the class of signal-dependent degradations, the majority of these techniques lie in the specification of a nonstationary model for the description of the noise [4], [5], resulting in a genre of nonstationary filters that operate independently at each pixel. This is intuitive with the notion of signal-dependent noise; however, these techniques also have a common feature in that they are linear. While it is well known that a nonlinear filter will generally perform better than a linear filter for non-Gaussian environments, the formulation of one is also difficult, if not impossible. Nevertheless, because the property of Poisson noise is such that it is known to have information in the higher order moments, it is useful to consider nonlinear estimators for the recovery of signal parameters.

In this correspondence, a class of nonlinear systems that can be described by a Volterra expansion [6] is investigated for the filtering of images containing quantum noise. Specifically, a set of linear-quadratic filters is derived using higher order statistics for suppressing the quantum noise without the need to know the exact probability density function (pdf) governing the noise samples.

Manuscript received August 18, 1993; revised October 28, 1994. This work was supported in part by a grant from Siemens. The associate editor coordinating the review of this paper and approving it for publication was Dr. Hsueh-Ming Hang.

C. L. Chan was with the Department of Electrical Engineering and Computer Science, Northwestern University. He is currently with PAR Government Systems Corporation, La Jolla, CA 92037-4146 USA.
A. K. Katsaggelos and A. V. Sahakian are with the Department of Electrical Engineering and Computer Science, Robert R. McCormick School of Engineering and Applied Science, Northwestern University, Evanston, IL 60208-3118 USA.

IEEE Log Number 9413721.

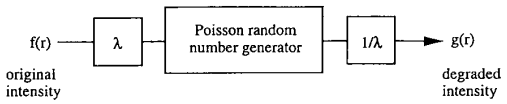

Fig. 1. Poisson noise simulator, after [8].

II. Background

A. Observation Model

In this section, a nonstationary additive observation model is proposed to be used for the nonlinear filtering of images degraded by quantum noise. Consider first the intensity function $\mathbf{f}$ of an object to be imaged. The observed counting process, $\mathbf{N}$, at an image intensifier's input in an imaging chain is characterized by the pdf $p(\mathbf{N} \mid \mathbf{f})$. For the quantum-limited scenario, which is the focus of our work, the number of photon counts in each pixel becomes statistically independent [2], [4], [7], i.e.

$$p(\mathbf{N} \mid \mathbf{f}) = \prod_{\mathbf{r} \in A} p(N(\mathbf{r}) \mid f(\mathbf{r}))$$

where the conditional density function of $N(\mathbf{r})$ is Poisson-distributed for a given realization of $f(\mathbf{r})$. The conditional density of the number of photons at each pixel is thus given by

$$p(N(\mathbf{r}) \mid f(\mathbf{r})) = \frac{(\lambda f(\mathbf{r}))^{N(\mathbf{r})} e^{-\lambda f(\mathbf{r})}}{N(\mathbf{r})!} \tag{1}$$

where the constant λ is a proportionality factor relating the displayed image intensity to the assumed number of photon counts present as shown in Fig. 1 [8]. The displayed image intensity, $g(\mathbf{r})$, where $g(\mathbf{r}) = \alpha N(\mathbf{r})$ and $\alpha = \frac{1}{\lambda}$, has the pdf [8], [9]

$$p_g(g(\mathbf{r}) \mid f(\mathbf{r})) = \lambda \frac{(\lambda f(\mathbf{r}))^{\lambda g(\mathbf{r})} e^{-\lambda f(\mathbf{r})}}{\lambda g(\mathbf{r})!}.$$

An equivalent and more useful model for minimum mean square error (MMSE) filtering is to express the noise as an additive signal-dependent term [4], [10]; that is

$$g(\mathbf{r}) = f(\mathbf{r}) + n_g(\mathbf{r}) \tag{2}$$

where the mean of the noise $n_g(\mathbf{r})$ is 0 and the variance is given by [2]

$$\sigma^2_{n_g}(\mathbf{r}) = \frac{1}{\lambda} E\{f(\mathbf{r})\}. \tag{3}$$