

Image Reconstruction from Contrast Information

Asim A. Khwaja, Roland Goecke
 Research School of Information Sciences & Engineering
 Australian National University, Canberra, Australia
 {asim.khwaja, roland.goecke}@anu.edu.au

Abstract

An iterative algorithm for the reconstruction of natural images given only their contrast map is presented. The solution is neuro-physiologically inspired, where the retinal cells, for the most part, transfer only the contrast information to the cortex, which at some stage performs reconstruction for perception. We provide an image reconstruction algorithm based on least squares error minimization using gradient descent as well as its corresponding Bayesian framework for the underlying problem. Starting from an initial image, we compute its contrast map using the Difference of Gaussians (DoG) operator at each iteration, which is then compared to the contrast map of the original image generating a contrast error map. This contrast map is processed by a non-linearity to deal with saturation effects. Pixel values are then updated proportionally to the resulting contrast errors. Using a least squares error measure, the result is a convex error surface with a single minimum, thus providing consistent convergence. Our experiments show that the algorithm's convergence is robust to initial conditions but not the performance. A good initial estimate results in faster convergence. Finally, an extension of the algorithm to colour images is presented. We test our algorithm on images from the COREL public image database. The paper provides a novel approach to manipulating an image in its contrast domain.

1 Introduction

Image reconstruction or restoration is an inverse problem which aims to recover the original image from its transformed version. It is an ill-conditioned problem at best and a singular problem at worst [1]. The transformation of the image could be the result of unwanted distortion like blurring, aliasing, or noise, or it could be a purposeful operation like compression, convolution, etc. Similarly, the transformation function may or may not be known. Furthermore, the transformation could be information-preserving or non-

preserving. Inverse problems are generally ill-posed in that they might not have a unique solution or a solution at all.

Colloquially, we often know that interesting information lies in the contrast. An image's contrast map contains within it everything that is needed for a faithful reconstruction. Here, we describe a technique which reconstructs an image given only its contrast map such that all the relative intensity values are restored. In addition, even the ambient levels get restored to their original values. But how is that possible, one may ask, as the mere process of calculating contrast discards the ambient component of the luminance in the image? It turns out that even though the ambient component is removed, its information gets encoded in the relative intensity levels as the contrast map of any natural image is a highly constrained system when computed in an overlapping series of image patches. This enables a high fidelity reconstruction of the images from only their contrast information.

2 Motivation and Background

Humans are visual beings. Our visual system is not only the most elegant example of vision in nature, it is the foremost example we have. If it were not for our ability to see, we would have no idea whatsoever of what actually is a seeing experience. Hence, the motivation of this research is derived from discoveries in primate vision in the last fifty years.

Faced with a load of information, the primate retina with an approximate 130 million cells has to decide about transmitting the most important information, disregarding those that are not needed. It turns out that what is transmitted tells the cortex about changes in the pattern of information reaching the eye [7]. These changes are either in space, such as a border between a bright region and a dark one, or in time, such as a sudden increase or decrease in light intensity. Thus, the retina is primarily a detector of change, both spatially and temporally. Hence, the main function of the primate retina, in doing spatial analysis (as we are concerned), is to extract contrast from the luminance distri-

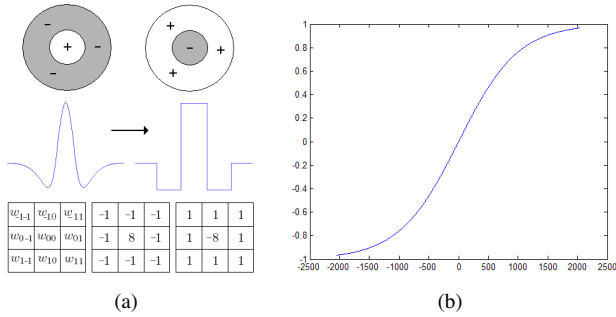


Figure 1: (a)*Top:* Retinal receptive fields. On-centre (left) and off-centre (right). *Middle:* Continuous DoG (left) modelling the retinal receptive field approximated by its discretized version (right). *Bottom:* a general 3x3 mask (left) and its on-centre and off-centre weights (centre and right) (b)Tangent sigmoidal squashing function with $\alpha = 1000$. With a mask of (a), the values will be in the range of -2040 to 2040 .

bution, thus making the percept largely independent of the ambient illumination level. This is achieved through a layered architecture of bipolar and ganglion cells that perform spatial differentiation (broadly symmetric) on their inputs using a finite sized, roughly circular, receptive field with antagonistic centre and surround. Two such types of cells are found to exist in the retina which are termed as the on- and the off-centre cells. The on-centre cells are activated when the centre of their receptive fields are brighter than their surround and deactivated otherwise. The off-centre cells work the opposite way by turning on when the surround is brighter than the centre and off otherwise. Together these two cell types capture all the spatial information that is available in an image.

This, then, brings up an interesting question as to how we perceive surfaces with uniform illumination - where the contrast is zero, as nothing apparently is being sent to the brain about the brightness levels of those regions (spatially speaking only, ignoring time effects) [7][4][14]. For this the brain must somehow and in some way perform reconstruction to get the relative brightness values of different regions of the image. Although we don't know what the internal representation of the reconstructed image in the brain is, we know that it must be done for providing us proper perception. The current research reports results and algorithms that we developed to perform an image reconstruction from just those contrast maps. We are, however, not suggesting that the brain does this in a similar way.

3 Related Work

The term *image reconstruction* as it appears in the literature has generally been used to imply some form of image

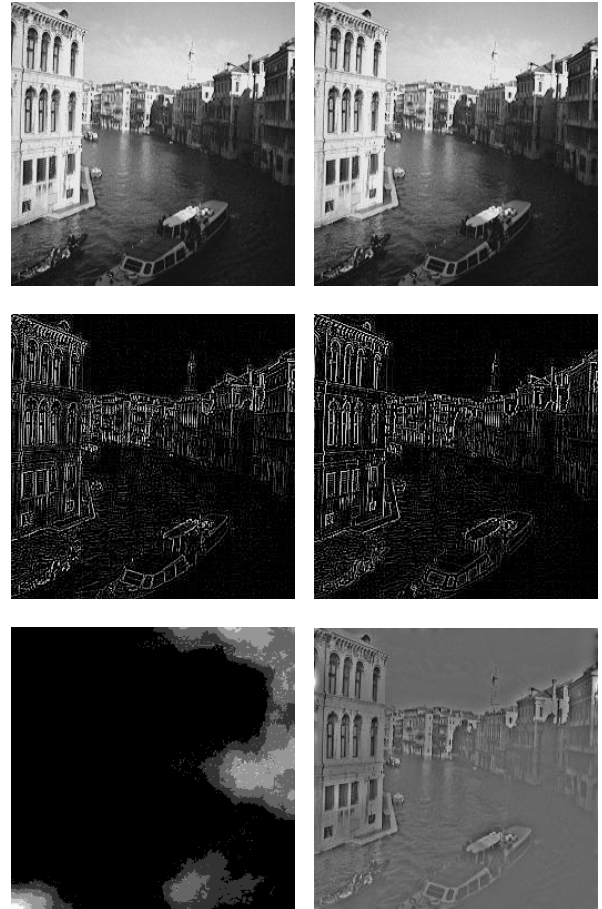


Figure 2: (*Top, left*): Original image (*top, right*): Reconstruction using our algorithm (*mid-left&right*): Its on & off centre map (*bot, left*): A 50x scaled error image of our reconstruction giving a PSNR of 48.82db (*bot, right*): Reconstruction using Wiener filter deconvolution.

filtration or reverse filtration to retrieve a denoised version of it. Deconvolution is usually performed for image restoration in many applications. Andrews and Hunt [1] describe in detail the linear algebraic restoration methods that include inverse filters, least-squares or Wiener filters, constrained least-squared filters and maximum entropy filters. Bates and McDonnell [3] describe these in terms of multiplicative and subtractive deconvolutions and other such techniques. In all these the point spread function (PSF) function is assumed to be known a priori. In cases where a PSF is unknown, attempts are made to restore the original image without the knowledge of the underlying noise function - a technique called blind deconvolution [2, 8]. However, central to all deconvolution methods is the underlying assumption that the image at hand is a blurred and noised image and so they majorly aim at performing deblurring and denoising. Besides, the selection of an appro-

appropriate technique depends upon the form and extent of the PSF, the nature of the original image and the extent of distortions in it [3]. Furthermore, as [3] points out that in order to apply standard deconvolution methods effectively, it is almost always necessary to preprocess the given image in some way including enhancement, sectioning, windowing etc. We show a comparison of our algorithm in Fig. 2 with that of the Wiener filter deconvolution. As can be seen, the deconvolutional output does not restore the brightness levels and introduces additional noise. In addition, the non-blind and blind versions of Richardson-Lucy deconvolution failed to reconstruct the image at all.

Some authors have mentioned image reconstruction from contrast information as an interesting issue in their work without providing any solution. Hubel, for one, raises this question in his book [7] in which he wonders how we are able to see uniformly lit surfaces when all that our cells respond to is the local intensity differences. However, without proposing any answers to it, he quickly disposes off the topic by supposing that the centre-surround architecture results in it somehow.

Other related work is perhaps the work of Land and McCann on *Retinex Theory* [9] which was followed up by that of Horn [6] and Marr [12]. The same are also cited by Palmer [14] when discussing the importance of intensity edges and luminance ratios. Retinex theory is a theory of colour constancy and brightness constancy in which luminance ratios are taken at edges and integrated across the visual field. Land [9] mostly talks about colour constancy and about intensity calculations in one dimension. Horn [6] takes it to two dimensions but mentions that since the component related to ambient light is differentiated out, there would be no way of calculating the brightness values of uniformly illuminated surfaces, while Marr [12] focuses mostly on the primate retina.

Sadr *et al.* [15] investigate local ordinal encoding and constraints. The authors show that such an approach can encode the most stable characteristics of an image and that it is neurally plausible, while [16] takes the neurophysiological data recorded from the lateral geniculate nucleus (LGN) of cats and attempts to reconstruct the image to some extent with the motivation of unveiling the coded information in the action potentials of neurons.

The work presented here differs from the rest as we put forward a non-linear method of reconstruction given only the contrast maps of an image. This opens up avenues whereby images can be transformed to the contrast domain, manipulated and then transformed back. Calculating the contrast of an image significantly loses information and the contrast map does not lend itself to such enhancement techniques as needed to make it suitable to apply classical deconvolutional methods. Furthermore, our approach differs from [15] and [16] in that through the use of an iterative,

non-linear technique we are able to reconstruct the image to a higher fidelity which in a small number of iterations produces an image which is visually indistinguishable to that of the original. Furthermore, we have developed a Bayesian framework for our method. *Bayesian Inferencing* has become a popular tool lately in image processing and computer vision, partially because vision [5] now is believed by some to be a Bayesian inferencing process [11].

4 The Algorithm

We compute on- and off-centre contrast maps from the original image. If \mathbb{M} represents the mask and \mathbb{I} the image, then a contrast map is given by

$$\mathbb{C} = \mathbb{M} * \mathbb{I} \quad (1)$$

where $*$ is the convolution operator and \mathbb{C} is the composite contrast map combining values from both on- and off-centre maps. This composite map without any additive noise is used in the algorithm for reconstruction. Algorithm 1 depicts the step-by-step reconstruction procedure. The following sections a detailed description of the algorithm.

Algorithm 1 Image Reconstruction From Contrast Information

```

1: img_in ← input
2: rf ← receptive_field_mask
3: contr_d ← compute_image_contr(img_in, rf)
4: eta ← 0.8
5: img_out ← initial_value
6: while stopping_condition ≠ true do
7:   for all [x, y] in img_out do
8:     contr_a ← compute_pixel_contr(x, y, rf)
9:     contr_e ← contr_d[x, y] − contr_a
10:    if contr_e ≠ 0 then
11:      img_out[x, y] ← img_out[x, y] + eta * contr_e
12:      if img_out[x, y] < 0 then
13:        img_out[x, y] ← 0
14:      end if
15:      if img_out[x, y] > 255 then
16:        img_out[x, y] ← 255
17:      end if
18:    end if
19:  end for
20: end while

```

5 Least Squares Error Minimisation

Using two 2D Gaussians with zero means,

$$G_{\sigma_1}(u, v) = \frac{1}{\sqrt{2\pi\sigma_1^2}} e^{-\frac{(u^2+v^2)}{2\sigma_1^2}} \quad (2)$$

$$G_{\sigma_2}(u, v) = \frac{1}{\sqrt{2\pi\sigma_2^2}} e^{-\frac{(u^2+v^2)}{2\sigma_2^2}}$$

we define our centre-surround receptive field R as a Difference of Gaussians (DoG) at any point (u, v) as

$$R(u, v) = G_{\sigma_1}(u, v) - G_{\sigma_2}(u, v) \quad . \quad (3)$$

We define the contrast at any pixel i at location (x, y) as a convolution of the receptive field R with the image intensities I in the neighbourhood of the pixel (x, y)

$$C(x, y) = I(x, y) * R(u, v) \quad (4)$$

$$C(x, y) = \int_v \int_u I(x - u, y - v) R(u, v) du dv \quad (5)$$

For the discrete case, the receptive field $R(u, v)$ is approximated by the mask $M(u, v)$ and the convolution then becomes

$$C(x, y) = \sum_v \sum_u I(x - u, y - v) M(u, v) \quad (6)$$

where u, v range for the size of the mask. However, because the combined input representing the entire neighbourhood influence can exceed limits, one must also specify a squashing function to keep resultants within reasonable range. We use a tangent sigmoidal non-linearity (Fig. 1(b)) as a squashing function

$$y(x) = \frac{1 - e^{-2x/\alpha}}{1 + e^{-2x/\alpha}} \quad (7)$$

where α is the constant for range adjustment and is dependent on a mask of particular size and weights. In certain applications, α could be made to vary as a function of local contrast serving as a dynamic range adjustment factor. Our cost function is defined as

$$C_{SE}(x, y) = \frac{1}{2} [C_D(x, y) - C_A(x, y)]^2 \quad . \quad (8)$$

We call the original image contrast map our desired contrast C_D and the contrast of the reconstructed image as the actual contrast C_A which is evaluated at every iteration and their squared difference is the squared contrast error measure C_{SE} at any pixel (x, y) . Based on this error measure, we update the actual pixel intensities until C_A converges to C_D .

The gradient of this error w.r.t. the actual pixel intensity I_A at (x, y) is

$$\frac{\partial C_{SE}(x, y)}{\partial I_A(x, y)} = -[C_D(x, y) - C_A(x, y)] \frac{\partial C_A(x, y)}{\partial I_A(x, y)} \quad (9)$$

so that,

$$\frac{\partial C_A(x, y)}{\partial I_A(x, y)} = R(0, 0) \quad . \quad (10)$$

But $R(0, 0) \approx M(0, 0) = w_{00}$ for a mask of finite size, giving us the following error expression

$$\frac{\partial C_{SE}(x, y)}{\partial I_A(x, y)} = -[C_D(x, y) - C_A(x, y)] w_{00} \quad . \quad (11)$$

Using gradient descent, the update rule then becomes

$$\Delta I_A(x, y) = \eta w_{00} C_E(x, y) \quad (12)$$

where η is the update constant and $C_E(x, y) = [C_D(x, y) - C_A(x, y)]$ is the contrast error at pixel (x, y) . Starting with an initial image, local changes are performed using this pixel update rule until some convergence criterion is met.

6 Bayesian Framework

Let us now develop a Bayesian framework for the least-squares error minimisation of Section 4. Given the desired contrast map (of the original image) C_D , the posterior distribution of the reconstructed image at each iteration, given Bayes' rule, is

$$p(\mathbb{I}_A | C_D, \mathbb{M}) = \frac{p(C_D | \mathbb{I}_A, \mathbb{M}) p(\mathbb{I}_A, \mathbb{M})}{p(C_D | \mathbb{M})} \quad . \quad (13)$$

The term $p(C_D | \mathbb{I}_A, \mathbb{M})$ is the probability distribution of obtaining the desired contrast map from the reconstructed image at any iteration. Since the contrast at any pixel i depends only on the image intensities at i and its defined neighbourhood and is independent of contrasts calculated at any other pixel, the likelihood, as well, at every pixel is independent of the others, hence

$$p(C_D | \mathbb{I}_A, \mathbb{M}) = \prod_i p(C_{D_i} | I_{A_{i \in N}}, \mathbb{M}) \quad . \quad (14)$$

where \mathbb{I}_A is the image pixel intensity, \mathbb{M} is the mask approximating the DoG (Eq. 3), C_{D_i} is the desired contrast at any pixel i , I_{A_i} is the actual intensity at pixel i and $i \in N$ represents the pixels in the neighborhood N of i . The contrast error at the i -th pixel, defined in terms of the desired contrast (C_{D_i}) and the actual contrast (C_{A_i}), is

$$C_{E_i} = C_{D_i} - C_{A_i} \quad (15)$$

having a probability distribution function given by a normal distribution with zero mean $p(C_{E_i} | \sigma^2) = N(0, \sigma^2)$, so that

$$p(C_{D_i} | I_{A_{i \in N}}, \mathbb{M}) = N(C_{A_i}, \sigma^2) \quad (16)$$

$$p(C_D | \mathbb{I}_A, \mathbb{M}) = \prod_i \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(C_{D_i} - C_{A_i})^2}{2\sigma^2}} \quad . \quad (17)$$

Using a normal distribution here could, theoretically, cause pixel values to cross the valid range of 0 – 255 because it is unbounded while an image is a strictly bounded system.

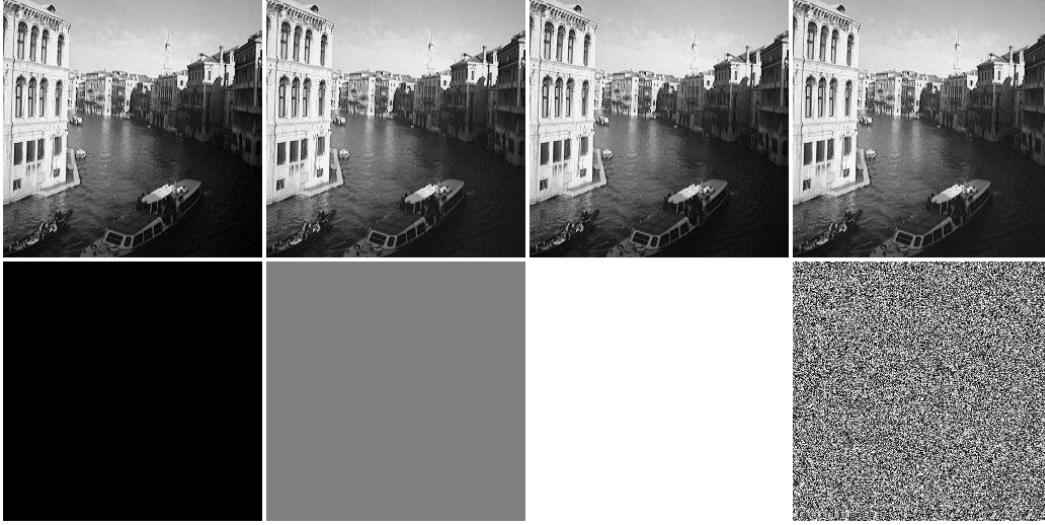


Figure 3: Reconstructed images in the top row with their corresponding initial images in the bottom row.

Although other distributions exist, like the beta distribution that may be more appropriate in such circumstances, they are generally harder to manipulate. However, we impose the desired boundedness algorithmically by limiting all pixels to 0 on the low side and to 255 on the high side.

Since we do not have any prior information about what the reconstructed image will look like, the best prior, in this case, would be a uniform random distribution in the range of 0 – 255

$$p(\mathbb{I}) = \prod_i p(I_{A_i}) \quad (18)$$

where

$$p(I_{A_i}) = \begin{cases} 0 & \text{for } x < 0 \\ \frac{1}{256} & \text{for } 0 \leq x \leq 255 \\ 0 & \text{for } x > 255 \end{cases} \quad (19)$$

Now, maximising $p(C_D | \mathbb{I}_A, \mathbb{M})$ is the same as minimising the negative log likelihood

$$-\log [p(C_D | \mathbb{I}_A, \mathbb{M})] = \frac{N}{2} \log(2\pi\sigma^2) + \frac{1}{2\sigma^2} \sum_{i=1}^N (C_{D_i} - C_{A_i})^2 \quad (20)$$

The first term in Eq. 20 is independent of the pixel intensity and so the second term is the sum of squared error for the entire image and for a single pixel, minimising the squared error would maximise the log likelihood

$$-\log [p(C_{D_i} | I_{A_i}, \mathbb{M})] = \frac{1}{2\sigma^2} (C_{D_i} - C_{A_i})^2 \quad (21)$$

This is essentially a least squares expression. For the case when the probability distribution function is normal

and there is no prior information available, the Bayesian paradigm reduces to the maximum likelihood which gives the same result as the least-squares method.

7 Implementation

7.1 Initialisation Values

The algorithm works by starting from some initial values for the pixels and progressively modifying them until converging to an image with the same contrast map as the given one. While the choice of initial values does not influence the final convergence result, a good initial guess can help in reducing the number of iterations needed for convergence. We experimented with a few different initialisations. As shown in Fig. 3, these were all black pixels, all white pixels, all mid-level gray pixels and pixel values from a uniform random distribution in the range 0 to 255. It turned out that our algorithm converges relatively faster when an initial image comprising of pixels from all mid-level gray or uniform random distribution is chosen.

7.2 Stopping Condition

Let the following be the images at different iterations:

$$\mathbb{I}_0, \mathbb{I}_1, \mathbb{I}_2, \dots, \mathbb{I}_{n-1}, \mathbb{I}_n \quad (22)$$

and let the absolute differences in these images be defined by the sequence

$$|\Delta\mathbb{I}_0|, |\Delta\mathbb{I}_1|, |\Delta\mathbb{I}_2|, \dots, |\Delta\mathbb{I}_{n-1}| \quad (23)$$

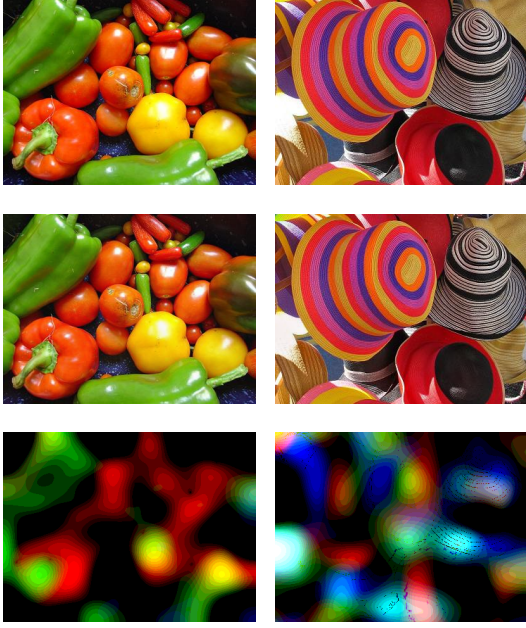


Figure 4: Original (top), reconstruction (middle) and 30x scaled error image (bottom) with a PSNR of 34.60db and 38.43db, respectively.

where $|\Delta I_{n-1}| = |\Delta I_n| - |\Delta I_{n-1}|$. Then for some threshold θ , we define the stopping condition as

$$\frac{|\Delta I_{n-1}|}{|\Delta I_0|} \leq \theta \quad (24)$$

7.3 Convergence

Although the least squares error cost function used here is non-linear, it is still a convex uni-modal function. In addition, our choice of α for the squashing function defined in Eq. 7 constrains most of the allowable values to the quasi-linear range around the origin, with few values curving off on the extremes as in Fig. 1(b). Together, these factors constrain the system of equations to solve in such a way that a unique solution is provided and convergence to the solution is guaranteed.

7.4 Normalised Masks

The pixel update rule in Eq. 12 depends on the centre weight of the convolution mask. This has the disadvantage that, if the mask is changed, the update will get severely affected. To make the update rule independent of such factors, we can use normalised masks with a centre weight of unity. This gives us the following update rule

$$\Delta I_A(x, y) = \eta C_E(x, y) \quad . \quad (25)$$

Doing so will require changes in the range adjustment constant α as well that is used in the squashing function (Eq. 7).

7.5 Spectral Analysis

The centre-surround filter acts as a high-pass filter keeping the high frequency components in the contrast information while discarding a band of low frequencies. These low frequency components get reconstructed through a diffusion process. The high frequencies determine the detail of an image. As the details are already preserved, information from the contrast diffuses to nearby areas, filling in, thereby reconstructing the image. Fig. 5 depicts the spectrum of the first few iterations showing the iterative reconstruction of low frequencies.

7.6 Results

We performed tests on images from the COREL 1000 image database (Fig. 6), generally used for content-based image retrieval. Each image was 256x384 pixels in size and was converted to grayscale before processing. The threshold in the stopping criterion used (Sec. 7.2) during the experiments was $\theta < 0.5\%$ which generated in a typical mean square error of less than 100. This resulted in reconstructions that are visually indistinguishable to the original image. However the algorithm requires a huge number of iterations to reconstruct such high fidelity images and typical values range anywhere from a few hundred iterations to a few thousand. This is due to the fact that there is a kind of diffusion process in effect where changes diffuse out from areas of high contrast to those of low or zero contrast. Similarly, images that are composed of large uniformly illuminated areas require an unusually high iteration count to produce reconstructions that are anywhere near the original.

8 Colour Extensions

The above methodology can be naturally extended to colour images by applying it to each of the three colour planes individually. Each colour plane of the RGB system can be treated as a grayscale image in itself. An on-centre and off-centre map for each plane is computed resulting in six contrast maps. This approach involves three times more computations than for grayscale images. We show some initial results in Fig. 4. It should be noted, however, that this is not the way a primate brain is known to handle colour information [10].

9 Conclusion and Future Work

We have presented an algorithm by which images can be reconstructed given only their contrast information. We

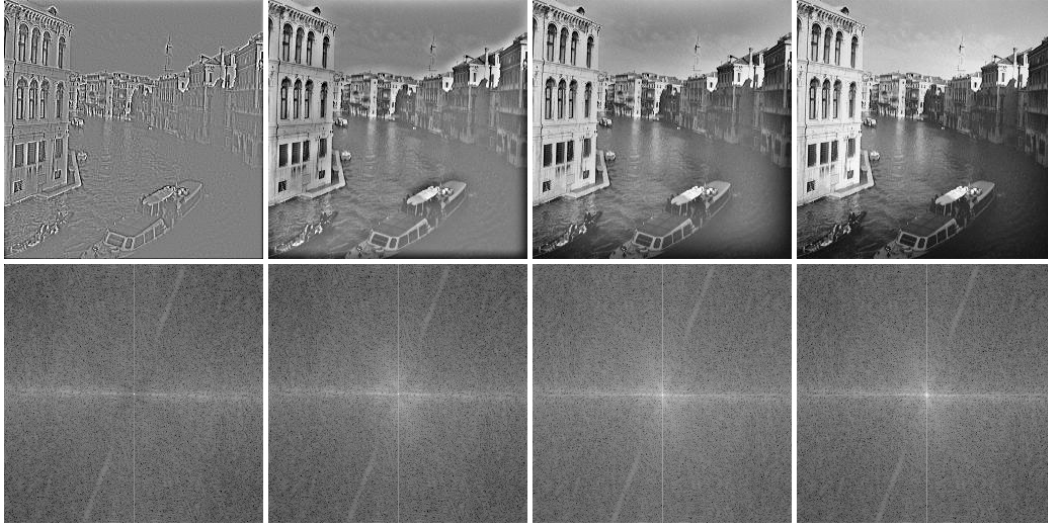


Figure 5: Reconstructed images and their FFT at iteration 1, 10, 100 and 500 showing the gradual reconstruction of low frequency components.

start with an initial guess and progressively update it by comparing its contrast map with that of the original image and using the difference as a cost function. We took inspiration from biological vision in primates where, by an unknown process, the brain is performing some kind of reconstruction internally to perceive not only the relative but to a large degree the absolute brightness values as well. Even though our method may be neurophysiologically plausible, as it uses only local computations, we refrain from making any such claims and as such use the neurophysiological approach only for inspiration. Our method can be viewed as providing a transformation and its inverse by which images can be transformed into their respective contrast domain, manipulated in interesting ways, and then transformed back into the image domain.

One of the potential future applications of the current work is dynamic range adjustment of images. In this paper, we have used the sigmoidal non-linearity with the same, constant parameters for all pixels. By making these parameters variable for different pixels that vary in accord with their neighbourhood, contrast values can be selectively expanded or compressed providing contrast enhancements and dynamic range adjustments.

References

- [1] H. C. Andrews and B. R. Hunt. *Digital Image Restoration*. Prentice-Hall Inc., 1977.
- [2] G. R. Ayers and J. C. Dainty. Iterative blind deconvolution method and its applications. *Optics Letters*, 13(7):547–549, 1988.
- [3] R. H. T. Bates and M. J. McDonnell. *Image Restoration and Reconstruction*. Oxford Press, 1986.
- [4] M. F. Bear, B. W. Connors, and M. A. Paradiso. *Neuroscience: Exploring the Brain*. Lippincott Williams & Wilkins, 3 edition, 2007.
- [5] K. M. Hanson. Introduction to bayesian image analysis. *Medical Imaging: Image Processing*, 1898:716–731, 1993. Proc SPIE.
- [6] B. K. P. Horn. Determining lightness from an image. *Computer Graphics and Image Processing*, 3(1):277–299, 1974.
- [7] D. H. Hubel. *Eye, Brain, and Vision*. Scientific American Library, 1995.
- [8] D. Kundur and D. Hatzinakos. Blind image deconvolution. *IEEE Signal Processing Magazine*, 13(3):43–64, 1996.
- [9] E. H. Land and J. J. McCann. Lightness and retinex theory. *Journal of the Optical Society of America*, 61(1):1–11, 1971.
- [10] M. Livingstone. *Vision and Art: The Biology of Seeing*. Harry Abrams Inc., 2002.
- [11] P. Mamassian, M. Landy, and L. T. Maloney. Bayesian modelling of visual perception. *Probabilistic Models of the Brain: Perception and Neural Function*, pages 13–36, 2002.
- [12] D. Marr. The computation of lightness by the primate retina. *Vision Research*, 14(12):1377–1388, 1974.
- [13] B. A. Olshausen and D. J. Field. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.
- [14] S. E. Palmer. *Vision Science - Photons to Phenomenology*. MIT Press, 1999.
- [15] J. Sadr, S. Mukerjee, K. Thoresz, and P. Sinha. The fidelity of local ordinal encoding. *Advances in Neural Information Processing Systems*, 14, 2002.
- [16] G. B. Stanley, F. F. Li, and Y. Dan. Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus. *The Journal of Neuroscience*, 19(18):8036–8042, 1999.
- [17] D. K. Warland, P. Reinagel, and M. Meister. Decoding visual information from a population of retinal ganglion cells. *Journal of Neurophysiology*, (78):2336–2350, 1997.



Figure 6: COREL images: (Left): Original, (centre): Their on- & off-centre contrast maps and (Right): Their reconstructions.