PERCEPTION BASED ADAPTIVE IMAGE RESTORATION

Stuart W. Perry and Ling Guan

Department of Electrical Engineering
University of Sydney
Sydney, NSW 2006.
Australia

ABSTRACT

This paper presents an image restoration technique which uses a cost function based on a novel image error measure. The cost function presented here takes into account local statistical information of the image when performing restoration. It is shown that this technique compares favourably with other techniques, especially when applied to colour images.

1. INTRODUCTION

The restoration of an image degraded in some way is usually approached in terms of a multidimensional optimisation problem. Theoretically, a degraded image may be restored by the creation of an algorithm which minimises a measure of image quality such as Mean Square Error (MSE). Developing such an algorithm can be difficult if the nature of the image degradation is complicated or not completely known. When degraded images are being filtered for the purposes of image segmentation, then the usefulness of the segmentation obtained from the filtered image provides a measure of quality. However, often an image is filtered for the purpose of greater visual quality or clarity as perceived by humans, such as an old photograph or a television transmission [1]. In these cases a restoration algorithm is attempting to use mathematics to produce an image which human beings will find visually pleasing. For an algorithm to produce an image which humans will find pleasant it must posses a method of quantifying an image's quality which takes into account human visual preferences.

Classical image error measures such as mean square error or Signal to Noise Ratio (SNR) compare images on a pixel to pixel basis, and in effect make statements about the power of the noise signal created by the subtraction of the two images to be compared. This kind of information is mathematically useful. However these measures favour slow variations in the image and bear very little relationship to the manner in which humans view the differences between two images. Humans tend to pay more attention to sharp differences in intensity within an image, for example edges or noise in background regions. Hence an error measure should take into account the concept that low variance regions in the original image should remain low variance regions in the enhanced image, and high variance regions in the original image should likewise remain high variance regions in the enhanced image. This implies that noise should be at a minimum in background regions, where it is most noticable, but noise suppression should not be as important in highly textured regions where image sharpness should be the dominate consideration. In this paper we present a novel image error measure which attempts to quantify the statistical differences between regions in an image rather than the differences between individual pixels. This image error measure is incorporated into a restoration cost function which is used to restore degraded colour images.

2. THE BASIC RESTORATION MODEL

Consider an N by M input image, and a linear image degradation model described by the equation [1,2,3]:

$$\boldsymbol{g} = \boldsymbol{H}\boldsymbol{f} + \boldsymbol{n} \qquad (1)$$

where $\boldsymbol{f}$ and $\boldsymbol{g}$ are lexicographically organised original and degraded image vectors of length NM, respectively, $\boldsymbol{H}$ is a matrix distortion operator and $\boldsymbol{n}$ is an additive noise vector. In the case of a colour image, (1) can be assumed to be applied to each colour plane separately. Vectors $\boldsymbol{f}$ and $\boldsymbol{g}$ are created by either row or column scanning the image. The matrix $\boldsymbol{H}$ is an arrangement of the elements in the degrading point spread function (PSF) such that equation (1) holds. In the case of spatially invariant distortion, $\boldsymbol{H}$ has a Block-Toeplitz form. When attempting to deconvolve a distorted image, one method is to minimise an error measure such as the constrained least square error function:

$$SE = \|\boldsymbol{g} - \boldsymbol{H}\hat{\boldsymbol{f}}\|^2 + \left(\frac{1}{2}\right)\lambda\|\boldsymbol{D}\hat{\boldsymbol{f}}\|^2 \qquad (2)$$

where $\hat{\boldsymbol{f}}$ is the restored image estimate and λ is the constraint factor. The matrix $\boldsymbol{D}$ is of the same nature as $\boldsymbol{H}$ and acts as a high pass filter. The first term in the equation is minimised when $\hat{\boldsymbol{f}}$ is equal to the original image whereas the second term in (2) increases in the presence of noise and is minimised for a smooth image estimate. Hunt and Kubler [2] showed that in general information in the different colour planes will be correlated. However, in this investigation we assume that the correlations are small enough that each plane may be restored independently and then brought together for a final result. Equation (2) applies the same λ value and $\boldsymbol{D}$ matrix to all pixels in the image and hence a value of λ large enough to fully suppress noise in the smooth regions of the image may blur fine details. A great deal of research has been done into algorithms which minimise (2), [3,4]. However as stated above, (2) does not take into account the human visual system, hence a modified cost function would be advantageous.

3. MODIFIED COST FUNCTION

The measurement we introduce is termed Local Standard Deviation Mean Square Error (LSMSE). LSMSE is calculated by comparing the local standard deviations in the neighbourhood of each pixel in the images we wish to compare to each other. The mean square error between these two standard deviations gives an indication of the degree of similarity between the two images.

Define the local standard deviation in an A by A neighbourhood of pixel (x, y) in image $\boldsymbol{f}$ as:

$$\sigma_A(\boldsymbol{f}(x,y)) = \sqrt{\sum_{i=x-\frac{A}{2}}^{x+\frac{A}{2}} \sum_{j=y-\frac{A}{2}}^{y+\frac{A}{2}} \frac{\left(\boldsymbol{f}(i,j) - M_A(\boldsymbol{f}(x,y))\right)^2}{A^2}} \quad (3)$$

where the local mean of the A by A neighbourhood of pixel $\boldsymbol{f}(x, y)$ in image $\boldsymbol{f}$ is defined as:

$$M_A(\boldsymbol{f}(x,y)) = \sum_{i=x-\frac{A}{2}}^{x+\frac{A}{2}} \sum_{j=y-\frac{A}{2}}^{y+\frac{A}{2}} \frac{\boldsymbol{f}(i,j)}{A^2} \quad (4)$$

Using the above conventions, we define the LSMSE between two NxM images $\boldsymbol{f}$ and $\boldsymbol{g}$ as:

$$\text{LSMSE}_A(\boldsymbol{f},\boldsymbol{g}) = \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \frac{\left(\sigma_A(\boldsymbol{f}(x,y)) - \sigma_A(\boldsymbol{g}(x,y))\right)^2}{NM} \quad (5)$$

LSMSE in effect requires the matching of homogeneous statistical regions between the two images to be compared. Hence background regions should remain as noise free as possible and highly textured regions should not be smoothed by the enhancement procedure. LSMSE by itself can be used as a measure of image quality, however the authors wished to investigate modifying (2) to incorporate a LSMSE-like term.

The new cost function we suggest is given by:

$$E = \|\boldsymbol{g} - \boldsymbol{H}\hat{\boldsymbol{f}}\|^2 + \lambda\|\boldsymbol{D}\hat{\boldsymbol{f}}\|^2 + \theta \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \frac{\left(\sigma_A^2(\hat{\boldsymbol{f}}(x,y)) - \sigma_A^2(\boldsymbol{g}(x,y))^*\right)^2}{NM} \quad (6)$$

where $\sigma_A^2(\hat{\boldsymbol{f}}(x,y))$ is the variance of the local region surrounding pixel (x,y) in the image estimate and $\sigma_A^2(\boldsymbol{g}(x,y))^*$ is the variance of the local region surrounding pixel (x,y) in the degraded image scaled to predict the variance in the original image. Note that local variances are used in the cost function rather than local standard deviations as this simplifies the minimisation of the cost function. If the degraded image has been blurred then image variances in $\boldsymbol{g}$ will be lower than the corresponding variances in the original image. In this case the variances $\sigma_A^2(\boldsymbol{g}(x,y))^*$ would be scaled larger than $\sigma_A^2(\boldsymbol{g}(x,y))$ to reflect the decrease in variance due the blurring function. In general, if we consider an image degraded by a process which is modelled by (1), then we find that a suitable approximation is:

$$\sigma_A^2(\boldsymbol{g}(x,y))^* = K(x,y)\left(\sigma_A^2(\boldsymbol{g}(x,y)) - J(x,y)\right) \quad (7)$$

where $J(x, y)$ is a function of the noise added to the degraded image at point (x, y) and $K(x, y)$ is a function of the degrading point spread function at point (x, y). Although it may appear difficult to accurately determine the optimal values of $K(x, y)$, the algorithm is extremely tolerant of variations in this factor and only a rough guess is required. This is due to the fact that the LSMSE term in (6) has it's greatest effect on the low variance background regions of the image where noise is the most noticable. In such regions $\sigma_A^2(\boldsymbol{g}(x,y))^*$ merely has to be small to suppress noise since details are not noticable. In highly textured regions of the image where the preservation of image details are most important, the LSMSE term requires that the variance of the region be large, and the first two terms of (6) ensure the sharpness and accuracy of the image features.

The first step in the development of the algorithm is a change in notation. For an N by M image let $\boldsymbol{f}$ represent the lexicographically organized image vector of length NM as per the model given by (1). Consider setting all pixels in vector $\boldsymbol{f}$ to zero except those which are in the two dimensional A by A neighbourhood of pixel i. We denote this vector $\boldsymbol{f}^i$. Using this notation the average pixel value in the A by A region surrounding pixel i is given by:

$$M_A(i) = \frac{1}{A^2} \sum_{j=1}^{NM} \boldsymbol{f}_j^i$$

Let $\beta_i = \sum_{j=1}^{NM} \left(\boldsymbol{f}_j^i\right)^2$ and $\gamma_i = \sum_{j}^{NM} \boldsymbol{f}_j^i$ then the variance of the A by A region surrounding pixel i is given by:

$$V^i = \frac{\beta_i}{A^2} - \frac{\gamma_i^2}{A^4} \quad (8)$$

The LVMSE (Local Variance Mean Square Error) between the image estimate, $\hat{\boldsymbol{f}}$, and the original image, $\boldsymbol{f}$, may then be written as:

$$\text{LVMSE}(\hat{\boldsymbol{f}},\boldsymbol{f}) = \frac{1}{NM} \sum_{i=1}^{NM} \left(V^i(\hat{\boldsymbol{f}}) - V^i(\boldsymbol{f})\right)^2 \quad (9)$$

Let $V^i(\boldsymbol{f})$ be denoted by $V\!f^i$. $V\!f^i$ is the estimate of the local variance of pixel i in the original image based on the degraded image and knowledge of the degrading point spread function as per equation (7). $V\!f^i$ is calculated before the algorithm commences and remains a constant throughout the restoration procedure.

To optimise (6) we require an equation for the gradient of (6) as well one for the change in energy given a change in neuron state. Without the LSMSE term in (6) these equations are obtained easily by using a neural network approach and have been derived by Paik and Katsaggelos [3]. For each of these equations an extra term must be added to take into account LSMSE. We derive these extra terms below.

The gradient of (9) is given by:

$$\frac{\partial}{\partial \hat{f}_i}\mathrm{LVMSE} = \frac{2}{NM}\left(V^i(\hat{\boldsymbol{f}}) - Vf^i\right)\frac{\partial}{\partial \hat{f}_i}\left(V^i(\hat{\boldsymbol{f}}) - Vf^i\right) \quad (10)$$

Note that this formula is an approximation of the gradient which ignores the contributions of the local variances of the pixels adjacent to *i* to the overall LVMSE of the image.

$$\frac{\partial V^i}{\partial \hat{f}_i} = \frac{2\hat{f}_i}{A^2} - \frac{2\gamma_i}{A^4} \quad (11)$$

Taking note of the fact that $\hat{f}^i_i = \hat{f}_i$. Substituting (11) into (10) we obtain:

$$\frac{\partial}{\partial \hat{f}_i}\mathrm{LVMSE} = \frac{2}{NM}\left(\frac{\beta_i}{A^2} - \frac{\gamma_i^2}{A^4}\right)\left(\frac{2\hat{f}_i}{A^2} - \frac{2\gamma_i}{A^4}\right)$$

Therefore

$$\frac{\partial}{\partial \hat{f}_i}\mathrm{LVMSE} = \frac{4}{NMA^2}\left(\frac{\hat{f}_i\beta_i}{A^2} - \frac{\hat{f}_i\gamma_i^2}{A^4} - \hat{f}_i Vf_i - \frac{\beta_i\gamma_i}{A^4} + \frac{\gamma_i^3}{A^6} + \frac{\gamma_i Vf_i}{A^2}\right) \quad (12)$$

Given a change in the value of pixel *i*, the resultant change in LVMSE is given by:

$$\Delta\mathrm{LVMSE} = \frac{1}{NM}\left(\left(V^i_{new} - Vf^i\right)^2 - \left(V^i_{old} - Vf^i\right)^2\right) \quad (13)$$

The algorithm first computes the negative direction of the gradient using (12) which gives an indication of whether increasing or decreasing the current neurons value will result in a net decrease in energy. Once the negative gradient is found the neurons value is changed in unit steps and the resultant energy decrease after each step is computed with the aid of (13). This ends when no further energy minimisation is possible. A further problem to be overcome is the fact that the third term in (6) is not quadratic in nature. When the local variance in the image estimate is much lower than the projected local variances of the original image the LSMSE term becomes large and may force the pixel values to an extreme of the range of acceptable values in order to create a high variance region. The LSMSE term should never completely dominate over the first term in (6) since the LSMSE term only attempts to match regions, not pixels and fine structure within the region will be lost. To remedy this situation the pixel values are not allowed to change by more than a set amount per iteration. This method appears to work well in practice and the pixel values converge to a solution after a finite number of iterations.

4. RESULTS

For this experiment colour images were used consisting of three colour planes, red, green, and blue. The images in this paper have been converted to grayscale to comply with ICASSP requirements. The image was degraded by a 5 by 5 Gaussian PSF of standard deviation 2.0 applied to each of the colour planes. In addition additive noise of variance 369.31 was also added to each colour plane. Figure 1 shows the original image and Figure 2 shows the degraded image. The degraded image has a SNR of 19.81dB and a LSMSE of 313.05. The SNR was calculated by adding together the signal to noise ratio of each colour plane:

$$SNR = 20\log\left(\frac{\sigma_o^r}{\sigma_n^r} + \frac{\sigma_o^g}{\sigma_n^g} + \frac{\sigma_o^b}{\sigma_n^b}\right) \quad (14)$$

Similarly the LSMSE for the entire image was calculated by summing the LSMSEs of each colour plane. A 9 by 9 neighbourhood was used for calculating the local variance. We compared our algorithm with a Wiener filter approach. In this investigation, we assumed that each colour plane in our test image does not have a high level of correlation and so applied a Wiener filter to each colour plane separately. The Wiener restored image is shown in Figure 3 and has a SNR of 16.65 dB and a LSMSE of 859.80. The image was also restored using the neural network algorithm without the LSMSE term. A constraint factor of $\lambda = 0.001$ was chosen. The non-LSMSE restored image is shown in Figure 4 and has a SNR of 17.26 dB and a LSMSE of 634.04. The same degraded image was also restored using the LSMSE modified cost function. In the LSMSE modified cost function the value of λ was set to 0.0005. The factor θ was set to be 0.00001 and the image local variance estimate was computed as:

$$\sigma_A^2(\boldsymbol{g}(x,y))^* = 2\left(\sigma_A^2(\boldsymbol{g}(x,y)) - 200\right)$$

This image is shown in Figure 5 and has a SNR of 19.89 dB and a LSMSE of 180.81. Although the grayscale images in this paper do not demonstrate this algorithms results as clearly as the colour images, the important features of the colour images can still be discerned. By visual observation it can be seen that Figure 5, produced by the LSMSE modified cost function, displays better noise suppression in background regions and is at the same time sharper than Figure 3 and Figure 4, produced by the Wiener and non-LSMSE neural network approaches. Figure 5 also displays a better SNR and LSMSE than Figures 3 and 4. Although the LSMSE restored image is visually closer to the original image than the degraded image, it’s SNR is only slightly higher than the degraded image. This is not surprising in view of the arguments above that SNR does not correspond well with human visual perception. However LSMSE does match with human observation assigns a much lower value to Figure 5.

5. SUMMARY

A novel error measure was introduced in this paper which compares two images by consideration of their regional statistical differences rather than their pixel-level differences. It was found that this error measure more closely corresponds to human visual perception of image quality. Using the new error measure a modified neural network cost function was developed. This cost function was shown to perform well when applied to colour images. A future research direction would be incorporating into the algorithm information about the correlation between the colour coordinates and comparing this with existing techniques such as the well known Wiener filter proposed by Hunt and Kubler [2].

6. REFERENCES

[1] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*. Addison-Wesley, 1992.

[2] B. R. Hunt, O. Kubler, “Karhunen-Loeve multispectral image restoration, Part I:Theory,” *IEEE Trans. Acoust., Speech, Signal Processing,* vol. ASSP-32, no. 3, pp. 595-600, June 1984.

[3] J. K. Paik and A. K. Katsaggelos, “Image restoration using a modified Hopfield network.,” *IEEE Trans. Image Proc.*, vol. 1, no. 1, pp. 49-63, Jan. 1992.

[4] S. W. Perry and L. Guan, “Image Restoration using a neural network with an Adaptive Constraint Factor.,” *Proc. IEEE-SMC. CESA’96 IMACS Multiconference, Symp. Control, Optimization and Supervision*, Lille, France, July 9-12, 1996, pp. 854-859.

Figure 1. Original image.

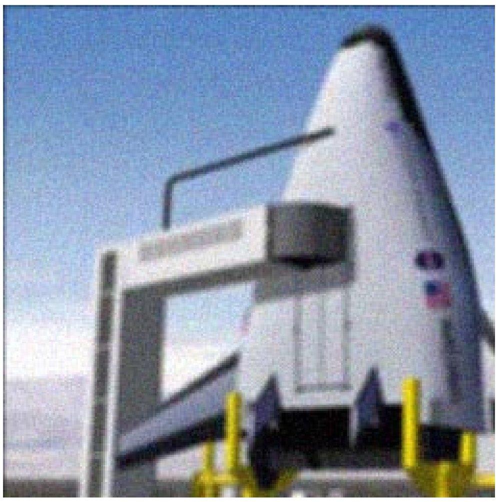

Figure 2. Degraded image.

Figure 3. Image restored using Wiener algorithm.

Figure 4. Image restored using non-LSMSE neural network approach

Figure 5. Image restored using LSMSE-based algorithm.