

A Nonlinear Filter for Film Restoration and Other Problems in Image Processing¹

STUART GEMAN AND DONALD E. McCLURE

Division of Applied Mathematics, Brown University, Providence, Rhode Island 02912

AND

DONALD GEMAN

Department of Mathematics and Statistics, University of Massachusetts, Amherst, Massachusetts 01003

Received December 11, 1990; accepted February 27, 1992

A filter is proposed for removing noise and other types of degradation. An application is explored to enhancement of frame sequences, motivated by the problem of film restoration for the movie industry. Experiments are performed on sequences from a degraded black and white copy of a recently released movie. Both temporal and spatial information are used in the restoration. Temporal information is obtained from the preceding unprocessed frame; spatial information enters by smoothing the current frame. The smoothing must be *nonlinear* in order to preserve boundaries. © 1992 Academic Press, Inc.

1. INTRODUCTION

This paper is about a nonlinear filter designed to remove noise and other types of degradation from single images or temporal sequences while at the same time preserving important discontinuities, such as those due to boundaries, creases, and shadows. Basically, the idea is to define the filtered image as the minimum value of a cost functional which consists of two terms, one incorporating *a priori* smoothness constraints and the other insuring fidelity to the data. A version of this filter was introduced in [9] for emission tomography. In that version, the smoothness constraints, which promote regions of *constant* intensity, are applied to the reconstructed image, and the data consist of arrays of photon counts. Here we suggest some extensions: the use of "higher order" constraints, which accommodate planar or quadratic surfaces and hence eliminate the "patchiness" associated with first order roughness penalties; and the inclu-

sion of a *temporal* component in the data term, motivated by the application to film restoration in the movie industry. In [7], these same higher order constraints are explored in a further application, to deblurring.

There is a growing literature on the processing of image sequences, motivated by such problems as reducing noise in TV frames for both enhancement and efficient coding [4, 5, 11, 13], sharpening ultrasonic and infrared imagery, and frame sequence reformatting, which arises in television standards conversion, and may involve spatial interpolation to alter the number of scan lines or temporal interpolation to vary the frame rate; see [12] and the references therein. The basic ingredient of most algorithms for restoration and enhancement is temporal filtering along motion trajectories, which are derived by estimating the displacement field; see, e.g., [4, 11, 12]. Motion-compensated temporal filtering may be preceded or followed by spatial (i.e., intraframe) filtering. Generally, the processing is nonlinear (e.g., median filter) and often recursive.

In contrast, our approach is nonmodular and optimization-based: the processed frame sequence is defined outright as the minimizer of an image functional rather than as the final result of a series of filters or other operations. (Actually, motion compensation is treated as a preprocessing step, but could be incorporated into the cost functional.) In particular, our filter is not separable, and spatial and temporal processing are performed at the same time: the brightness value at a pixel in the processed frame reflects a compromise between the values suggested by the spatial neighbors in the same frame and the values suggested by the motion-compensated "neighbors" in the temporal sequence. Similarly, in this framework, edge sharpening and noise suppression can be accomplished *simultaneously*, which may be advantageous since the former enhances noise and the latter degrades

¹ Supported by Army Research Office Contract DAAL03-86-K-0171 with the Center for Intelligent Control Systems, National Science Foundation Grant DMS-8813699, Office of Naval Research Contract N00014-88-K-0289, and the General Motors Research Laboratories.

edges. On the other hand, our model is certainly complex alongside such standbys as the median filter, and our approach is often more computationally intensive than those in [11, 12] for example. In our view, the trade-offs between efficiency and performance are as yet unclear, and likely to be application-dependent.

The filter is introduced in Section 2, and the application to film restoration is discussed in Section 3. Sections 4 and 5 explore ways to use the largely redundant information in successive movie frames for the restoration task. Results of experiments with some degraded black and white movie frames are presented in Section 6.

2. A NONLINEAR FILTER

The filtered image is defined as the collection of grey level values, $\mathbf{x} \doteq \{x_s\}_{s \in S}$, that minimizes a cost functional of the form

$$U(\mathbf{x}) + V(\mathbf{x}, \mathbf{y}), \quad (1)$$

where $\mathbf{y} \doteq \{y_t\}_{t \in T}$ is the *data* (e.g., one or more “raw,” or unprocessed, pictures), and S and T are the pixel lattices of the restored and raw picture(s), respectively. (In the movie application, for example, the raw data may include a previous frame as well as the current, to-be-processed, frame, in which case T is two copies of S .) The functional V penalizes restorations that are not faithful to the data (see Section 4), whereas U enforces anticipated regularities by assigning high values to undesirable configurations.

The functional in (1) can be thought of as arising either from a Bayesian model [2, 3, 8, 10] or, equivalently, from a likelihood model with regularization [14, 16]. From either viewpoint, $V(\mathbf{x}, \mathbf{y})$ is a negative log likelihood: The probability of the data, given \mathbf{x} , is proportional to $\exp\{-V(\mathbf{x}, \mathbf{y})\}$. From the Bayesian viewpoint, U then defines a prior distribution via the formula $\pi(\mathbf{x}) = Z^{-1} \exp\{-U(\mathbf{x})\}$, and the filtered image is the *maximum a posteriori* (MAP) estimator of \mathbf{x} given \mathbf{y} . Alternatively, from the likelihood viewpoint, we can think of U as a regularization functional, introduced mainly to smooth the (sometimes unacceptable) maximum likelihood estimator.

We want to remove noise, artifacts (for example, scratches), and other degrading effects, while simultaneously preserving boundaries and other authentic features of the original scene. Similar requirements arise in emission tomography and many other image processing problems. In [9] we introduced a class of cost functionals designed to promote sharp boundaries and smooth regions. These are of the form

$$U(\mathbf{x}) = \alpha \sum_{(s,t)} \phi(x_s - x_t),$$

where α is a classical “smoothing parameter” governing a trade-off between regularization and faithfulness to the data. Here, $\sum_{(s,t)}$ denotes summation over all nearest neighbor pairs in the pixel lattice S , and $\phi(u)$ is a symmetric function, nondecreasing on $[0, \infty)$. The minimization of (1) thereby favors “smooth” images; in fact, $U(\mathbf{x})$ is minimized when \mathbf{x} is constant. If $\phi(u)$ grows “slowly,” at least for large u , then the discontinuities associated with boundaries are not unduly penalized. In this regard, the *quadratic* function, $\phi(x_s - x_t) = (x_s - x_t)^2$, is not suitable. Instead, we have used functions of the form

$$\phi(u) = -\left(1 + \left|\frac{u}{\delta}\right|^\gamma\right)^{-1}. \quad (2)$$

Roughly speaking, when the exponent, γ , is small (say $\gamma \leq 1$), sharp boundaries are favored over gradual transitions, although larger values of γ permit more variation within otherwise homogeneous regions. In all of our experiments γ is set to one. The setting of α and δ is discussed in Section 6.

The cost functional

$$U(\mathbf{x}) = -\alpha \sum_{(s,t)} \left(1 + \left|\frac{x_s - x_t}{\delta}\right|^\gamma\right)^{-1} \quad (3)$$

favors regions of *constant grey level*. To the extent that grey level images of real scenes have homogeneous regions, these regions are better defined by constant *gradient*, or even constant *curvature*, then by constant grey level. That is, planar and quadric surfaces are better local descriptions in typical grey level images. This suggests that filters based upon (3) will introduce an artificial “patchiness” or “mottling,” which is exactly what we have observed in various restoration experiments.

The functional $U(\mathbf{x})$ in (3) penalizes large “derivatives.” Thinking of this as a “first order” model it is natural to generalize to “second order” and “third order” models by penalizing, respectively, large second and third derivatives, thereby allowing planar and quadric surfaces, *in addition to* flat ones. To guess what terms should be involved, let us think, for a moment, about a continuum formulation in which there is a grey level surface $x(\xi_1, \xi_2)$, defined on the (ξ_1, ξ_2) -plane. The *first order model* favors solutions $x(\xi_1, \xi_2) = c$, defined by the (necessary and sufficient) conditions

$$\frac{\partial}{\partial \xi_1} x(\xi_1, \xi_2) = 0, \quad \frac{\partial}{\partial \xi_2} x(\xi_1, \xi_2) = 0.$$

The discrete analog, on the (ξ_1, ξ_2) -integer lattice, is

$$\begin{aligned} x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2) &= 0 \\ x(\xi_1, \xi_2 + 1) - x(\xi_1, \xi_2) &= 0 \end{aligned}$$

and this is exactly what minimizes (3), if we declare the neighbors of a lattice site to be the four nearest sites, two vertical and two horizontal. In particular, on an $N \times N$ lattice, the first order model is

$$U(\mathbf{x}) = \alpha \sum_{\xi_1=1}^{N-1} \sum_{\xi_2=1}^N \phi(x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2)) + \alpha \sum_{\xi_1=1}^N \sum_{\xi_2=1}^{N-1} \phi(x(\xi_1, \xi_2 + 1) - x(\xi_1, \xi_2)). \quad (4)$$

We can make a similar analysis to derive the second order model, promoting planar surfaces $x(\xi_1, \xi_2) = \alpha\xi_1 + \beta\xi_2 + c$. The corresponding differential equations are

$$\begin{aligned} \frac{\partial^2}{\partial \xi_1^2} x(\xi_1, \xi_2) &= 0, & \frac{\partial^2}{\partial \xi_2^2} x(\xi_1, \xi_2) &= 0, \\ \frac{\partial}{\partial \xi_1} \frac{\partial}{\partial \xi_2} x(\xi_1, \xi_2) &= 0, & \frac{\partial}{\partial \xi_2} \frac{\partial}{\partial \xi_1} x(\xi_1, \xi_2) &= 0, \end{aligned} \quad (5)$$

although the last two are redundant. As with the first order model, each differential equation corresponds to a difference equation for lattice-based grey levels, and each of these contributes a term to the second order model:

$$\begin{aligned} U(\mathbf{x}) &= \alpha \sum_{\xi_1=1}^{N-2} \sum_{\xi_2=1}^N \phi(x(\xi_1 + 2, \xi_2) - 2x(\xi_1 + 1, \xi_2) + x(\xi_1, \xi_2)) \\ &+ \alpha \sum_{\xi_1=1}^N \sum_{\xi_2=1}^{N-2} \phi(x(\xi_1, \xi_2 + 2) - 2x(\xi_1, \xi_2 + 1) + x(\xi_1, \xi_2)) \\ &+ \alpha \sum_{\xi_1=1}^{N-1} \sum_{\xi_2=1}^{N-1} \phi(x(\xi_1 + 1, \xi_2 + 1) - x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2 + 1) + x(\xi_1, \xi_2)) \\ &+ \alpha \sum_{\xi_1=1}^{N-1} \sum_{\xi_2=1}^{N-1} \phi(x(\xi_1 + 1, \xi_2 + 1) - x(\xi_1, \xi_2 + 1) - x(\xi_1 + 1, \xi_2) + x(\xi_1, \xi_2)). \end{aligned} \quad (6)$$

(Because of the redundancy in (5), the last two terms are identical. It would perhaps have been more logical, although most likely of very little practical consequence, to include this term only once.)

Finally, quadric surfaces, $x(\xi_1, \xi_2) = a\xi_1^2 + b\xi_2^2 + c\xi_1\xi_2 + d\xi_1 + e\xi_2 + f$, correspond to the equations

$$\begin{aligned} \frac{\partial^3}{\partial \xi_1^3} x(\xi_1, \xi_2) &= 0, & \frac{\partial^3}{\partial \xi_2^3} x(\xi_1, \xi_2) &= 0, \\ \frac{\partial}{\partial \xi_2} \frac{\partial^2}{\partial \xi_1^2} x(\xi_1, \xi_2) &= 0, & \frac{\partial}{\partial \xi_1} \frac{\partial^2}{\partial \xi_2^2} x(\xi_1, \xi_2) &= 0, \end{aligned}$$

These lead to the third order functional

$$\begin{aligned} U(\mathbf{x}) &= \alpha \sum_{\xi_1=1}^{N-3} \sum_{\xi_2=1}^N \phi(x(\xi_1 + 3, \xi_2) - 3x(\xi_1 + 2, \xi_2) + 3x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2)) \\ &+ \alpha \sum_{\xi_1=1}^N \sum_{\xi_2=1}^{N-3} \phi(x(\xi_1, \xi_2 + 3) - 3x(\xi_1, \xi_2 + 2) + 3x(\xi_1, \xi_2 + 1) - x(\xi_1, \xi_2)) \\ &+ \alpha \sum_{\xi_1=1}^{N-2} \sum_{\xi_2=1}^{N-1} \phi(x(\xi_1 + 2, \xi_2 + 1) - 2x(\xi_1 + 1, \xi_2 + 1) + x(\xi_1, \xi_2 + 1) - x(\xi_1 + 2, \xi_2) + 2x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2)) \\ &+ \alpha \sum_{\xi_1=1}^{N-1} \sum_{\xi_2=1}^{N-2} \phi(x(\xi_1 + 1, \xi_2 + 2) - 2x(\xi_1 + 1, \xi_2 + 1) + x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2 + 2) + 2x(\xi_1, \xi_2 + 1) - x(\xi_1, \xi_2)). \end{aligned}$$

All of the models accommodate discontinuities, via the bounded growth of ϕ . For example, in the second order model, very large jumps in either intensity or gradient incur little more penalty than moderate jumps. In this way, boundaries are better preserved than with the quadratic and other conventional "stabilizers."

As we have already pointed out, restorations with the first order model (4) often have an unnatural patchy or mottled look. In this regard, it is our experience that generally better results are obtained with the second and third order models. In our experiments on movie frames, results with second and third order models were visually indistinguishable, despite the *conceptual* advantage of the latter, namely that it supports a richer world model. Moreover, the second order model is computationally less demanding.

3. THE FILM RESTORATION PROBLEM

Most movies are made with 35-mm film, projected at 24 frames per s. The frames are usually acetate-based, although many older films are nitrate-based. Nitrate film presents special difficulties, both because it decays rapidly and because it is highly flammable. Either type is susceptible to degradation, through scratches and the ac-

cumulation of foreign material, mostly dirt. (The difficulties in keeping film "clean" are legendary.) Furthermore, both types of film deteriorate by drying and flaking. The result is a variety of artifacts, of which some persist from frame to frame (e.g., vertical scratches) whereas others are principally confined to *single* frames (e.g., lint or burst noise). In addition, there are generic degrading effects, such as uncorrelated noise. In current practice, depending on the production values, some artifacts are removed by a rather painstaking process, which involves stepping through the frames at video workstations, using electronic paintbrushes and other "manual" devices.

On the other hand, the deterioration can be halted and movies can be *preserved* by the production of master video copies. Of course, defects already present in the film are inherited in the video, and in fact most defects are actually accentuated in the viewing of video in comparison to film.

It is possible that some of the degradation can be automatically reversed, prior to making master copies. Ideally, frames would be directly digitized, then restored, and finally transferred to video. However, it is more common, and (with current technology) more convenient, to digitize frames from video, the latter having been made more or less directly from film. There is a rather complicated interleaving process (which we shall not go into) that makes 30 video frames out of every 24 film frames. Suffice it to say that we experimented with digital frames produced by the standard procedure, i.e., film to video to digital.

Our experiments were on "Frankenstein Unbound," a recent release from The Mount Company, which exhibits mostly intraframe artifacts. A common step in editing is the production of black and white film from color film, producing what is known as a "black and white dupe." This black and white copy is heavily used and accumulates scratches and foreign material, thereby degrading in the same fashion as older films. The sequences used in our experiments were selected and digitized from a video cassette made from one such black and white copy.

4. DATA TERM

The plan is to successively restore one frame at a time, based upon, at the very least, the information contained in the unprocessed "current" frame. Of course there is additional relevant information, both in the preceding and ensuing frames. It is unlikely that a scratch or dirt particle will be found in identical locations on two successive frames. (Indeed, the fact that ordinary noise is uncorrelated from frame to frame is the principal motivation for (purely) temporal filtering [5].) To the extent that a scene is more or less unchanged in a $\frac{1}{24}$ second interval, the information in neighboring frames is relevant to the restoration. Many ideas come to mind: use several pre-

ceding frames, somehow weighted by the length of time to the current frame; use processed, instead of unprocessed, previous frames; use *both* previous and ensuing frames. In our experiments, restorations were based simply upon the current and immediately preceding *unprocessed* frames. We view this as a decidedly minimal use of the available information.

Thus the data, $\mathbf{y} = \{y_i\}_{i \in T}$, consist of two images, $\mathbf{y} = (\mathbf{y}^c, \mathbf{y}^p)$, where \mathbf{y}^c is the current unprocessed frame, and \mathbf{y}^p is the preceding unprocessed frame. Both arrays, \mathbf{y}^c and \mathbf{y}^p , are indexed by the same pixel lattice S that indexes the restored image \mathbf{x} : $\mathbf{y}^c = \{y_s^c\}_{s \in S}$ and $\mathbf{y}^p = \{y_s^p\}_{s \in S}$, so that $T = S \cup S$. The (negative) log likelihood term, V , in (1) is made up of two corresponding pieces:

$$V(\mathbf{x}, \mathbf{y}) = V^c(\mathbf{x}, \mathbf{y}^c) + V^p(\mathbf{x}, \mathbf{y}^p).$$

In each term, we "tie" the restoration to the data on a pixel-by-pixel basis:

$$V^\kappa(\mathbf{x}, \mathbf{y}^\kappa) = \alpha^\kappa \sum_{s \in S} \phi^\kappa(x_s - y_s^\kappa), \quad \kappa = c \text{ or } p.$$

If, for example, we were to model the degradation as independent Gaussian noise, then we would take $\phi(u)$ to be quadratic. The quadratic, however, is inappropriate, both when $\kappa = c$ and when $\kappa = p$. Scratches and dust, for example, do not present themselves with a particular preference for the "true" grey level. Also, if we follow a pixel through successive frames then we will occasionally see a large change in grey level, associated, for example, with motion or a scene change. These observations argue against a quadratic function, and in fact for a bounded, or at least slowly growing, function of the general form used in constructing the *spatial* term $U(\mathbf{x})$ in (1).

We settled on the function ϕ in (2). For the experiments we used $\gamma = 2$. Perhaps there is no good reason for favoring this value over, say, $\gamma = 1$, which was used for the spatial term. In any case, we do not expect that the results would be very different with $\gamma = 1$ (or, for that matter, with $\gamma = 2$) throughout; no systematic evaluation was made. Putting together the above pieces we arrive at a data term of the form

$$V(\mathbf{x}, \mathbf{y}) = -\alpha^c \sum_{s \in S} \left(1 + \left| \frac{x_s - y_s^c}{\delta^c} \right|^2 \right)^{-1} - \alpha^p \sum_{s \in S} \left(1 + \left| \frac{x_s - y_s^p}{\delta^p} \right|^2 \right)^{-1}. \quad (7)$$

There are four new parameters, α^c , α^p , δ^c , and δ^p . These were chosen more or less "by hand," as discussed in Section 6.

5. MOTION COMPENSATION

One reason for using a bounded ϕ (see (2)) when tying the current reconstruction to the previous frame is to accommodate discontinuities due to motion. Of course scratches and dirt result in the same kinds of "temporal" discontinuities. This creates a dilemma: Concerning motion, we hope to suspend, or at least limit, the influence of the previous frame at those pixels which undergo large changes in grey level in the transition to the current frame, whereas these same pixels are to be *exploited* in our efforts to remove defects. In principle, the problem is mitigated by the *spatial* (or prior) functional U , which assumes relatively large values in the vicinity of small or thin structures such as dirt or scratches. But this suggests a rather delicate balance, and indeed we were unable to adjust parameters in such a way as to remove defects without introducing some artifact due to motion. An example of motion artifact is given in Section 6.

Evidently, motion compensation is necessary to properly register the frame sequences so that temporal bonding is based on matching scene locations rather than pixel locations. In fact, estimation of the "displacement field" is a standard preprocessing step in most algorithms for frame sequence restoration and enhancement; see [5, 11, 12, and 15]. There are several predominant methods, all based upon certain simplifying assumptions, mainly that there exists a pixel-based correspondence between successive frames that preserves brightness and represents *uniform* motion. In "region-matching," one further assumes that the displacement vector is *constant* over small regions and estimates these shifts separately for each region by minimizing some pixel-based error measure. Another technique is to compute the velocity field by utilizing the "spatio-temporal constraint equation," which is a linear equation relating the two velocity components at a given location and the spatial and temporal partial derivatives of the intensity function y , or, more accurately, its continuum analog. As in region-matching, some additional constraints are necessary, such as constancy over small regions, which then leads to a linear system; see [12]. Obviously, the simplifying assumptions promote inaccuracies when the motion is more complex, for example due to zoom or rotation, or when objects are uncovered. In addition, errors in estimating derivatives are accentuated by noise.

We have chosen the most elementary form of region-matching: a simple *rigid* shift, bringing the previous frame y^p into registration with the current frame y^c . Specifically, we chose the vector τ to minimize

$$\sum_{s \in S} (y_s^c - y_{s+\tau}^p)^2,$$

and then, in (7), replaced $\{y_s^p\}_{s \in S}$ by $\{y_{s+\tau}^p\}_{s \in S}$. This choice

was motivated by two factors. First, we have only worked with 192×192 windows selected from within the full 720×486 frames. Second, our primary goal here is to investigate the feasibility of our approach rather than to develop a full-blown algorithm amenable to commercial utilization.

Obviously, extensions will be needed for accommodating more complex displacements and for automatically processing full frames. Perhaps some form of region-matching will be adequate if the "boundary problem" can be solved by "gluing together" the restorations of small windows, in each of which motion is well-approximated by a simple shift. A more ambitious plan is to perform enhancement and motion compensation *at the same time*, allowing pending restorations to influence the displacement estimates and vice-versa. Still another possibility, perhaps more workable, is to estimate more or less nonparametrically a "background deformation," in the manner of Amit, Grenander, and Piccioni [1]. This would provide a global nonlinear map for registering a pair of successive frames. Finally, another issue is scene changes, for which no transformation will adequately register the previous frame with the current frame. However, these should be easy to spot, automatically, and the obvious remedy is to then remove the temporal term y^p in (7). In any case, a fully operational procedure for restoring entire movies would require considerably more work on this aspect of the problem.

6. EXPERIMENTS

All of the results shown below were obtained under the second order model. Actually, we experimented with all three models: first, second, and third order. As we mentioned earlier, the results with the second order model were generally better than those with the first order model, but we found no corresponding improvement in going to the third order model. Of course, since our experiments only involve processing several successive frames, such comparisons are based on observing the restored frames individually rather than as a *video display*; ultimately, the performance of any algorithm must be tied to the final objectives. Still, it is evident that certain types of noise and artifacts are indeed removed (see below), and natural boundaries and contours appear to be preserved; in particular, there is little if any evidence of the "patchiness" associated with first order smoothing. Whether any unforeseen artifacts appear in the video remains to be seen.

The restoration is the result of an iterative algorithm designed to minimize the cost functional (1). Specifically, under the second order model (6), with motion-compensated data, the cost functional used was

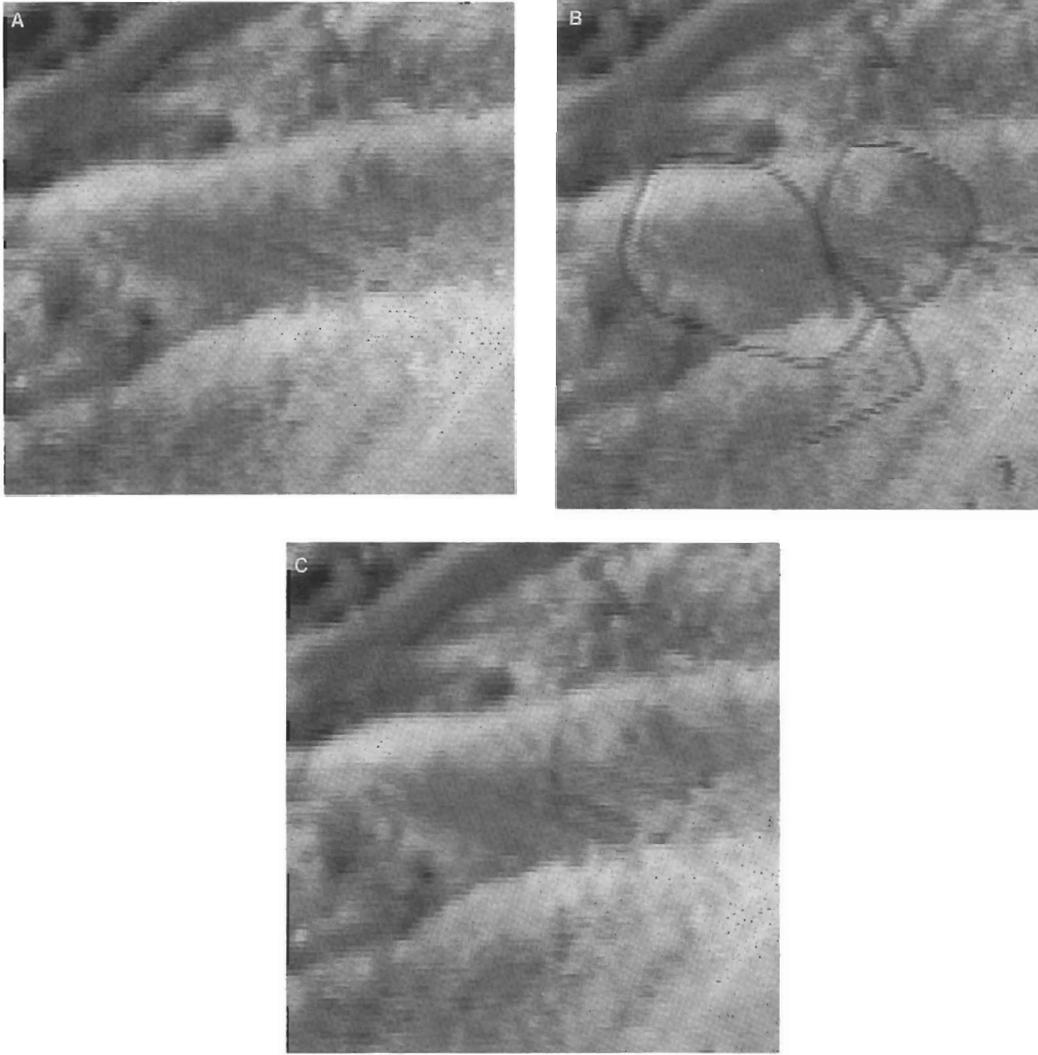


FIG. 1. Panels A and B: Successive frames. Panel B is contaminated by the hair. Panel C: Restoration of B based upon B and A. The hair is almost completely removed.

$$\begin{aligned}
 & \alpha \sum_{\xi_1=1}^{N-2} \sum_{\xi_2=1}^N \phi(x(\xi_1 + 2, \xi_2) - 2x(\xi_1 + 1, \xi_2) + x(\xi_1, \xi_2)) \\
 & + \alpha \sum_{\xi_1=1}^N \sum_{\xi_2=1}^{N-2} \phi(x(\xi_1, \xi_2 + 2) - 2x(\xi_1, \xi_2 + 1) \\
 & + x(\xi_1, \xi_2)) \\
 & + 2\alpha \sum_{\xi_1=1}^{N-1} \sum_{\xi_2=1}^{N-1} \phi(x(\xi_1 + 1, \xi_2 + 1) + x(\xi_1, \xi_2) \\
 & - x(\xi_1 + 1, \xi_2) - x(\xi_1, \xi_2 + 1)) \\
 & + \alpha^c \sum_{\xi_1=1}^N \sum_{\xi_2=1}^N \phi^c(x(\xi_1, \xi_2) - y^c(\xi_1, \xi_2)) \\
 & + \alpha^p \sum_{\xi_1=1}^N \sum_{\xi_2=1}^N \phi^p(x(\xi_1, \xi_2) - y^p(\xi_1 + \tau_1, \xi_2 + \tau_2)). \quad (8)
 \end{aligned}$$

The shift, $\tau = (\tau_1, \tau_2)$, was chosen by least squares, as explained in Section 5. In all experiments $N = 192$. The functions ϕ , ϕ^c , and ϕ^p are all of the form (2), with $\gamma = 1$, $\gamma = 2$, and $\gamma = 2$, respectively (see Sections 2 and 4), and have three additional parameters δ , δ^c , and δ^p . Thus there are six parameters to be specified: α , α^c , α^p , δ , δ^c , and δ^p .

Our past experience has been that results are not particularly sensitive to the scale parameters δ , δ^c , and δ^p . These were chosen to be 5, 10, and 10, respectively, and never changed. This left three parameters, α , α^c , and α^p , but since we were interested in *minimizing* (8) (rather than computing, for example, a *posterior mean*) there was no harm in setting $\alpha = 1$. After that, the procedure was basically trial and error, although good “ballpark” guesses were made by carrying out a few simple thought experiments. Essentially, this is the method of “repara-

metrization" (see [6]). For example, one imagines being faced with raw data for which there are two reasonable, but quite different, restorations. By choosing which should be favored, or by specifying that the two are equally favorable, one derives an inequality, or equality, in the free parameter values. We ended up with $\alpha^c = 6$ and $\alpha^p = 10$.

Pixels (ξ_1, ξ_2) were visited successively and the current grey level value $x(\xi_1, \xi_2)$ was replaced by that value which minimized (8), fixing the values $x(\xi'_1, \xi'_2)$ at all other pixels $(\xi'_1, \xi'_2) \neq (\xi_1, \xi_2)$. In the Bayesian paradigm, this is the Iterated Conditional Mode algorithm that Besag recommends (see [2]). In order to minimize any artifacts induced by the order of visitation, two pixels contributing to a common summand in (8) were never visited in succession. This was done as follows. Such pixels are "neighbors" relative to the graph structure inherent in

(8). The $N \times N$ square lattice was first "colored": colors were assigned to pixels in such a way that pixels of the same color were never neighbors in the sense above. A sweep, then, consisted of a loop through the different colors, updating, for each color, all associated pixels.

The results of such "greedy" algorithms are often highly dependent on the initialization. Unfortunately, this was the case in our experiments. If, for example, we initialized with the previous (unprocessed) frame, $\mathbf{x} = \mathbf{y}^p$, then the restoration sometimes inherited inappropriate features from this initialization. We tried many variations, including random initialization, and convex combinations of the previous and current frames. Overall, we did best by simply using the current unprocessed frame: $\mathbf{x} = \mathbf{y}^c$. This is what was done for all of the experiments reported below.

The first experiment was with the raw data shown in



FIG. 2. Three successive frames. Panel B is damaged by the scratch. There is a trace of a scratch in the lower left corner of Panel A.



FIG. 3. Panel A: Restoration of Fig. 2B, based upon 2B and 2A, *without motion compensation*. Although the scratch is mostly removed, there are motion artifacts. Panel B: Same as A, but with motion compensation. Again, the scratch is mostly removed; this time there are no discernible motion artifacts. Panel C: Restoration of 2C based upon 2C and 2B. Note that the scratch from 2B is not pulled into the restoration.

Fig. 1, panels A and B. These are two successive unprocessed frames. The picture in panel A is therefore \mathbf{y}^p , and the one in panel B is \mathbf{y}^c . Obviously, the frame in B is defective; there was apparently a hair stuck to it before the conversion to video. The restoration is shown in panel C. The hair is mostly removed, and there is no detectable artifact. The result, of course, reflects both the panel A and panel B frames. One way to measure the respective contributions is through distances from one frame to another. We used an L_1 norm: the average absolute grey level difference over the N^2 pixels. Using $\mathbf{x} = \{x_s\}_{s \in S}$ to represent the restored picture, and $\|\cdot\|$ to rep-

resent the L_1 norm, the results are as follows: $\|\mathbf{y}^c - \mathbf{y}^p\| = 4.21$, $\|\mathbf{x} - \mathbf{y}^c\| = 2.40$, and $\|\mathbf{x} - \mathbf{y}^p\| = 2.02$.

A second set of experiments was performed on the *three* successive frames shown in Fig. 2, panels A, B, and C. The actor (Raul Julia) is falling, and there is significant frame-to-frame movement. Note the scratch in panel B, and in the lower left of panel A. Figure 3A shows the restoration of 2B, based upon the frames in 2A and 2B (\mathbf{y}^p and \mathbf{y}^c respectively), *except that no compensation for motion was made*. Although the scratch is removed, motion artifacts are clearly introduced. Figure 3B is the result of the same experiment, incorporating motion

compensation. The between-picture distances are $\|y^c - y^p\| = 3.23$, $\|x - y^c\| = 1.52$, and $\|x - y^p\| = 1.93$.

Finally, it may be suspected that the restored image will carry along undesirable properties from the previous frame, since the restoration is evidently some kind of mixture of both the current and the previous frames. Figure 3, panel C, is the restoration of the last frame in Fig. 2. This time, 2B is the preceding frame. The scratch is not introduced into the restoration. The distances are $\|y^c - y^p\| = 2.80$, $\|x - y^c\| = 1.20$, and $\|x - y^p\| = 1.83$.

ACKNOWLEDGMENTS

Jay Cassidy from the Mount Film Company, Los Angeles, California, provided the film sequences and much more in the way of technical advice and information on the film restoration problem, and Yali Amit provided valuable advice on motion compensation.

REFERENCES

1. Y. Amit, U. Grenander, and M. Piccioni, Structural image restoration through deformable templates, *J. Am. Statist. Assoc.* **86**, 1991, 376–388.
2. J. Besag, On the statistical analysis of dirty pictures (with discussion), *J. R. Statist. Soc. Ser. B* **48**, 1986, 259–302.
3. J. Besag, Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. R. Statist. Soc. Ser. B*, **36**, 1974, 192–236.
4. T. J. Dennis, Nonlinear temporal filter for television picture noise reduction, *IEEE Proc.* **127**, 1980, 52–56.
5. E. Dubois and S. Sabri, Noise reduction in image sequences using motion-compensated temporal filtering, *IEEE Trans. Commun.* **32**, 1984, 826–831.
6. D. Geman and S. Geman, Bayesian image analysis, in *Disordered Systems and Biological Organization* E. Bienenstock, F. Fogelman, and G. Weisbuch, Ed., Springer-Verlag, Berlin, 1986.
7. D. Geman and G. Reynolds, Constrained restoration and the recovery of discontinuities, *IEEE Trans. Pattern Anal. Mach. Intelligence*, **14**, 1992, 367–383.
8. S. Geman and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. Pattern Anal. Mach. Intelligence* **6**, 1984, 721–741.
9. S. Geman and D. E. McClure, Bayesian image analysis: An application to single photon emission tomography, in *1985 Proceedings of the Statistical Computing Section, American Statistical Association, 1985*, pp. 12–18.
10. U. Grenander, *Tutorial in Pattern Theory*, Technical Report, Division of Applied Mathematics, Brown University, 1983.
11. T. S. Huang, *Image Sequence Analysis*, Springer-Verlag, Berlin, 1981.
12. J. S. Lim, *Two-Dimensional Signal and Image Processing*, Prentice-Hall, Englewood Cliffs, NJ, 1990.
13. J. Limb and J. Murphy, Estimating the velocity of moving image in television signals, *Comput. Graphics Image Process.* **4**, 1975, 311–327.
14. J. L. Marroquin, S. Mitter, and T. Poggio, Probabilistic solution of ill-posed problems in computational vision, *J. Am. Statist. Assoc.* **82**, 1987, 76–89.
15. R. Paquin and E. Dubois, A spatio-temporal gradient method for estimating the displacement field in time-varying imagery, *Comput. Vision Graphics Image Process.* **21**, 1983, 205–221.
16. T. Poggio, V. Torre, and C. Koch, Computational vision and regularization theory, *Nature* **31**, 1985, 314–319.