JOURNAL OF VISUAL COMMUNICATION AND IMAGE REPRESENTATION
Vol. 9, No. 4, December, pp. 352–365, 1998
ARTICLE NO. VC980392

Denoising Color Images Using Regularization and "Correlation Terms"

Daniel Keren[1] and Anna Gotlib

Department of Computer Science, University of Haifa, Haifa 31905, Israel

Received February 5, 1998; accepted July 28, 1998

The problem addressed in this work is restoration of images that have several channels of information. We have studied color images so far, but hopefully the ideas presented here apply to other types of images with more than one channel. The suggested method is to use a probabilistic scheme which proved rather useful for image restoration and to incorporate into it an additional term which results in a better correlation between the color bands in the restored image. Results obtained so far are good; typically, there is a reduction of 20 to 40% in the mean square error, compared to standard restoration carried out separately on each color band. The contributions suggested in this work are the introduction of "correlation terms," which augment "standard" regularization, and the process of choosing *two* regularization hyperparameters. Also, a relation between the algorithm suggested here and the recently introduced ideas of smoothing by diffusion in color space is explored.

1. INTRODUCTION AND PREVIOUS WORK

The last years have witnessed a growth in the amount of research on multichannel image restoration; see, for instance, [24, 10, 18, 11, 4, 5, 9, 16, 13, 14]. In these works, the between-channel correlation was used to restore multichannel images. Various tools were called to task, including regularization [24, 10, 9], Kalman filtering [4], least square restoration [5], Wiener filtering [16], stochastic methods using Markov random fields [18], differential geometry [13, 14, 17], and total variation methods [1].

Difficulties in some of the aforementioned approaches are the necessity of manipulating very large matrices of a nontrivial structure, and the problem of estimating the image's autocorrelation function. In [24], it was suggested overcoming these problems by using regularization. This was accomplished by adjusting the "smoothness term" so that it will force smoothness not only in the spatial domain of each color, but also between the distinct color channels. Specifically, the Laplacian operator, whose norm is often used to estimate the "roughness" of a grey-level image, was replaced by a three-dimensional Laplacian, which also incorporates between-channel smoothness. Also, the GCV method for estimating the regularization parameter (see Section 4) was extended to handle multichannel regularization.

Our approach resembles that of [24]; however, we suggest measuring the smoothness in color space by different methods. This is motivated by the observation that spatial smoothness and "color smoothness" are, in our opinion, rather different entities. For instance, an image can have a very rough spatial structure, but be very smooth in color space, and vice versa. "Color smoothness" is naturally defined by the change in the chrominance, not the brightness.

The difference between spatial smoothness and color smoothness can be demonstrated by a simple example. Suppose P_1 and P_2 are adjacent pixels. In case (a), let the color values of P_1, in the red-green-blue (RGB in the sequel) channels, be (r, g, b), and those of P_2 be $(r + \delta, g + \delta, b + \delta)$. In case (b), the values are (r, g, b) and $(r + \delta, g - \delta, b + \delta)$, respectively. In both cases, the spatial smoothness, as measured separately in the three color channels, is the same; however, in case (b), the clique of P_1 and P_2 is less smooth in color space.

We applied two different tools to measure color smoothness:

- A Bayesian measure. Here, the average colors and color covariance matrix at each pixel were estimated from a small neighborhood; the process was bootstrapped from the "simple" denoised image, that is, the image consisting of the three color channels obtained by independently denoising each channel of the original image. Then, the "color smoothness" was estimated by computing the probability of each color pixel in the Gaussian distribution with the aforementioned average and covariance.
- A "geometric measure." Here, the following intuitive idea was adopted: an image is smooth in color space if, on the average, the angle between adjacent color pixels—when viewed as vectors in RGB space—is small. As a

[1] Corresponding author. E-mail: dkeren@mathcs2.haifa.ac.il.

1047-3203/98 $25.00

measure of the angle, the squared norm of the vector product between adjacent pixels was used. The advantage over using the angle itself is that this norm is a more manageable function of the color values. In the Appendix, we offer a possible explanation as to why the geometric measure is suitable for restoring color images; namely, it is shown that it may be viewed as a natural extension of ''standard'' regularization, which was extensively used to restore gray level images.

While the two methods are different, they share a similar feature: the ''cost functional'' to be minimized is the sum of the standard ''data fidelity'' and ''spatial smoothness'' terms, and a novel ''color correlation term.'' The difference is in the latter; the Bayesian measure is image (and spatial) dependent and uses an estimate of the local image structure. The geometric measure is defined globally and is more closely related to standard regularization terms—although it introduces a more complicated, nonquadratic cost functional.

The geometric measure was clearly superior to the Bayesian measure; in all the images we tested, it yielded restorations which were better both according to the mean square error criterion and appearance. This may well indicate that the geometric measure is quite suitable to measure smoothness in color space.[2]

1.1. *A Relation between ''Correlation Terms'' and Smoothing by Diffusion*

Recently, it was proposed to process color images by subjecting them to a diffusion process [13, 14, 17]. The color image is viewed as a two-dimensional manifold in fifth-dimensional Euclidean space (two dimensions for the spatial coordinates of the image, and three for the RGB values). In [13, 14], the image is then subjected to a diffusion process (the ''Beltrami flow'') and, in the limit, converges to a minimal surface—that is, a surface whose mean curvature is everywhere zero. Such surfaces are known to be extremal points of the area function [2]. The intuitive idea is that the smaller the area, the ''nicer'' the image is; hence, this diffusion process is a natural generalization of regularization to higher dimensions. The results are superior to those obtained by applying regularization separately in every channel.

In the Appendix we implement this idea, by directly optimizing a cost functional which includes the surface area as a summand. A relation between the surface area and the augmented ''smoothness term,'' which includes the color correlation terms introduced here, is noted. However, on the images tested, restoration by using the surface area as a measure for image quality resulted in inferior results (according to the mean square error criterion) as compared to those obtained when using the correlation terms. This may be due to the fact that the expression for the area contains the squared first derivatives of the color values (or the so-called first-order smoothness term); it has been our experience that using this term is too restrictive and that better results on natural images are obtained when the second-order term (using second derivatives) is used [12].

However, the results obtained by using the surface area as a measure for image quality are, in general, pleasing to the eye. This touches a fascinating and yet unresolved question in image processing—by what criterion should different restoration methods be compared?

2. EXTENDING THE BAYESIAN PARADIGM TO COLOR IMAGES

A rather general formulation of the restoration problem is the following: Given some partial information D on an image F, find the best restoration for F. Obviously, there are many possible ways in which to define ''best.'' One way, which proved quite successful for a wide range of applications, is probabilistic in nature: Given D, one seeks the restoration $\hat{F}$ which maximizes the probability $\Pr(F/D)$. Following Bayes' rule, this is equivalent to maximizing $\Pr(D/F)\ \Pr(F)/\Pr(D)$. The denominator is a constant once D is measured; $\Pr(D/F)$ is usually easy to compute. $\Pr(F)$ is more interesting and more difficult to define. Good results have been obtained by following the physical model of the Boltzman distribution, according to which the probability of a physical system to be at a certain state is proportional to the exponent of the negative of the state's energy—that is, low-energy, or ''ordered'' states, are assigned higher probabilities than high-energy, or ''disordered,'' states [6, 20]. It is common to define the energy of a signal by its ''smoothness''; the energy of a two-dimensional signal F is often defined as $\int\int (F_{xx}^2 + 2F_{xy}^2 + F_{yy}^2)\, dx\, dy$. Such integrals are usually called ''smoothing terms,'' as they force the resulting restoration to be smooth [12, 21, 8, 19].

To see how the probabilistic approach naturally leads to restoration by so-called ''smoothing,'' or regularization, let us look at the problem of restoring a two-dimensional image from samples which are corrupted by additive noise. Suppose the image is sampled at the points $\{(x_i, y_i)\}$, the sample values are z_i, and the measurement noise is Gaussian with variance σ^2. Then

$$\Pr(D/F) \propto \exp\left(-\sum_{i=1}^{n} \frac{[F(x_i, y_i) - z_i]^2}{2\sigma^2}\right)$$

[2] It is interesting to note here that we have also applied this geometric measure to the *Demosaicing* problem, in which the color image is subsampled in color space—only one color is given at each pixel, in a mosaic pattern. For this problem, too, the geometric measure yielded better results than other methods we tested.

and, based on the idea of the Boltzman distribution, one can define $\Pr(F)$ as being proportional to

$$\exp\left(-\lambda \iint (F_{xx}^2 + 2F_{xy}^2 + F_{yy}^2)\, dx\, dy\right)$$

for some positive constant λ. So, the overall probability to maximize is

$$\exp\left(-\left(\sum_{i=1}^{n} \frac{[F(x_i, y_i) - z_i]^2}{2\sigma^2} + \lambda \iint (F_{xx}^2 + 2F_{xy}^2 + F_{yy}^2)\, dx\, dy\right)\right)$$

which is, of course, equivalent to minimizing

$$\sum_{i=1}^{n} \frac{[F(x_i, y_i) - z_i]^2}{2\sigma^2} + \lambda \iint (F_{xx}^2 + 2F_{xy}^2 + F_{yy}^2)\, dx\, dy. \quad (1)$$

This leads, via calculus of variations, to a partial differential equation, which can be effectively solved using finite element methods [21].

Now, suppose we are given a color image, with RGB channels, which was corrupted by additive noise with known variance σ^2 (assume for the meantime that $\sigma = 1$). One obvious way to restore the image is to apply the denoising algorithm described above for each of the separate channels and to combine the restored channels into a color image. Such an approach, however, does not work well in general. Usually, the resulting image is of low quality, and contaminated by false colors; that is, certain areas contain streaks of colors which do not exist in the original image. This problem is more acute in highly textured areas.

The proposed solution is to incorporate into the probabilistic scheme a ''correlation term,'' which will result in a better correlation between the RGB channels. If $C_{x,y}$ is the covariance matrix of the RGB values at a pixel (x, y), and $(\overline{R}, \overline{G}, \overline{B})$ the average colors in the pixel's vicinity, then, assuming a normal distribution, the probability for the combination of colors $(R(x, y), G(x, y), B(x, y))$ is proportional to $\exp(-\frac{1}{2}(R(x, y) - \overline{R}, G(x, y) - \overline{G}, B(x, y) - \overline{B})C_{x,y}^{-1}(R(x, y) - \overline{R}, G(x, y) - \overline{G}, B(x, y) - \overline{B})^t$. Multiplying over all the pixels results in adding these terms in the exponent's power. Exactly as in the interpolation problem above, this exponential term combines with the other exponentials, and we get a combined exponential that has to be maximized; therefore, we have to minimize the negative of the power, which simply results in adding the ''correlation term,'' $\iint (R(x, y) - \overline{R}, G(x, y) - \overline{G}, B(x, y) - \overline{B})C_{x,y}^{-1}(R(x, y) - \overline{R}, G(x, y) - \overline{G}, B(x, y) - \overline{B})^t\, dx\, dy$, to the expression of Eq. (1). In effect, this term makes use of the fact that, in natural and synthetic images, the RGB channels are usually highly correlated. The ''correlation term'' penalizes deviations from this correlation, thus ''pushing'' the restored image towards one whose channels are ''correctly correlated.''

The combined expression to minimize is the following extension of Eq. (1):

$$\begin{aligned}\|F - H\|^2 + \lambda_1 \Bigg(&\iint (R_{xx}^2 + 2R_{xy}^2 + R_{yy}^2)\, dx\, dy \\ &+ \iint (G_{xx}^2 + 2G_{xy}^2 + G_{yy}^2)\, dx\, dy \\ &+ \iint (B_{xx}^2 + 2B_{xy}^2 + B_{yy}^2)\, dx\, dy\Bigg) \\ &+ \lambda_2 \iint (R(x, y) - \overline{R}, G(x, y) - \overline{G}, B(x, y) - \overline{B})C_{x,y}^{-1} \\ &\quad (R(x, y) - \overline{R}, G(x, y) - \overline{G}, B(x, y) - \overline{B})^t\, dx\, dy.\end{aligned} \quad (2)$$

Here, H is the measured (noised) color image and F is the restoration, which is composed of the R, G, and B channels. λ_1 and λ_2 are positive constants (see Section 4 for a discussion on how to choose these ''hyperparameters'').

We have implemented a straightforward iterative scheme for minimizing this functional. The covariance matrices are estimated, in a bootstrapping fashion, from the color image obtained by restoring each channel separately. Further iterations using the covariance matrices of the new images have not improved the results.

We do not include results obtained when using the Bayesian color correlation term, as they were inferior to those obtained using the method presented below. Nonetheless, the Bayesian paradigm justifies the use of ''color correlation terms.'' The method described in the following section can be derived from the Bayesian paradigm by using a different measure for the ''energy'' of a color image—one determined by the angles between adjacent pixels when viewed as vectors in RGB space.

3. THE GEOMETRIC COLOR CORRELATION TERM

A substantial improvement—both in quality and speed—over using the Bayesian color correlation term, was obtained by using a different term, defined as the sum of squared norms of the vector products between neighboring pixels, when viewed as vectors in $\mathscr{R}^3$. The underlying intuition is straightforward: since natural images are generally smooth both in the spatial and color spaces, one can expect that neighboring color pixels will have similar directions in color space—hence, their vector product will be small. We have tried using the scalar product for the same goal; however, the results were inferior to those obtained with the vector product. In the Appendix,

a more concrete relation between the vector products and the geometric properties of the color image is derived.

The functional to minimize is (as defined on the discrete image)

$$\begin{aligned}\|F-H\|^2 &+ \lambda_1\Bigg(\sum_{i,j}(R_{xx}(i,j)^2+2R_{xy}(i,j)^2+R_{yy}(i,j)^2)\\ &+\sum_{i,j}(G_{xx}(i,j)^2+2G_{xy}(i,j)^2+G_{yy}(i,j)^2)\\ &+\sum_{i,j}(B_{xx}(i,j)^2+2B_{xy}(i,j)^2+B_{yy}(i,j)^2)\Bigg)\\ &+\lambda_2\sum_{i,j}\sum_{k,l/(k,l)\in N(i,j)}\|P_{i,j}\times P_{k,l}\|^2,\end{aligned}\tag{3}$$

where F is the sought image, H is the measured image, $P_{i,j}=(R(i,j),G(i,j),B(i,j))$ is the pixel of F at location (i,j), the xx, xy, yy subscripts denote the partial derivatives by x and y, and $N(i,j)$ is the 3×3 set of $P_{i,j}$'s neighbors.

This functional resembles the one in Eq. (2), but it uses the new correlation term. Both correlation terms force a ''nice'' behavior on the image in color space; however, the second one is universal in nature and does not depend on the color covariance matrices of the specific image.

It is interesting to note that both correlation terms are nonquadratic; the Bayesian one uses the pixel's values to determine the covariance matrices, and the geometric one contains fourth powers of the color values. This may be viewed as a liability, since it is more difficult to minimize such expressions; however, in all the examples we tested, the iterative minimization scheme converged quite fast (Section 5.1). This is probably due to the fact that the set of measurements is dense.

Our experiments suggest that the advantage of using a more general optimizing functional outweighs the difficulty incurred by minimizing the nonquadratic expression.

We note here that the vector product term is biased towards areas with higher color values. This can be overcome, for instance, by restoring the logarithm of the image. Experiments have not shown any significant difference when this is done. Also, it may be argued that this bias is possibly an advantage, as the areas with the higher color values are the more important ones.

4. CHOOSING THE HYPERPARAMETERS λ_1 AND λ_2

A problem which attracted a great deal of attention in regularization theory is the choice of the regularization hyperparameter, usually denoted by λ, which determines the trade-off between the smoothness and the fidelity to the data (see Eq. (1)); as λ increases, the solution becomes smoother, but may diverge from the measured data. If λ is too small, the solution may ''overfit'' the noise. Some approaches for choosing the correct λ are presented in [7, 15, 3, 23, 22, 20]. In [24], regularization is extended to deal with color images, and the generalized cross validation (GCV) method is also extended to choose an appropriate λ.

The idea of cross validation is to choose a λ so that the data points ''predict one another.'' Using the notations of Section 2, one proceeds as follows: for each sample point (x_k,y_k), $1\le k\le n$, $\hat{F}_k$ is defined as the function minimizing

$$\sum_{i\neq k}^{n}[F(x_i,y_i)-z_i]^2+\lambda\iint(F_{xx}^2+2F_{xy}^2+F_{yy}^2)\,dx\,dy,$$

i.e. the restoration obtained by considering all the data points but the ''removed'' kth point. A function $V_0(\lambda)$ is then defined as $\sum_{k=1}^{n}[\hat{F}_k(x_k,y_k)-z_k]^2$, and the λ chosen is the one minimizing $V_0(\,)$. This algorithm is called ordinary cross validation (OCV).

An improvement of this method is the GCV algorithm; see [3, 24] for further details. Here, we have applied OCV.

A more straightforward method can be used if σ^2, the variance of the noise N, is known: if $H=F+N$ is the measured signal, choose a λ which results in a restoration $\hat{F}$ such that $E(\|\hat{F}-H\|^2)=\sigma^2$.

For the method suggested in this work, we need to find *two* hyperparameters, λ_1 and λ_2. Using only one of the criteria will, therefore, not suffice, as it will give a curve in $\lambda_1-\lambda_2$ space. In order to solve this problem, we have used the two criteria simultaneously. First, the pairs of hyperparameters for which $E(\|\hat{F}-H\|^2)=\sigma^2$ were found. Among them, we chose the pair which minimizes the OCV function.

The OCV idea of ''leaving one out'' was implemented as follows. In the functional of Eq. (3), every color element (every color in every pixel), in turn, was ''removed''; that is, the corresponding term in the sum which compromises the data fidelity term $\|F-H\|^2$ was taken out. Then, the image was restored, and the resulting value in that location was subtracted from the original value. These differences are squared and summed to obtain the OCV function.

This process is, however, very time-consuming for large images, as it requires restoring the entire color image $3N$ times, where N is the number of image pixels. A heuristic which worked quite well was to confine the restoration to a small (for instance, 11×11) neighborhood of the pixel whose color element is currently removed. This considerably speeded up the computation of the OCV function. However, choosing the correct hyperparameters is still, computationally, a nontrivial problem (see Section 5.2). It is not clear if the sophisticated techniques used to recover the correct λ [24, 3] can be directly applied here, if at all, due to the complicated, nonquadratic form of the cost functional. At the end of Section 5.2, we offer a very quick heuristic for choosing good values for the hyperparameters.

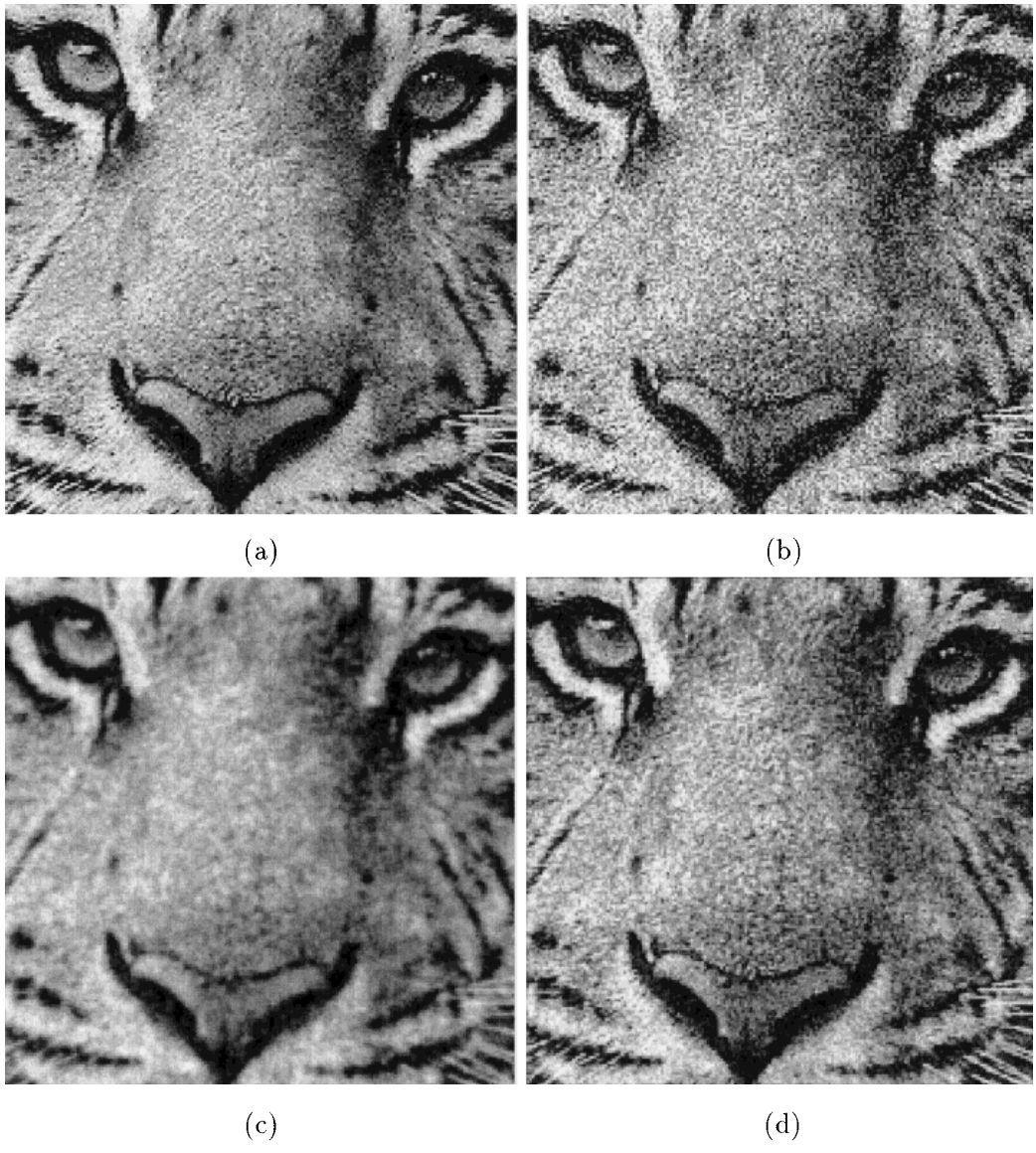


FIG. 1. (a) Red channel of original tiger image; (b) after addition of noise; (c) result of separate regularization in every channel; (d) result after incorporating the color correlation term.

5. IMPLEMENTATION AND RESULTS

In this section, the derivation of the iterative restoration scheme will be presented, as well as some results on running the algorithm with the aforementioned method for choosing the hyperparameters. In the images we tested, the OCV criterion chose "good" values from those satisfying $\|\hat{F} - H\|^2 = \sigma^2$, in the sense that they give good approximations to the original image in the mean square error norm (see Figs. 7b and 7d).

5.1. *Deriving the Iterative Scheme*

The iterative scheme is derived using a simple finite element method. Since the finite-element paradigm tries to annihilate the derivatives of the cost functional, it is enough to look at the neighborhood of a pixel P which consists of pixels that appear in the partial derivatives of the cost functional by the RGB values at P:

	0	1	2	3	4	
0	r_{00}, g_{00}, b_{00}	r_{01}, g_{01}, b_{01}	r_{02}, g_{02}, b_{02}	r_{03}, g_{03}, b_{03}	r_{04}, g_{04}, b_{04}	$\cdots$
1	r_{10}, g_{10}, b_{10}	r_{11}, g_{11}, b_{11}	r_{12}, g_{12}, b_{12}	r_{13}, g_{13}, b_{13}	r_{14}, g_{14}, b_{14}	$\cdots$
2	r_{20}, g_{20}, b_{20}	r_{21}, g_{21}, b_{21}	$(P)\ r_{22}, g_{22}, b_{22}$	r_{23}, g_{23}, b_{23}	r_{24}, g_{24}, b_{24}	$\cdots$
3	r_{30}, g_{30}, b_{30}	r_{31}, g_{32}, b_{31}	r_{32}, g_{32}, b_{32}	r_{33}, g_{33}, b_{33}	r_{34}, g_{34}, b_{34}	$\cdots$
4	r_{40}, g_{40}, b_{40}	r_{41}, g_{42}, b_{41}	r_{42}, g_{42}, b_{42}	r_{43}, g_{43}, b_{43}	r_{44}, g_{44}, b_{44}	$\cdots$
.	.	.	.	.	.	
.	.	.	.	.	.	
.	.	.	.	.	.	

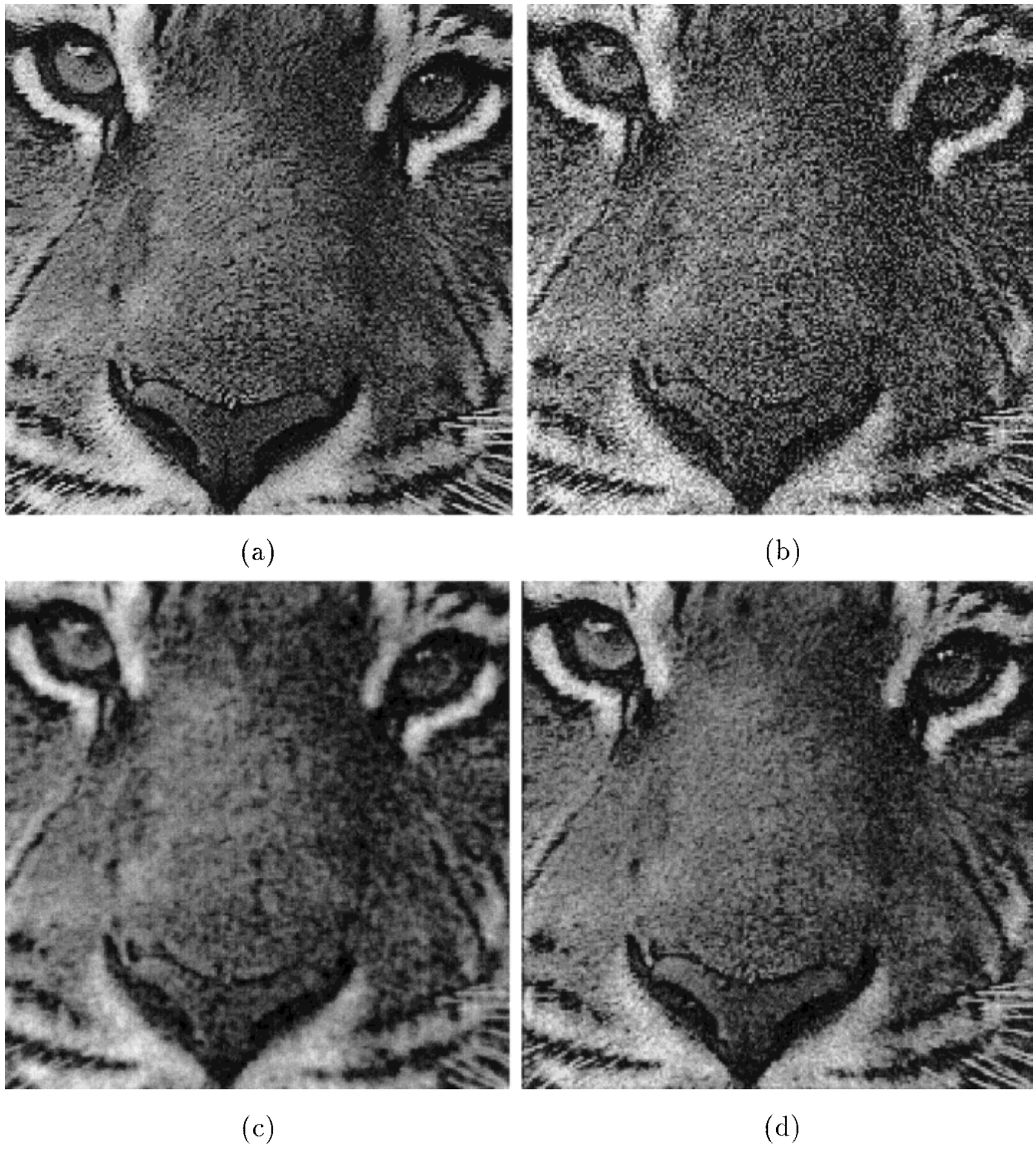


FIG. 2. Same as Fig. 1, for green channel.

The cost functional of Eq. (3), restricted to that small neighborhood, is the sum of following terms:

• *Data fidelity term* is equal to the sum of the squared differences between the candidate image F and the measured image H; the only part which has a nonzero partial derivative with respect to P is $(r_{22} - R_{22})^2 + (g_{22} - G_{22})^2 + (b_{22} - B_{22})^2$, where r, g, b denote the sought values, and R, G, B the measured values.

• *Smoothness term* is equal to the sum of squared discrete approximations to the second partial derivatives. For instance, R_{xx}^2 at P is represented as $(r_{21} - 2r_{22} + r_{23})^2$, etc.

• *Color correlation term* is the sum of squared norms of the vector products of P with its neighbors, for instance, $\|(r_{22}, g_{22}, b_{22}) \times (r_{21}, g_{21}, b_{21})\|^2$.

Adding all these terms gives the relevant part of the cost functional. Annihilating the partial derivatives by P's components (r_{22}, g_{22}, b_{22}) gives a set of equations, and solving these equations using a Gauss–Seidel iterative method results in the iterative scheme for restoring the color image. Since the resulting expressions are nontrivial, the Maple symbolic computation package was used to derive them. The resulting iterative step for updating the value of r_{22}, P's red component, is given below, $r_{22}^{(n+1)}$ is the new value, while the (n) superscript denotes the previous iteration. If one substitutes $\lambda_2 = 0$ in this expression, it reduces to a standard iterative scheme for denoising:

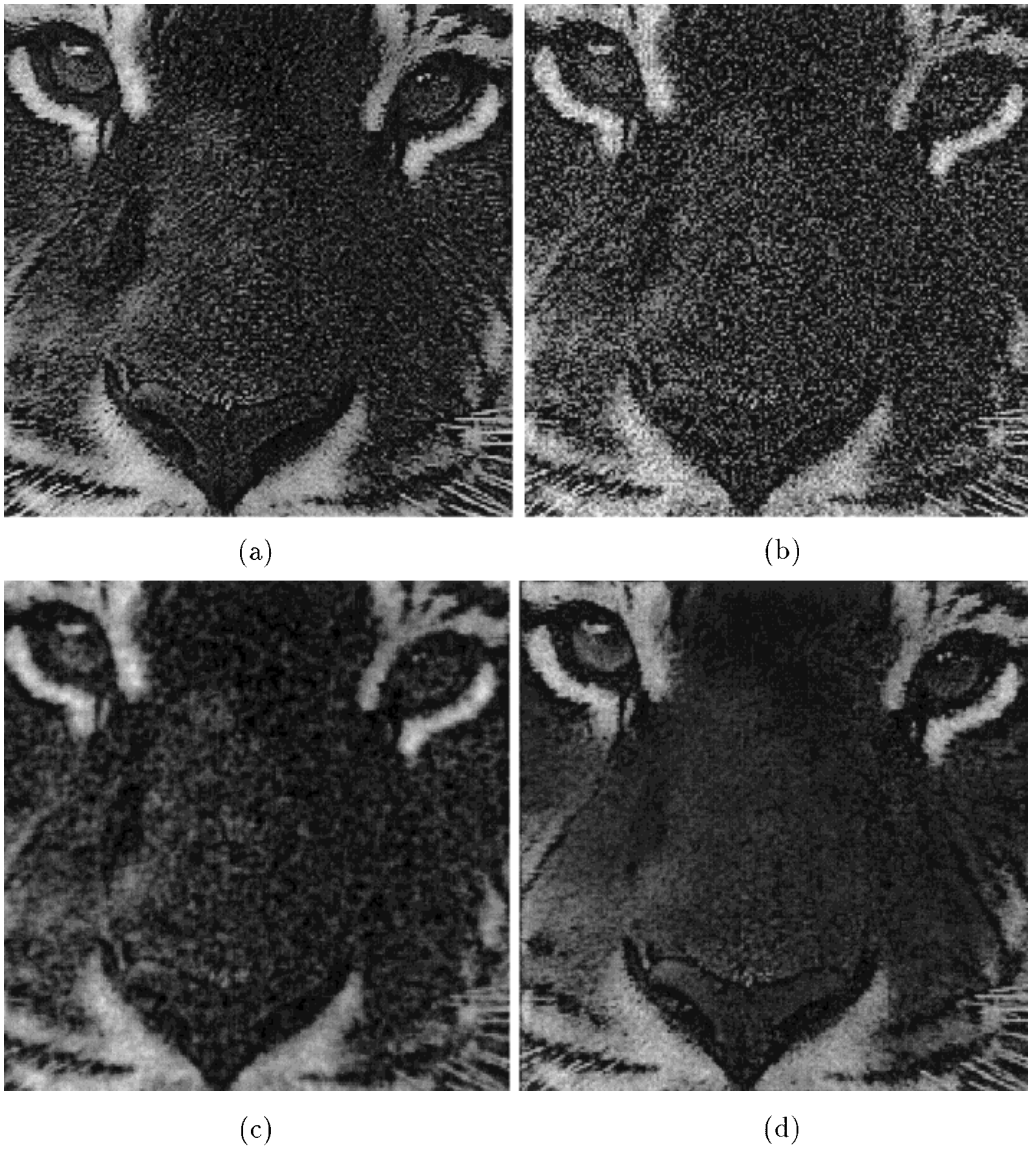


FIG. 3. Same as Fig. 1, for blue channel.

$$
\begin{aligned}
r_{22}^{(n+1)} = \frac{1}{2}\Bigg(& 2R_{22} - \lambda_1 r_{02}^{(n)} + 8\lambda_1 r_{12}^{(n)} - \lambda_1 r_{20}^{(n)} + 8\lambda_1 r_{21}^{(n)} \\
& + 8\lambda_1 r_{32}^{(n)} + 8\lambda_1 r_{23}^{(n)} - \lambda_1 r_{24}^{(n)} - \lambda_1 r_{42}^{(n)} - \frac{\lambda_1 r_{00}^{(n)}}{2} \\
& - \frac{\lambda_1 r_{04}^{(n)}}{2} - \frac{\lambda_1 r_{40}^{(n)}}{2} - \frac{\lambda_1 r_{44}^{(n)}}{2} + 2\lambda_2 r_{11}^{(n)} b_{22}^{(n)} b_{11}^{(n)} \\
& + 2\lambda_2 r_{11}^{(n)} g^{(n)}22 g^{(n)}11 + 2\lambda_2 r_{21}^{(n)} b_{22}^{(n)} b_{21}^{(n)} \\
& + 2\lambda_2 r_{21}^{(n)} g^{(n)}22 g^{(n)}21 + 2\lambda_2 r_{31}^{(n)} b_{22}^{(n)} b_{31}^{(n)} \\
& + 2\lambda_2 r_{31}^{(n)} g^{(n)}22 g^{(n)}31 + 2\lambda_2 r_{12}^{(n)} b_{22}^{(n)} b_{12}^{(n)} \\
& + 2\lambda_2 r_{12}^{(n)} g^{(n)}22 g^{(n)}12 + 2\lambda_2 r_{32}^{(n)} b_{22}^{(n)} b_{32}^{(n)} \\
& + 2\lambda_2 r_{32}^{(n)} g^{(n)}22 g^{(n)}32 + 2\lambda_2 r_{13}^{(n)} b_{22}^{(n)} b_{13}^{(n)} \\
& + 2\lambda_2 r_{13}^{(n)} g^{(n)}22 g^{(n)}13 + 2\lambda_2 r_{23}^{(n)} b_{22}^{(n)} b_{23}^{(n)} \\
& + 2\lambda_2 r_{23}^{(n)} g^{(n)}22 g^{(n)}23 + 2\lambda_2 r_{33}^{(n)} b_{22}^{(n)} b_{33}^{(n)} \\
& + 2\lambda_2 r_{33}^{(n)} g^{(n)}22 g^{(n)}33\Bigg) \Bigg/ (\lambda_2 b_{31}^{(n)^2} + \lambda_2 g_{31}^{(n)^2} \\
& + \lambda_2 b_{21}^{(n)^2} + \lambda_2 g_{11}^{(n)^2} + \lambda_2 b_{11}^{(n)^2} + \lambda_2 g_{33}^{(n)^2} + \lambda_2 g_{21}^{(n)^2} + 1 \\
& + \lambda_2 b_{12}^{(n)^2} + \lambda_2 g_{12}^{(n)^2} + \lambda_2 b_{32}^{(n)^2} + \lambda_2 g_{32}^{(n)^2} + \lambda_2 b_{13}^{(n)^2} \\
& + \lambda_2 g_{13}^{(n)^2} + \lambda_2 b_{23}^{(n)^2} + \lambda_2 g_{23}^{(n)^2} + \lambda_2 b_{33}^{(n)^2} + 13\lambda_1).
\end{aligned}
$$

Similar expressions were derived for the green and blue

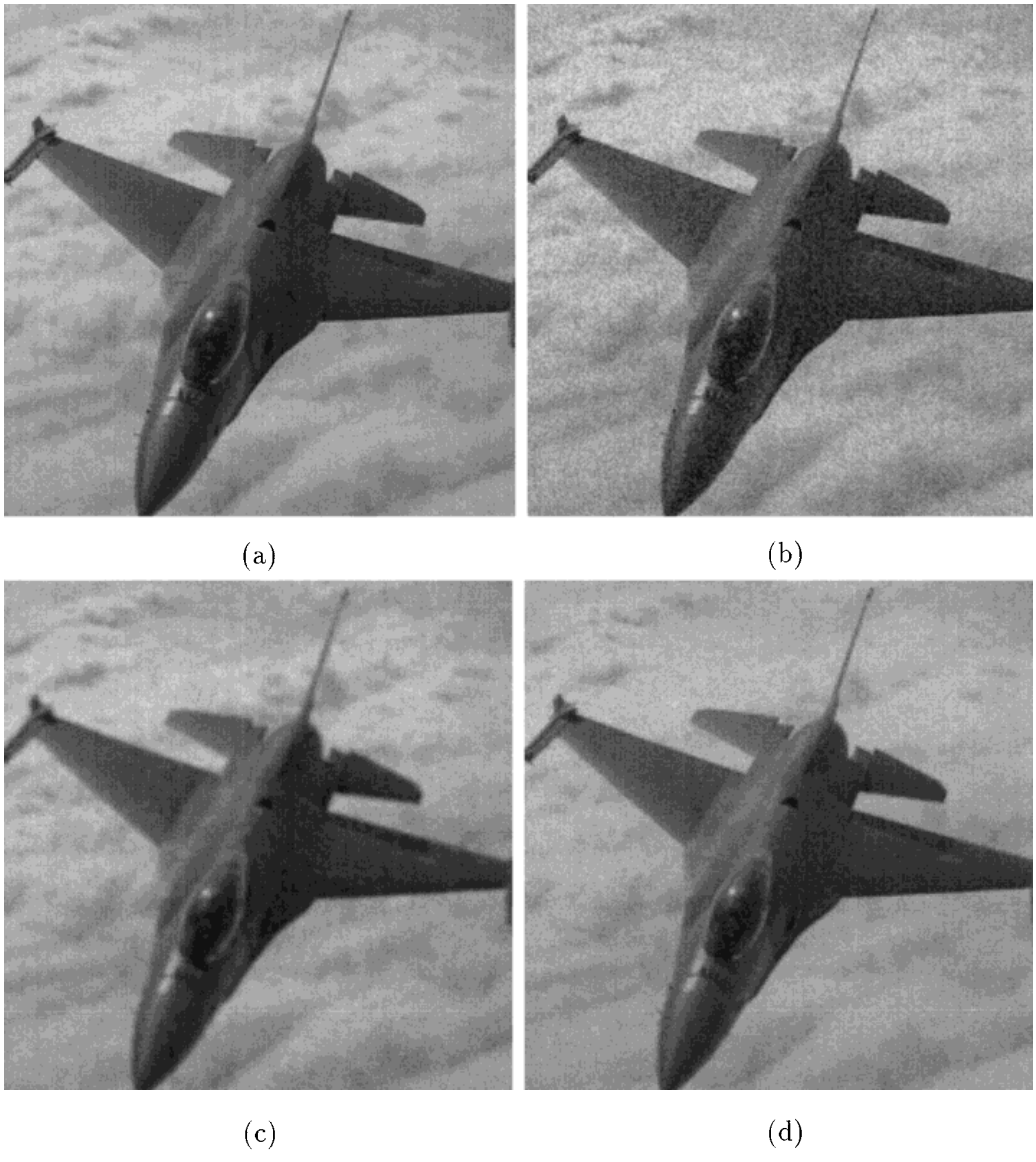


FIG. 4. (a) Red channel of original airplane image; (b) after addition of noise; (c) result of separate regularization in every channel; (d) result after incorporating the color correlation term.

components. The iterative scheme can be stopped when there is no significant change between successive iterations; for the images we tested, this was usually after 20 iterations or so.

5.2. *Choosing the Hyperparameters*

Let $\hat{F}_{\lambda_1,\lambda_2}$ denote the restoration obtained using the aforementioned iterative scheme, with λ_1, λ_2 as hyperparameters. The first step in the process of choosing the optimal hyperparameters is the construction of the curve in $\lambda_1 - \lambda_2$ space, for which $E(\|\hat{F}_{\lambda_1,\lambda_2} - H\|^2) = \sigma^2$. This proceeds as follows: first, the λ_1 for which $E(\|\hat{F}_{\lambda_1,0} - H\|^2) = \sigma^2$ is determined. The most straightforward way to do this is by binary search over $E(\|\hat{F}_{\lambda_1,0} - H\|^2)$; denote this λ_1 by λ_1^0. Next, for a fixed m (we have used $m = 30$), the values $i\lambda_1^0/m$, for $0 \le i < m$, are each chosen as λ_1, and, for each of them, the value of λ_2 satisfying $E(\|\hat{F}_{\lambda_1,\lambda_2} - H\|^2) = \sigma^2$ is found by binary search. The set of m points thus obtained defines the sought curve in $\lambda_1 - \lambda_2$ space (see Figs. 7a and 7c). We have not succeeded in deriving an analytic form for this curve, and it is not clear if it exists at all (even choosing a single λ using the OCV or GCV criterion is a non-trivial numerical task; moreover, the color correlation term is non-quadratic, and therefore difficult to analyze).

After the m points on the $\lambda_1 - \lambda_2$ curve are found, the OCV function is computed at each. The one yielding the

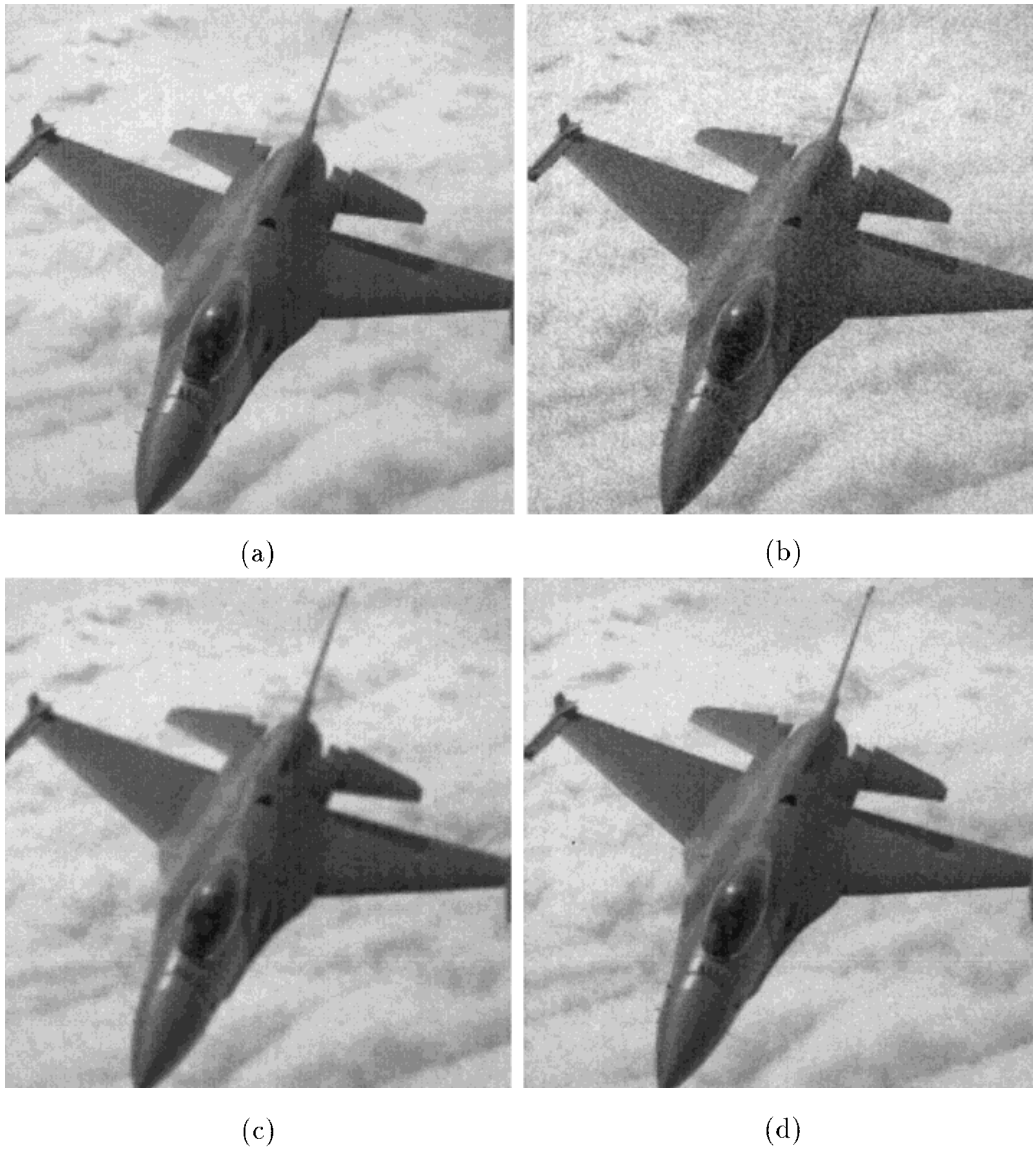


FIG. 5. Same as Fig. 4, for green channel.

minimal value is chosen as the optimal pair of hyperparameters used to restore the color image. In the case we tested, there was a rather good correlation between the OCV function and the mean square error; that is, the chosen hyperparameters resulted in a restoration which gave nearly the minimal mean square error, among the pairs which satisfy $E(\|\hat{F}_{\lambda_1,\lambda_2} - H\|^2) = \sigma^2$ (see Figs. 7b and 7d).

A "quick and dirty" heuristic for a good value of the hyperparameters, is to choose λ_1 as $\lambda_1^0/3$, and then to choose λ_2 satisfying $E(\|\hat{F}_{\lambda_1,\lambda_2} - H\|^2) = \sigma^2$. In all our experiments, this gave a value not far from the optimal one in terms of the mean square restoration error.

5.3. *Some Results*

Results are presented for two color images, one of a tiger's face and one of an airplane. The two images display rather different behavior in the frequency domain (the tiger image has much more energy in the high frequencies). The image of the tiger was corrupted by noise with $\sigma^2 = 1200$, the airplane with $\sigma^2 = 33$ (these are for the noise at each channel; from here on we shall refer to the mean square error summed over the three channels).

In Fig. 7a, the curve in hyperparameter space which satisfies $E(\|\hat{F}_{\lambda_1,\lambda_2} - H\|^2) = \sigma^2$ for the tiger image is shown, together with the curve showing the OCV function and the mean square error (y-axis) versus λ_1 (x-axis) (Fig. 7b). The optimal parameters (that is, minimal value of the OCV function) selected were $\lambda_1 = 0.06$ and $\lambda_2 = 0.00023$. Note the good correlation between the OCV and mean square error function; the chosen hyperparameters result in a value of the mean square error which is very close to the optimal one.

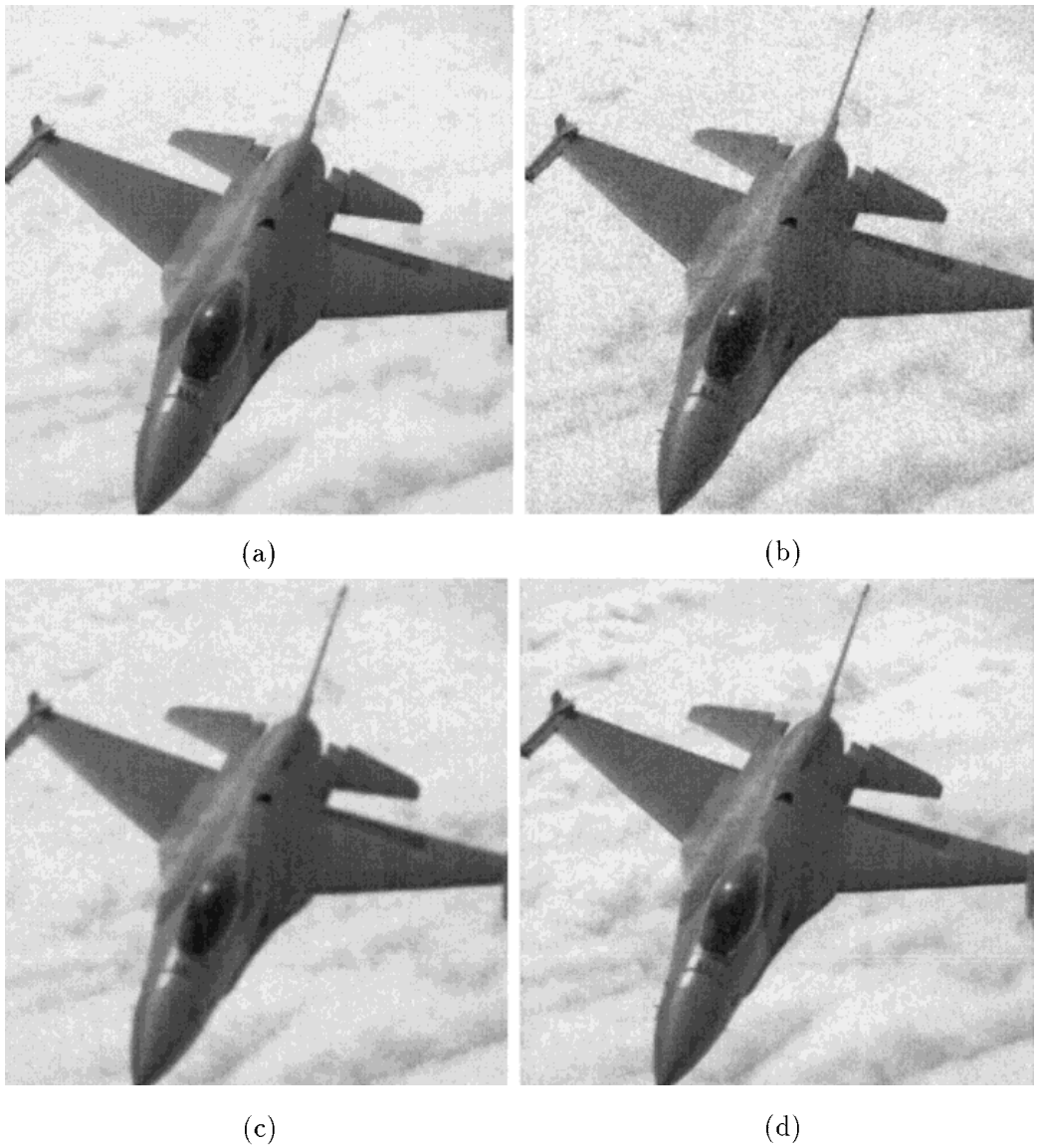


FIG. 6. Same as Fig. 4, for blue channel.

In Figs. 7c and 7d, the $\lambda_1 - \lambda_2$ and mean square error and OCV curves are shown for the airplane image. The optimal values were $\lambda_1 = 0.12$ and $\lambda_2 = 0.00005$.

The restoration results are presented in Figs. 1–6 and 8. For comparison, we show the original image, the noised image, the optimal restoration result without the color correlation term (that is, $\lambda_2 = 0$, and λ_1 the one satisfying $E(\|\hat{F}_{\lambda_1,0} - H\|^2) = \sigma^2$), and the restoration with the optimal hyperparameters. The advantage in incorporating the color correlation term is obvious, both in terms of the mean square error (a reduction of 18.6% for the tiger image and 40.4% for the airplane image; the difference is probably due to the fact that the tiger image is harder to restore because of its substantial high frequency content), and in the appearance of the images: the resulting images when the color correlation term was incorporated are much sharper. Intuitively, this is because a smaller λ_1 is used, which results in a smaller penalty on image sharpness; however, if $\lambda_2 = 0$, these small values of λ_1 will result in a very noisy restoration. The color correlation term reduces the noise by forcing a "good" correlation between the channels.

In Fig. 8, we show, for the tiger image, the same set of four images (but in color), together with the result obtained when using the surface area of the color image in fifth-dimensional Euclidean space as a quality measure (see Sections 1.1 and 7).

6. CONCLUSIONS AND FURTHER RESEARCH

A novel method for denoising color images was presented. The two suggested contributions are the incorpora-

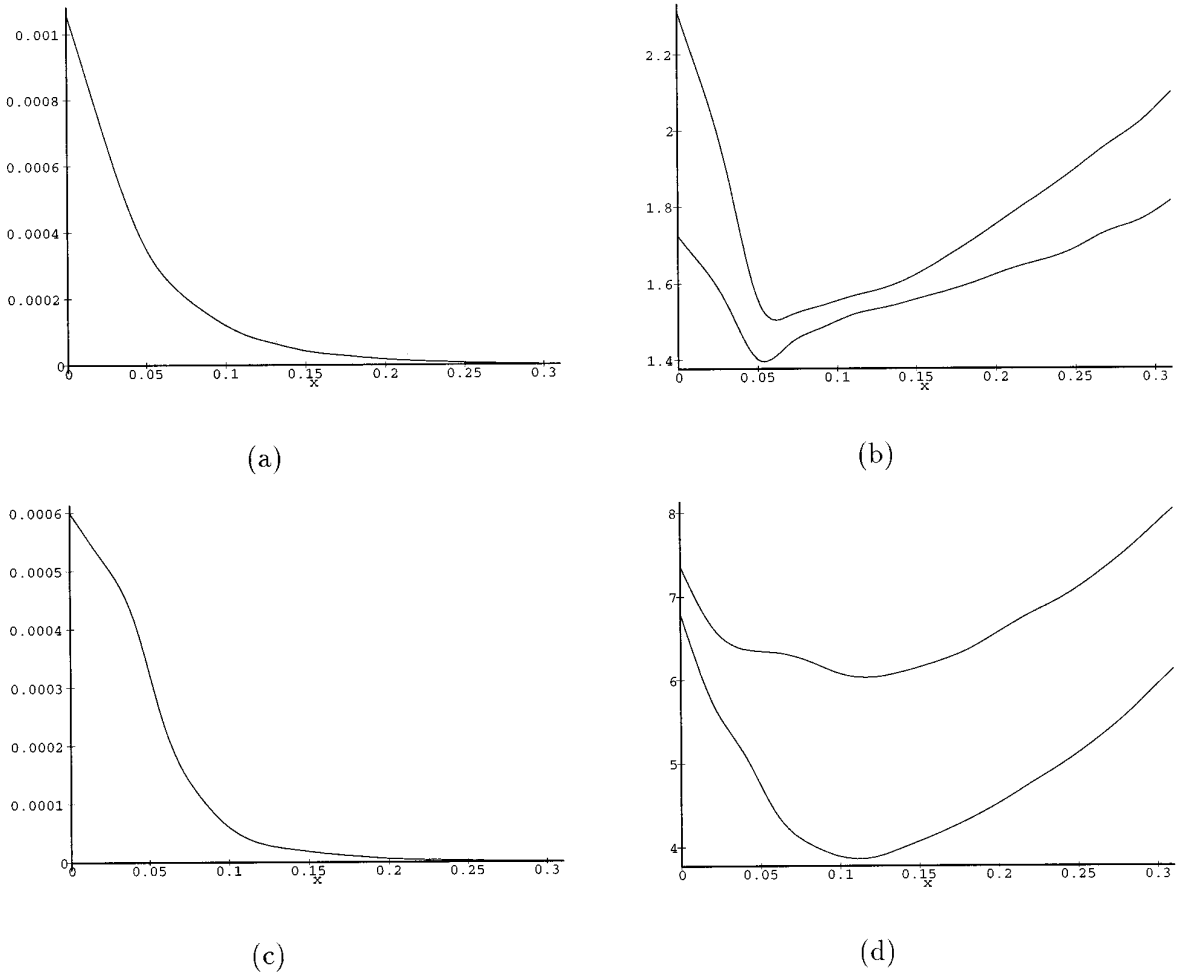


FIG. 7. (a) The curve in $\lambda_1 - \lambda_2$ space, for tiger image, for which $E(\|\hat{F}_{\lambda_1,\lambda_2} - H\|^2) = \sigma^2$ (the x-axis represents λ_1, y-axis represents λ_2). (b) The mean square error (summed over the RGB channels) for the curve in (a), normalized by 0.001 (lower), and the OCV function curve (upper). The x-axis represents λ_1. (c) and (d) are the same as (a) and (b), respectively, for the airplane image. In (d) the mean square error is normalized by 0.1. For both images, the OCV function attains its minimum for a (λ_1, λ_2) pair very close to the one for which the mean square error is minimal.

tion of "color correlation terms," two of which were investigated, and a method for choosing two regularization hyperparameters, by incorporating two previous methods for choosing a single hyperparameter. A simple iterative scheme was then derived for the denoising problem. Results were presented for two color images, which represent typical results for the method described here.

In the future, we hope to address the following issues:

- Search for a more efficient method to determine the two optimal hyperparameters.
- How to proceed when the noise variance is not known? In that case, minimizing the OCV function may yield a curve of possible solutions in hyperparameter space, but it is not clear which point on that curve should be used.
- Apply the paradigm offered here to the more difficult problem of restoring a blurred and noised color image.

APPENDIX: THE "GEOMETRIC MEASURE" AND SURFACE AREA

In this section, we first show how "standard" regularization can be seen in the context of reducing the area of a restored gray level image, when it is viewed as a surface in $\mathscr{R}^3$. Then, it is shown that when the idea of reducing the surface area is extended to a color image, the resulting expression bears a resemblance to the "geometric measure" approach presented in this work. A short comparison of the methods follows.

Let $F(x, y)$ be a gray level image, or any single-channel signal for that matter. Some times, instead of the second-order smoothness term used here, $\iint (F_{xx}^2 + 2F_{xy}^2 + F_{yy}^2)\, dx\, dy$, the simpler first-order term, $\iint (F_x^2 + F_y^2)\, dx\, dy$, is used to measure the image's smoothness. It is well known from elementary calculus that the area of the graph of F

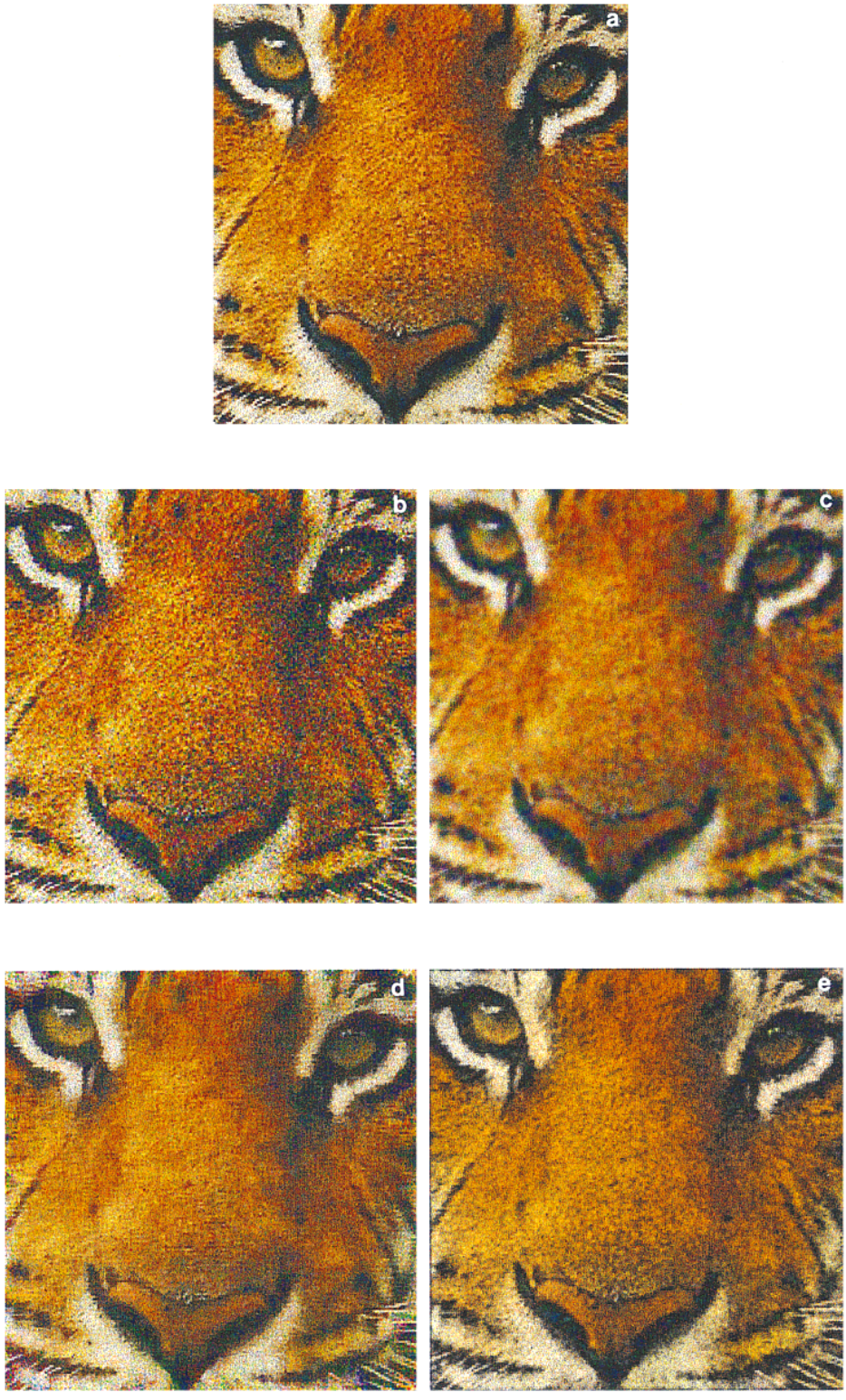


FIG. 8. (a) Original color image; (b) after adding noise; (c) result of optimal denoising in each channel separately; (d) result of optimal denoising using surface area in $\mathscr{R}^5$ as quality measure; (e) result of optimal denoising using the geometric correlation term suggested in this work.

(that is, the set of triplets $\{(x, y, F(x, y))\}$, when viewed as a surface in $\mathscr{R}^3$), equals $\iint\sqrt{1 + F_x^2 + F_y^2}\, dx\, dy$; thus, the smoothness is related to the area of the graph of the image, when viewed as a function from $\mathscr{R}^2$ to $\mathscr{R}$. It is interesting to note that this expression is related to $\iint (|F_x| + |F_y|)\, dx\, dy$, which is used in *total variation* methods, that were also applied to restore color images [1].

This notion of surface area can be extended to a color image $(R(x, y), G(x, y), B(x, y))$ [13, 14]. Its graph, the set of quintuplets $\{(x, y, R(x, y), G(x, y), B(x, y))\}$, is a surface in $\mathscr{R}^5$. Its area can be determined using tools from differential geometry, namely, the first fundamental form [2]. The area turns out to be

$$\iint (1 + R_y^2 + G_y^2 + B_y^2 + R_x^2 + R_x^2 G_y^2 + R_x^2 B_y^2 + G_x^2 + G_x^2 R_y^2 + G_x^2 B_y^2 + B_x^2 + B_x^2 R_y^2 + B_x^2 G_y^2 - 2R_x R_y G_x G_y - 2R_x R_y B_x B_y - 2G_x G_y B_x B_y)^{1/2}\, dx\, dy.$$

Let us denote the expression in the square root of the integrand by A_{element} (for "area element"). It can be verified by a straightforward computation that

$$A_{\text{element}} = 1 + R_x^2 + R_y^2 + G_x^2 + G_y^2 + B_x^2 + B_y^2 + \|(R_x, G_x, B_x) \times (R_y, G_y, B_y)\|^2.$$

Hence, the integrand in the expression for the area of the color image, viewed as a surface in $\mathscr{R}^5$, is the square root of A_{element}—an expression consisting of a spatial smoothness term $R_x^2 + R_y^2 + G_x^2 + G_y^2 + B_x^2 + B_y^2$, and the squared norm of the vector product between the two components of the gradient. Since, in a discrete image, the gradient is approximated by differences between the RGB values of adjacent pixels, A_{element} is resemblant to the expression used in this work—a (weighted) sum of a spatial smoothness term and the squared norms of vector products between adjacent pixels. One difference is that we have used a second-order smoothness term, while the expression for surface area contains a first-order term.

One can then formalize the following restoration problem: minimize $\|F - H\|^2 + \lambda A(F)$, where H is the measured color image, F is the sought color image, and $A(F)$ is the area of F. We have implemented an iterative scheme minimizing this functional and applied it to some images, including those of the tiger and airplane (see Section 6). The value of λ was chosen so that the criterion $E(\|\hat{F} - H\|^2) = \sigma^2$ holds, where $\hat{F}$ is the solution of the minimization problem. It should, however, be noted that we could choose a λ based on the cross validation method, even if the noise variance was not known; this is the advantage of using only one hyperparameter. In [14], the "Beltrami flow" leading towards a minimal surface was "manually stopped." However one may simply stop the flow when the condition $E(\|\hat{F} - H\|^2) = \sigma^2$ holds; so, the duration of time during which the flow is applied is the analogue of the regularization constant λ. See also [17] for other work on diffusion in color space.

On the examples we tested, the approach based on minimizing the surface area gave results which were, in general, pleasing to the eye; however, the mean square error was larger than the error when using the correlation term with vector products: for the airplane image the optimal λ chosen was 1.52; and the mean square error was 58.8 (47% higher than for the result presented in Section 6). For the tiger image, the λ chosen was 0.82, and the mean square error 2660 (89% higher than for the result presented in Section 6). This may well be due to the fact that the first-order smoothness term is not a good measure for natural images—the second-order term is far better [12]. This problem is more acute in textured regions: see results for the tiger image (Fig. 8). The area-based approach does not provide a good restoration in highly textured areas. It gave better results for the airplane image.

However, the area-based approach has an important advantage—it can be implemented even when the noise variance is not known. Also, as noted, it usually produces results which are pleasing to the eye; and it can be argued that, depending on the client for the restored image, this is a criterion no less important than the mean square error. The search for criterions to image quality and similarity is an exciting one, which is guaranteed to occupy researchers in the vision community for a long time.

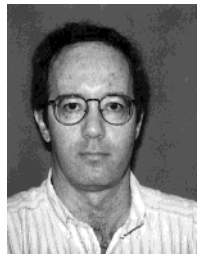
ACKNOWLEDGMENTS

The first author is grateful to the anonymous reviewers of a research proposal for their helpful and instructive comments, as well as for referencing work in the field of multichannel restoration. We thank Nir Sochen and Ron Kimmel for their comments on the Beltrami flow and its application to image restoration, and Guillermo Sapiro for discussions on diffusion in color space.

REFERENCES

1. P. Blomgren and T. F. Chan, *Color TV: Total Variation Methods for Restoration of Vector Valued Images*, technical report CAM96-5 UCLA, 1996. *IEEE Trans. Image Process*, to appear].
2. M. P. Do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliff, NJ, 1976.
3. P. Craven and G. Whaba, Optimal smoothing of noisy data with spline functions, *Numer. Math.* **31,** 1979, 377–403.
4. N. P. Galatsanos and R. T. Chin, Restoration of color images by multichannel Kalman filtering, *IEEE Trans. Signal Process.* **39**(10), 1991, 2237–2252.
5. N. P. Galatsanos, A. T. Katsagellos, R. T. Chin, and A. D. Hillery, Least squares restoration of multichannel images, *IEEE Trans. Acoustic Speech Signal Process.* **39**(10), 1991, 2222–2236.
6. S. Geman and D. Geman, Stochastic relaxation, gibbs distribution, and the bayesian restorat ion of images, *IEEE Trans. Pattern Anal. Mach. Intell.* **6,** 1984, 721–741.

7. P. Hall and I. Johnstone, Empirical functionals and efficient smoothing parameter selection, *J. R. Statist. Soc.* **54**(1), 1992, 475–530.
8. B. K. Horn and B. G. Schunck, Determining optical flow, *Artif. Intell.* **17,** 1981, 185–203.
9. M. G. Kang and A. K. Katsaggelos, General choice of the regularization functional in regularized image restoration, *IEEE Trans. Image Process.* **4,** 1995, 594–602.
10. M. G. Kang and A. K. Katsaggelos, Simultaneous multichannel image restoration and estimation of the regularization parameters, *IEEE Trans. Image Process.* **6**(5), 1997, 774–778.
11. A. K. Katsaggelos, K. T. Lay, and N. P. Galatsanos, A general framework for frequency domain multi-channel signal processing, *IEEE Trans. Image Process*. **2**(3), 1993, 417–420.
12. D. Keren and M. Werman, Probabilistic analysis of regularization, *IEEE Trans. Pattern Anal. Mach. Intell.* **15,** 1993, 982–995.
13. R. Kimmel, A natural norm for color processing, in *3rd Asian Conference on Computer Vision, Hong Kong, January 8–11, 1998*.
14. R. Kimmel, R. Malladi, and N. Sochen, Images as embedding maps and minimal surfaces: Movies, color, and volumetric medical images, in *IEEE CVPR, Puerto Rico, June 1997*.
15. D. Nychka, Choosing a range for the amount of smoothing in nonparametric regression, *J. Am. Statist. Assoc.* **86**(415), 1991, 653–664.
16. M. Ozkan, A. Erdem, M. I. Sezan, and A. M. Tekalp, Efficient multiframe Wiener filter restoration of blurred and noisy image sequences, *IEEE Trans. Image Process.* **4,** 1992, 453–476.
17. G. Sapiro and D. L. Ringach, Anisotropic diffusion of multivalued images with applications to color filtering, *IEEE Trans. Image Process.* **5,** 1996, 1582–1586.
18. R. R. Schultz and R. L. Stevenson, Stochastic modeling and estimation of multispectral image data, *IEEE Trans. Image Process.* **4**(8), 1995, 1109–1119.
19. J. Skilling, Fundamentals of maxent in data analysis, in *Maximum entropy in Action* (B. Buck and V. A. Macaulay, Eds.), Clarendon Press, Oxford, 1991.
20. R. Szeliski, *Bayesian Modeling of Uncertainty in Low-Level Vision*, Kluwer, Dordrecht, 1989.
21. D. Terzopoulos, Multi-level surface reconstruction, in *Multiresolution Image Processing and Analysis* (A. Rosenfeld, Ed.), Springer-Verlag, New York/Berlin, 1984.
22. A. M. Thompson, J. C. Brown, J. W. Kay, and D. M. Titterington, A study of methods of choosing the smoothing parameter in image restoration by regularization, *IEEE Trans. Pattern Anal. Mach. Intell.* **13,** (1991), 326–339.
23. G. Wahba, *Spline Models for Observational Data*, Society for Industrial and Applied Mathematics, Philadelphia, 1990.
24. W. Zhu, N. P. Galatsanos, and A. K. Katsaggelos, Regularized multichannel restoration using cross-validation, *Graph. Models Image Process.* **57**(1), 1995, 38–54.



DANIEL KEREN completed a Ph.D. in the field of computer vision at the Institute of Computer Science of the Hebrew University of Jerusalem, Israel. After that, he was a postdoctoral fellow at the Division of Engineering, Brown University, Providence, Rhode Island. Since 1994, he has taught at the Department of Computer Science of the University of Haifa, Israel. Dr. Keren's interests are mainly in regularization theory, invariants, and the application of implicit functions to describe free-form objects.

ANNA GOTLIB completed her B.Sc. (suma cum laude) in the Department of Computer Science of the University of Haifa in Israel, in 1996. Currently she is a graduate student there. Her main area of research is regularization theory and its applications to image and signal processing.