Building Perceived Colour Images

Xavier Otazu and Maria Vanrell
Computer Vision Center / Dept. Informàtica (UAB)
Edifici O, Campus UAB
Cerdanyola del Vallès, Barcelona, Spain

Abstract

Chromatic induction depends, among others, on the frequential content of the observed region.[5] As it is shown in Ref. [7], the two chromatic induction effects, i.e. chromatic contrast and assimilation, can be computationally simulated by blurring and sharpening operators, respectively. In this paper, we present a first unified approach to both effects using a wavelet decomposition approach. We propose a weighting function that modulates the multiresolution wavelet coefficients of any image point to perform either assimilation or contrast at every frequential level of the image. The recovered image present similar properties to a perceived colour image.

Introduction

Land et al.[2] showed how the colour perceived by the human visual system of a surface does not match with the physical light emitted by this surface. Then, the perceptual representation of the colour of a point depends on something else in the scene than just the physical properties of that point. There are simple and well known examples that easily show how colour appearance or perceived colour can change depending on the image content.[4]

In computer vision, researchers are in pursuit of automatic image understanding, the final goal is to put on computers the ability to act in front of any real scene in the same way as human visual system do. The automatisation of this process begins with the representation of a colour scene as a digital image. Colour on digital images is usually represented by the RGB responses of any commercial camera that is often used under uncontrolled conditions. Therefore, we would need to represent colour in such a way that simulates how colour is perceived by the human visual system. If we achieved this goal, further processes would be based on perceived colour.[3,7]

In colour literature all the changes in perceived colour caused by a nearby inducing stimulus are referred as colour induction mechanisms (see Figures 2 and 3). Smith et al.[5] measure the relationship between spatial frequency and colour induction effects, concretely on assimilation and contrast induction mechanisms. While the former behaves as a blurring effect, the latter is similar to a sharpening.

Considering these results, in this paper we propose a computational procedure as a first step to build perceived images where the effects of chromatic contrast and assimilation are produced regarding the local spatial frequency information of a digital image. Our unified approach is based on a wavelet decomposition of the colour image.

Wavelet Transform

Given an image, *I*, its wavelet decomposition is denoted by:

$$WT(f) \equiv \omega_{j,n}(I) = \left\langle I \middle| \psi_{j,n} \right\rangle = \int_{-\infty}^{\infty} I(z)\psi^*_{j,n}(z)dz \qquad (1)$$

where $\psi^*_{j,n}$ are the conjugate wavelet basis functions with parameters, j and n, related to the scale and pixel position respectively, and $\omega_{j,n}(I)$ is the decomposition coefficient of image I of pixel n and for the j wavelet plane. Given this decomposition, the original image can be completely recovered by integrating the coefficients with the basis functions. Although this is the general approach, in this work we will work on a particular case, it is the *à trous* algorithm.[1]

In the *à trous* algorithm, a sequence of images c_i is obtained by iteratively convolving these images by a low pass filter h. The difference between two consecutive images is the ω_j wavelet plane associated to a certain resolution j. This compact ω_j notation for the wavelet coefficients refers to the set of all the coefficients n, at a certain resolution j.

Using a one-dimensional notation for the sake of simplicity, we can see an initial discrete signal $c_0(k)$ (in the present case it would be an image, $I \equiv c_0(k)$) as a projection of continuous function $f(t)$ on a discrete V_0 space spanned by $\phi(t)$ basis functions, called *scaling functions*. The projection on a subspace $V_i \subset V_0$,

$$c_j = \left\langle f(t) \middle| \tfrac{1}{2^j}\phi\left(\tfrac{t}{2^j} - k\right) \right\rangle \qquad (2)$$

is then an approximation of c_0 at scale or resolution j. The approximation of coefficients c_{j+1} at scale $j+1$ can be calculated by means of the discrete convolution of coefficients c_j at scale j with a filter h,

$$c_{j+1}(k) = h(n) * c_j(k + n2^j)\,, \qquad (3)$$

and the wavelet coefficients can be calculated as the difference between two consecutive scales,

$$\omega_j(k) = c_{j-1}(k) - c_j(k)\,. \qquad (4)$$

This expression can be developed to show its recursive nature as a function of the original image I and the filters h_i:

$$\omega_0 = (I * h_0) - I$$

$$\omega_1 = ((I * h_0) * h_1) - (I * h_0)$$

$$\vdots$$

$$\omega_i = (\cdots(((I * h_0) * h_1) * h_2) * \cdots * h_i) - (\cdots((I * h_0) * h_1) * \cdots * h_{i-1})$$

In our case we have used a B_3 spline function for the scaling function $\phi(t)$, which leads to a *h(n)* function that can be approximated by a Gaussian kernel. The h_i filters are resampled versions of the original *h(n)* kernel. It is performed in order to accommodate, into the above convolution of this kernel with the original I image, the convolution of this kernel on the resampled $c_j(k + n2^j)$ data in equation (3).

The reconstruction of the original signal is simply the sum of all the wavelet coefficients plus the residual approximation $c_N(k)$,

$$I \equiv c_0(k) = \sum_{j=1}^{N} \omega_j(k) + c_N(k) \qquad (5)$$

From a computer vision point of view, the *à trous* algorithm can be understood as a multi-scale Laplacian filtering, where high frequency edges are separated on low index coefficients and low frequency edges appear on high index coefficients.

Perceived Image

Considering the results of previous section, now we can use the *à trous* algorithm as an efficient procedure to build a transformed image. In this framework we will call a perceived colour image, as a transformed image whose pixels have changed depending on its original colour and on spatial frequency properties; these chromaticity changes has to go in the same directions as induction mechanisms of the human visual system. This paper represents just one more step towards the final goal of building colour images that do not represent the physical properties of the light that reaches an acquisition device but the internal representation that human visual system builds for further processing. These perceptual representations must help in solving the ill-posed problems of computer vision.

Some induction mechanisms have easily been explained in terms of the image spatial frequency.[5] Colour assimilation is produced when colour appears in high frequency image regions, which makes we perceive the left side of Figure 2 as greenish. Colour contrast is produced when colour appears in low frequency image regions and it makes we perceive more contrast on blue and yellow regions of the right side of Figure 2.

Assimilation effects are usually simulated as a blurring of the image using Gaussian kernels, which means that high frequency features from the original image are diminished or eliminated. In contrast, chromatic contrast is implemented with Laplacians filters in order to increase differences between neighbour pixels.

To perform chromatic induction effects, i.e. assimilation and contrast, in the wavelet space we propose to introduce a weighting function that performs assimilation or contrast effects depending on the spatial frequency of the images, thus

$$I_{perceived} = \sum_{j=1}^{N} \alpha(j) \cdot \omega_j + \omega_r \, , \qquad (6)$$

where $\alpha(j)$ represents the *colour induction function*. It can perform a colour contrast effect, that is a sharpening effect if $\alpha(j) > 1$, and an assimilation effect if $\alpha(j) < 1$, that is a blurring effect. In this way, function $\alpha(j)$ can behave as a perceptual function that can simultaneously provide assimilation and contrast.

For higher frequencies (lower i values) we have to perform assimilation effects on the image, i.e. $\alpha(j) < 1$, and for lower frequencies (higher i values) we have to perform chromatic contrast, i.e. $\alpha(j) > 1$, which suggest an increasing function for $\alpha(j)$. An approximate generic profile for the $\alpha(j)$ function is shown in Figure 1. On the lower values of the wavelet plane, i.e. the higher frequency planes, the $\alpha(j)$ function has lower values in order to reduce the contribution of the higher frequency features, that is, to smooth the higher frequency features. On the higher wavelet planes, i.e. the lower frequencies, the $\alpha(j)$ function has higher values in order to increase the contribution of the lower frequency features, that is, to increase contrast between larger areas of the image.

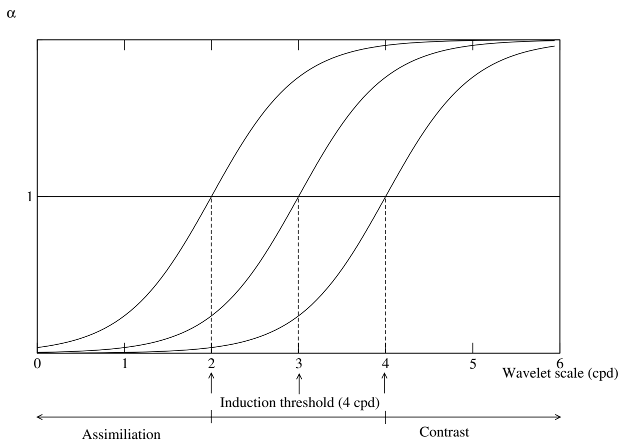

Figure 1. Approximate generic profile for the $\alpha(j)$ function.

There are several mathematical expressions that could be used for this induction $\alpha(j)$ function, e.g. truncated Gaussians, sigmoids, etc.; but a correct evaluation of this function should be obtained on the basis of psychophysical experiments.

An important problem we have to consider is where we put the threshold between assimilation and contrast on the $\alpha(j)$ function. Smith et al.[2] have proposed a threshold, hereafter induction threshold, where it is produced in the human visual system. They propose 4 cpd (cycles per degree) as the boundary between the two complementary induction mechanisms, i.e. assimilation and contrast.

Since the visual angle is the angle a feature presents from the human observer point of view, a feature shows a different visual angle depending on the distance from the observer. The larger the distance, the smaller the visual angle the feature shows. Hence, given a fixed visual angle, the number of image pixels inside this visual angle depends on the distance at which the image is observed.

The induction threshold value of 4 cycles/degree means that if an observer sees a feature which shows more than 4 cycles of variation inside a visual angle of 1 degree, this feature is blurred; on the opposite, if the observer sees a feature which shows more than 4 cycles/degree, the feature is enhaced.

We can define this induction threshold value in the image as a concrete spatial image frequency. For example, if the image is observed at a close distance, the threshold will be on the higher spatial image frequencies; since only the higher image frequencies will be contained into the 1 degree visual angle; on the opposite, if the image is observed at a large distance, the threshold will be on the lower frequencies.

Therefore, the $\alpha(j)$ function depends on the observation distance, and it is introduced as a shifting of the j parameter, that is, as a shifting of the $\alpha(j)$ function along the horizontal axis. This parameter will be denoted as j_{th}. When observing the image at lower distances, the $\alpha(j)$ function is shifted to the lower values of the wavelet scale (higher frequencies), i.e. to the left side of Figure 1. This way, the higher frequency planes (lower j values) of the $\alpha(j)$ function has higher values, that is, their values are increased, which means that they are not assimilated but contrasted. Lower frequency planes (higher j values) are contrasted since the $\alpha(j)$ function value is greater. On the other side, when observing at a large distance, the $\alpha(j)$ function is shifted to higher values of the wavelet plane (lower frequencies), i.e. to the right side of Figure 1. This way the higher frequency planes (lower j values) of the $\alpha(j)$ function has lower values, that is, their values are reduced, which means that they are strongly assimilated. Lower frequency planes are less assimilated or, in the case of the lower frequency planes, slightly contrasted.

Since several authors[4,6] suggest a maximum $m = \frac{3}{2}$ value for the chromatic contrast between features, the $\alpha(j)$ function has to be defined taking into account this restriction. Therefore, we have to define an $\alpha(j)$ function such that $0 \leq \alpha(j) \leq m$.

Opponent colour space is the colour representation usually used when modelling human colour visual system. Therefore, the above proposed algorithm has been applied on every opponent coordinate. This pose the question whether the same $\alpha(j)$ function has to be applied to every opponent colour coordinates. Since intensity channel contains most of the spatial resolution information and the other channels mainly contain the chromatic information, we should use different expressions of $\alpha(j)$ for every channel. Since psychophysical experiments should be performed to establish these differences, as a first approximation we applied the same function to every opponent channel.

Results

To test the behaviour of the model, two different kind of images where used. A synthetic image which contains two gratings with different spatial frequency, and some real scene images.

The synthetic image is used to test the behaviour of the method on grating images. In Figure 2. we show the original synthetic image with two different spatial frequencies on both sides. The perceived image obtained is shown in Figure 3. In Figure 4, we can see a profile from a row of the original synthetic image, and the values of this row when the perceptual operator is applied. High frequency features are clearly assimilated producing an almost uniform colour, showed as a reduction of the radiometric range values. In the low frequency right half of the image, situation is the opposite: wide stripes are contrasted, showed as an increased radiometric distance between them.

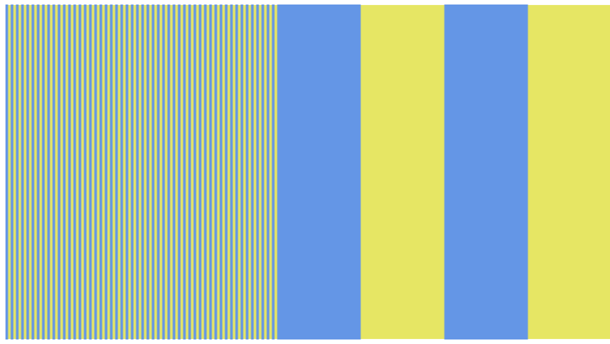

Figure 2. Synthetic image to apply the perceptual operator.

Figure 3. Perceptual image obtained from Figure 1, taking (jth=2).

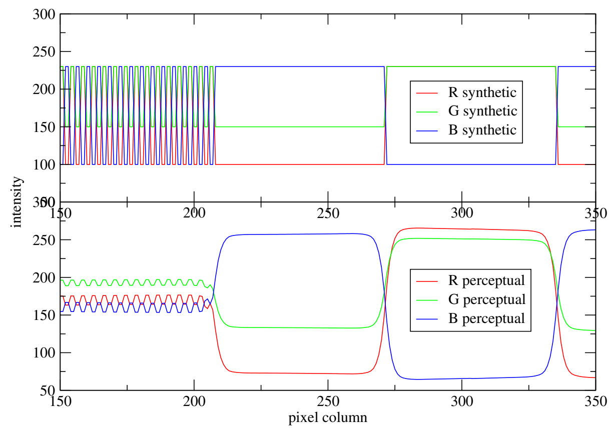

Figure 4. Profiles of a row from the original image (top) and the perceptual image (bottom).

In Figure 5 two real scene images are shown, which present high frequency features, candidates to be assimilated, as the forest leaves and the painting grains; and some global low frequency features showing contrast of colours between trees and painting features.

In Figure 6, different values for the induction threshold j_{th} have been applied on the forest image. Although it is inherently a high frequency image, its high frequency features are blurred depending on the induction threshold applied, but the color of the low frequency features is not affected by this blurring, always showing a highly contrasted yellow color.

In Figure 7 we show several perceptual images of the painting image, with the same effects that in Figure 6. Painting grains are assimilated but the contrast between more general features is enhanced, as the black and blue blobs.

Another example is shown on Figure 8, being both the original and a perceptual image on the upper row. On the lower row, details of these images are shown. The assimilation effect is clearly visible as a smoothing of the water surface at the same time that a contrast effect can be seen as a reddening of the red cork floats and a general colour contrast in the swimmer skin and water blue colour.

Some drawbacks of the method can be seen in Figures 9-10. The assimilation or contrast is performed based on the feature frequency, and the method does not distinguish between isolated high frequency features, edges or wide zones containing many high frequency features. Since the edge between the nose and the mandrill cheek is a high frequency feature, an assimilation is performed, hence, a blurring is obtained and the edge is smoothed. Another effect can be seen on the detail image of the single chin hair. This hair should be contrasted since it is in a dark background, but it is assimilated with the background, showing a blurred appearance. In figure 3, edges between wide blue and yellow stripes are smoothed, while only the left side narrow stripes should be assimilated.

The $\alpha(j)$ function only depends on the wavelet scale and the image observation distance, but it does not depend on the concrete pixel we are processing. That is, for a given j wavelet scale, the $\alpha(j)$ value is the same scalar value for all the pixels, so we are applying assimilation or contrast depending only on the global frequency we are dealing with. These problems suggest the need for a method that takes into account local information, that is, to perform assimilation or contrast depending not only on the feature frequency, but also on the information surrounding the feature. As shown in Ref. [5], high frequency features are assimilated when they are above some value of cycles/degree, that is, when more than eight vertical high frequency stripes are sighted inside 1 visual degree.

Conclusion

Multiresolution wavelet representation of colour images allows defining a computationally unified framework to simultaneously perform chromatic assimilation and contrast depending on the frequency content of the image. The $\alpha(j)$ function defines the chromatic induction process on this wavelet representation weighting the wavelet coefficients which describes the colour image.

One of the drawbacks of this model is that it does not preserve edges. Depending on its surround features, an edge can be associated to a high frequency feature, for example a line, or to a low frequency feature, for example the edge of a rectangle. Assimilation or contrast has to be applied to the edge depending on its surrounding information. Since this surround information has not been taken into account, the present model blurs all the high frequency features, even those which should be contrasted (as in the rectangle case).

The reason for this drawback is that weighting $\alpha(j)$ function has not been defined as a pixel dependent function, but it has only been defined as frequency dependent. This problem could be avoided defining an $\alpha(j)$ function which depends on the information from its surround area, and by extension on the pixel location. Hence, further improvement of this method is needed.

Figure 5. (Left) Image from a real scene. (Right) Detail from a painting.

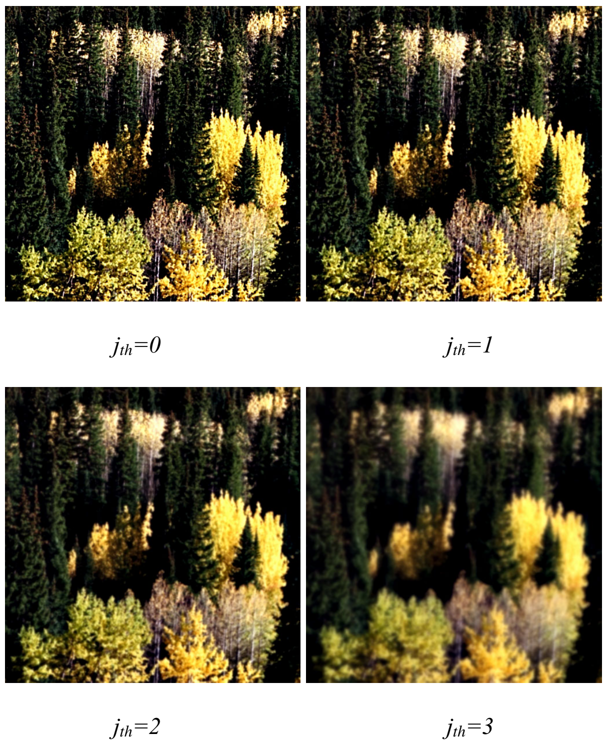

Figure 6. Perceptual images obtained from the real scene image in Figure 3 for several induction thresholds.

$j_{th}=0$ $j_{th}=1$

$j_{th}=2$ $j_{th}=3$

Figure 7. Perceptual images obtained from the painting image in Figure 3 for several induction thresholds.

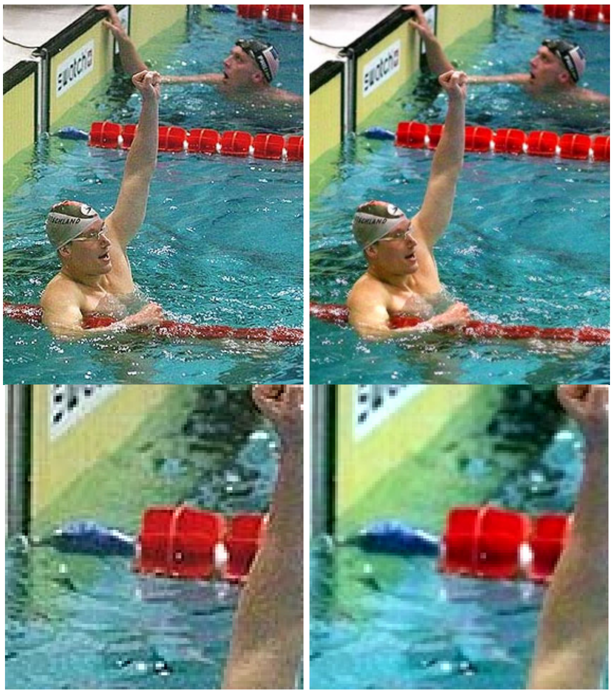

Figure 8. Original and perceptual images (upper, left and right, respectively), with $j_{th}=1$. Details of corresponding images (lower row).

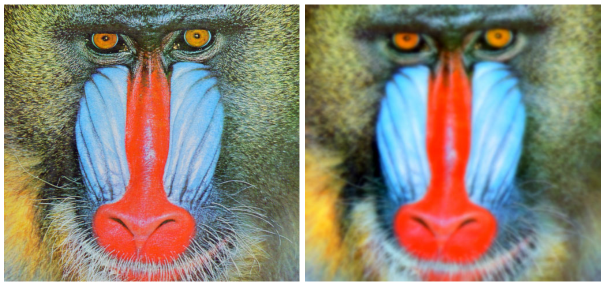

Figure 9. Original mandril image (left) and perceptual one (right) with $j_{th}=3$.

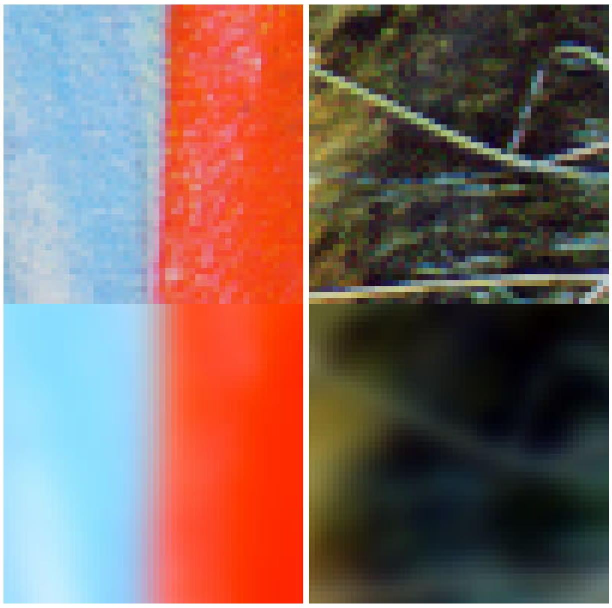

Figure 10. Details from original (top) and perceptual (bottom) images from mandrill images in Figure 7.

References

1. M. Holschneider, R. Kronland-Martinet, J. Morlet and P. Tchamitchian in Wavelet, Time-Frequency Methods and Phase Space, Springer-Verlag, 1989, pg. 289.
2. E. H. Land, Scientific American, 108 (1977).
3. M. Mirmehdi, M. Petrou. IEEE Trans. Pattern Anal. Mach. Intell. 22, 2 (2000).
4. R. Shapley and C. Enroth-Cugell, IEEE Trans. on Image Processing, 6, 7 (1997).
5. V. C. Smith, Phil Q. Ji and Joel Pokorny, Vision Research, 41, 1007 (2001).
6. H. Spitzer and S. Semo, Pattern Recognition, 32, 1645, (2002).
7. M. Vanrell et al. Computer Vision and Image Understanding, (in press).
8. X. Zhang, B.A. Wandell. SID (1996).

Aknowledgments

This work has been partially supported by project TIC2000-0382 of Spanish CICYT.

Xavier Otazu is grateful to the programme *Ramón y Cajal* funded by the Spanish Ministery of Science and Technology and Centre de Visió per Computador.

Biographies

Xavier Otazu received his B.S. degree in Physics from the University of Barcelona in 1994, a M.D. in Remote Sensing in 1999 from the Institute for Space Studies of Catalonia (IEEC), and a Ph.D. in Astrophysics from University of Barcelona in 2001. From 1998 to 2000 he worked in the Research and Development on Remote Sensing group at Catalonian Cartographic Institute (ICC). His Ph.D. work was focused on wavelet applications on astronomical and remote sensing image processing. His work on ICC was mainly focused on classification and fusion of multispectral data, and on texture-based analysis for segmentation of SAR data. He is presently at Computer Vision Center (CVC) working on texture and colour human perception. He is presently interested in the application of multiresolution wavelet methods on colour and texture human perception, and on other wavelet applications on remote sensing images.

Maria Vanrell joined the Computer Science Depart ment of the Universitat Autònoma de Barcelona (UAB) in 1990 and becomed associated professor in 1997. Currently, she is also a member of the research staff of the Computer Vision Center (CVC), a R&D institute founded by the UAB and the Generalitat de Catalunya (autonomous government of Catalonia). M. Vanrell is an active member of the Image Analysis and Pattern Recognition Spanish Association (AERFAI), a branch of the IAPR. Since 1996, when she received the Ph.D.degree at the UAB on work in Computational Texture Representation, her research interests are: texture perception and representation, computational colour representation, colour naming and the study of colour texture interactions. She has been the head and coordinator of several research and industrial projects and has published several papers in national and international conferences and journals.