

- [29] W. T. Hartmann and T. E. Madden, "Prediction of display colorimetry for digital video signals," *Journal of Imaging Technology*, vol. 16, pp. 11–21, Feb. 1990.
- [30] J. D. Foley, A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*. Addison-Wesley Publishing Company, 1996.
- [31] R. Ginosar and Y. Y. Zeevi, "Adaptive sensitivity / intelligent scan image processor," in *SPIE Visual Communications and Image Processing*, Nov. 1988.
- [32] J. J. Vos, "Colorimetric and photometric properties of a 2° fundamental observer," *COLOR research and application*, vol. 3, pp. 125–128, 1978.
- [33] W. K. Pratt, *Digital Image Processing*. John Wiley and Sons, 1978.
- [34] J. A. Saghri and A. G. Tescher, "Near-lossless bandwidth compression for radiometric data," *Society of Photo-Optical Instrumentation Engineers*, vol. 30, pp. 934–939, July 1991.
- [35] R. C. Gonzalez and P. A. Wintz, *Digital Image Processing*. Addison-Wesley, second ed., 1987.
- [36] B. R. Hunt and O. Kubler, "Karhunen-loeve multispectral image restoration, part 1: Theory," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-32, pp. 592–599, June 1984.
- [37] W. K. Pratt, "Spatial transform coding of color images," *IEEE Trans. on Communication Technology*, vol. COM-19, pp. 980–992, Dec. 1971.
- [38] A. K. Jain, *Fundamentals of Digital Image Processing*. Prentice Hall, Inc, 1989.
- [39] A. M. Rohaly and G. Buchsbaum, "Inference of global spatiochromatic mechanisms from contrast sensitivity functions," *Journal of the Optical Society of America A*, vol. 5, pp. 572–576, Aug. 1988.
- [40] A. Chaparro, C. F. Stromeyer(III), E. P. Huang, R. E. Kronauer, and R. T. Eskew(Jr), "Colour is what the eye sees best," *Nature*, vol. 361, pp. 348–350, Jan. 1993.
- [41] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE Trans. on Signal Processing*, vol. 5, pp. 1582–1586, Nov. 1996.

- [12] G. Wysecki and W. S. Stiles, *Color Science: Concepts and Methods, Quantitative Data and Formulae*. John Wiley and Sons, second ed., 1982.
- [13] S. L. Guth, "Model for color vision and light adaption," *Journal of the Optical Society of America A*, vol. 8, pp. 976–993, June 1991.
- [14] B. A. Wandell, *Foundations of Vision*. Sinauer Associates, Inc., 1995.
- [15] M. V. Srinivasan, S. B. Laughlin, and A. Dubs, "Predictive coding: A fresh view of inhibition in the retina," *Proceedings of the Royal Society of London B*, vol. 216, pp. 427–939, 1982.
- [16] Y. Tsukamoto, R. G. Smith, and P. Stirling, "Collective coding of correlated cone signals in the retinal ganglion cell," *Proc. Natl. Acad. Sci. USA*, vol. 87, pp. 1860–1864, Mar. 1990.
- [17] G. Buchsbaum and A. Gottschalk, "Trichromacy, opponent colours coding and optimum colour information transmission in the retina," *Proceedings of the Royal Society of London B*, vol. 220, pp. 89–113, 1983.
- [18] G. Buchsbaum, "Color signal coding: Color vision and color television," *COLOR research and application*, vol. 12, pp. 266–269, Oct. 1987.
- [19] J. B. Derrico and G. Buchsbaum, "A computational model of spatiochromatic image coding in early vision," *Journal of Visual Communication and Image Representation*, vol. 2, pp. 31–38, Mar. 1991.
- [20] S. G. Wolf, "Development of an adaptive colour model for colour image processing," Master's thesis, Technion-Israel Institute of Technology, Department of Electrical Engineering, June 1992.
- [21] R. Ginosar, Y. Y. Zeevi, and S. G. Wolf, "Apparatus & method for enhancing color images," Nov. 1995. USA patent 5,467,123.
- [22] V. R. Algazi, G. E. Ford, and E. Hildum, "Digital representation and storage of high quality color images by anisotropic enhancement and subsampling," in *ICASSP-89: Proceedings*, 1989.
- [23] E. Hering, *Outlines of a Theory of the Light Sense*. Harvard University Press, 1964. Translated by Hurvich and Jameson.
- [24] R. M. Boynton, *Human Color Vision*. Holt, Rinehart and Winston, 1979.
- [25] R. Gershon, "Aspects of perception and computation in color vision," *Computer Vision, Graphics, and Image Processing*, vol. 32, pp. 244–277, 1985.
- [26] W. N. Sproson, *Colour Science in Television and Display Systems*. Adam Hilger, Bristol, 1983.
- [27] R. S. Berns, R. J. Motta, and M. E. Gorzynski, "Crt colorimetry. part i: Theory and practice," *COLOR research and application*, vol. 18, pp. 299–314, Oct. 1993.
- [28] D. H. Brainard, "Calibration of a computer controlled monitor," *COLOR research and application*, vol. 14, pp. 23–24, Feb. 1989.

be performed in real time. The model and enhancement procedure together thus provide a simple spatio-chromatic color image enhancement procedure.

A natural extension of this model should integrate the three color components which depend not only on the *RGB* values at the same pixel location, but rather on some spatial neighborhood of that pixel such as a neighborhood resembling visual receptive field organization. This would implement a more complete spatio-chromatic model for color vision and processing.

The proposed processing may also be adapted for displaying and enhancing color information for individuals who have color deficiency (deuteranopes or protanopes). This is done by redefining the cone transformation T_{RV} such that the color image is mapped into that individual's cone response space.

Finally, better methods of enhancement, color image compression and quantization, color image segmentation and color filtering [41] ought to be investigated in the context of this model.

References

- [1] O. D. Faugeras, *Digital Color Image Processing and Psychophysics within the Framework of a Human Visual Model*. PhD dissertation, University of Utah, Department of Computer Science, June 1976.
- [2] O. D. Faugeras, "Digital color image processing within the framework of a human visual model," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. ASSP-27, pp. 380–393, Aug. 1979.
- [3] W. Frei and B. Baxter, "Rate-distortion coding simulation for color images," *IEEE Trans. on Communication Technology*, vol. COM-25, pp. 1385–1392, Nov. 1977.
- [4] M. Ardito, M. Barbero, M. Stroppiana, and M. Visca, "Compression and quality," in *Proceedings of the International Workshop on HDTV*, 1994.
- [5] C. J. van den Branden Lambrecht, "Color moving pictures quality metric," in *ICIP-96: Proceedings*, Aug. 1996.
- [6] D. Marr, *Vision*. Freeman, 1982.
- [7] J. S. Lim, *Two-Dimensional Signal and Image Processing*. Prentice-Hall, 1990.
- [8] A. F. Lehar and R. J. Stevens, "High-speed manipulation of the color chromaticity of digital images," *IEEE Trans. on Computer Graphics and Animation*, pp. 34–39, Feb. 1984.
- [9] I. Bockstein, "Color equalization method and its application to color image processing," *Journal of the Optical Society of America A*, vol. 3, pp. 735–737, May 1986.
- [10] R. N. Strickland, C. Kim, and W. F. McDonnell, "Digital color image enhancement based on the saturation component," *Optical Engineering*, vol. 26, pp. 609–616, July 1987.
- [11] S. Suzuki, M. Tokuyo, and M. Mori, *Color Change Method For Objects In Images*. Fujitsu Laboratories, Holography and Color-Imaging Laboratory, Fujitsu Laboratories Ltd., 10-1 Morinosato-Wakamiya, Atsugi 243-01, Japan.

Unfortunately owing to the limited scope of this paper we cannot illustrate all these results here.

7 Summary and Conclusions

We have developed a model for color image processing, based on the design of the HVS. First a colored image represented by calibrated red, green and blue planes, RGB , is transformed to the HVS cone response space using a linear approximation of this transformation. These linear approximations of the cone responses are then transformed to an opponent color space which has been shown to be similar to that of the retina [19]. This is done by modeling the retinal opponent transform by one that reduces redundant information inherent in natural images. Chromatic correlation (linear dependency between cone responses) reduction and energy compression is realized by using a multispectral Karhunen-Loève transform of the cone responses. Transforming an image in this way results in three new planes, a broadband (with respect to wavelength) achromatic channel (K_1), and two chromatic channels, red-green (K_2) and blue-yellow (K_3) as shown in Figure 3. Image processing is then performed in the perceptual space of K_1 , K_2 and K_3 .

It is interesting to note that natural images are characterized by low energy across most of the spatial frequency spectrum of the chromatic channels (K_2 and K_3). Higher energy is noticeable up to about 18 *cycles/image*. High spatial frequencies with noticeable energy are apparent only in the achromatic channel (K_1). This is illustrated in Figure 4.

The correspondence between the model response curves (Figure 3) in the case of natural images, and those response curves psychophysically predicted, indicates that the opponent process model well represents the behavior of the retina.

As a simple application we have investigated chromatic edge enhancement in the context of the HVS-based model. Based on evidence of both achromatic and *chromatic* retinal receptive fields, a biologically motivated processing scheme using the Laplacian operator is applied, as a simple example, to each of the channels. Surprisingly, in spite of the little spatial energy in the chromatic channels, chromatic edge enhancement does add significantly to the perception of image detail. This is in accordance with the recent finding that “Colour is what the eye sees best” [40]. This is contrary to the conventional wisdom that has motivated most of the research and technology to focus on achromatic enhancement only. Thus rather than treat a color image in the same way as an achromatic one and only enhance brightness, we have found that increasing chromatic high frequencies does add to image detail.

Experimentally we have also shown that chromatic edge enhancement in our HVS-based color space is visually superior to the same edge enhancement in other color spaces. This subjective result is explained by the fact that K_i are closely related to the HVS color-opponent space. Therefore HVS significant data is enhanced.

In addition, as the chromatic and achromatic channels are perceptually orthogonal, changes in one channel can be made without affecting the other channels. The resulting effect is then a combination of the effects exhibited by all three channels.

Since it is possible to simplify the opponent transform from a KLT per image to a constant transform for a wide class of images, the model can be efficiently realized in real time for video signals. Also, since the enhancement process can be based on small convolution masks, it too can

To discuss the enhancement we have prepared a reference image (Figure II) with regions of interest (A-G) marked.

Region-A contains tones of green and black. Little detail is improved with achromatic enhancement alone however chromatic enhancement in this region reveals hidden detail. This same effect is evident in region-F where details in the grass are much clearer in the chromatic enhanced version than in the achromatic.

Region-B illustrates how achromatic enhancement alone can degrade the chromaticity of an image. The pink flowers lose some of their color in the achromatic enhanced version although they are sharper. The color of the red flowers is unchanged as is their clarity. The chromatic enhanced version shows more colorful pinks and the red flowers are now clearer.

Region-C clearly shows how the yellow flowers and green stems are sharpened under chromatic enhancement while achromatic enhancement results in little change. The same is true of region-E where the red flowers are much more distinct in the chromatic enhanced image.

Region-D seems to be enhanced equally with both achromatic and chromatic enhancement. Chromatic enhancement sharpens the green areas and highlights pink spots on the flowers leaving the white (achromatic regions) almost unchanged. Because of the increased sharpness of the green area overall sharpness seems to be improved. Achromatic enhancement sharpens mainly the white/green border of the flower. This region contains distinct achromatic/chromatic boundaries and thus shows the effect of both achromatic and chromatic enhancement.

The effect in region-G is similar to that in region-C where chromatic enhancement sharpens the yellow and red flowers while achromatic enhancement has little effect.

Finally it should be noted that in the combined achromatic and chromatic enhanced image (Figure VI) the benefits of both type of enhancement are visible. It seems as expected as the two enhancements are orthogonal where change in one channel does not affect change in the other.

6.3 Comparison of Results to Other Methods

To compare results, the same Laplacian based edge enhancement scheme (unsharp masking) described in Section 6 was performed in various other color spaces (RGB , YIQ and YUV). This was performed on the test image of Figure III as well as to several other images [20]. The results using these techniques were inferior to our result as reported by subjective observation of a group of eight observers. We explain this surprising singularity by the fact that in these color spaces the channels are more highly correlated. Therefore each channel contains a relatively large achromatic component. Enhancing any of the correlated channels effectively yields preferential achromatic enhancement and in some cases even reduces details which are primarily chromatic. Even when we decorrelate the raw RGB data ((Karhunen-Loéve transformation) the results were inferior to ours. We explain this by the fact that our transformation takes the HVS into account and enhancements are effectively performed in a space resembling that of the HVS thereby emphasizing details significant to the HVS.

We also visually compared our method to a few methods that are based on the hue, saturation and intensity space including global saturation stretching [8], hue dependent saturation histogram equalization [9] and intensity edge enhancement based on saturation [10]. In all cases subjective quality in our method was judged to be superior.

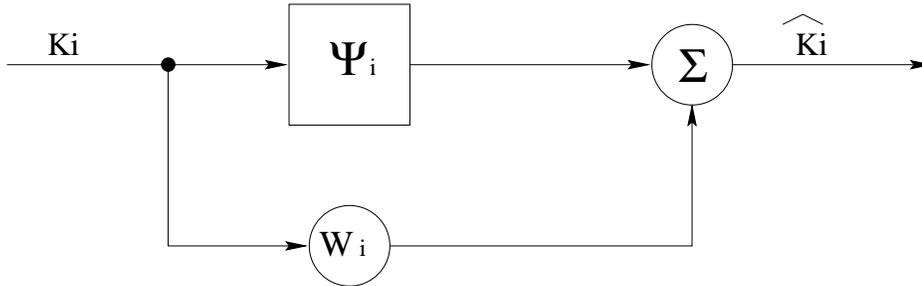


Figure 5: Generalized scheme for image processing in the context of the model. Ψ_i is an operator and W_i is a scalar weight applied to channel K_i . \widehat{K}_i is the processed channel.

ratios of α_1 to α_2 in the range of about 1 : 2 to 1 : 6 ensure more or less equivalent chromatic and achromatic high frequency enhancement. The enhanced channel $K_{i(enh)}$ is essentially a high frequency enhanced version of K_i .

Since Laplacian edge enhancement emphasizes high spatial frequencies, it also increases high frequency noise. In fact it can be shown [20] that for a stationary white noise process of average power σ_o^2 , the noise power increases as the square of α .

As we have shown however, the chromatic channels are essentially spatially low-pass and thus contain little noise. Higher values of gain α can thus be tolerated in these channels. On the other hand, the achromatic channel contains high frequencies at a significant level. To avoid noise increase we should thus avoid high frequency enhancement in this channel. As a tradeoff between enhancement and noise increase we have experimentally found that values of ($\alpha_1 = 1$) and ($\alpha_2 = 3$) or a ratio of $\alpha_1 : \alpha_2 = 1 : 3$ give good results for natural images (see Section 6.2).

6.2 Experimental Results

To illustrate¹ a possible use of the model the simple edge enhancement method discussed in Section 6.1 was applied to a number of images of natural scenes containing achromatic and chromatic regions.

The original of one of these images (Figure III) shows a garden containing red, green, blue, yellow and white regions. Figure I contains a 2×2 mosaic of four images showing the original image and three enhanced versions; achromatic enhanced only (K_1 enhanced); chromatic enhanced only (K_2 and K_3 in equal amounts); and the final result being a combination of achromatic and chromatic enhancement (K_1 , K_2 and K_3). The mosaic images are scaled down versions of the images in Figures III - VI where the scaling is a simple sub-sampling without further filtering.

The full size images are of dimension $640 \times 480 \times 24$ but have been printed at a larger scale, again with simple up-sampling and without other filtering. This has obviously resulted in the evident jaggedness but this is irrelevant to our discussion. It should also be noted that contouring (color banding) due to the limitations of the printer is evident. After processing each of the *RGB* planes are scaled by the *same* factor to avoid clipping thus global color shifts are not produced.

¹The enhancement results are available for viewing at www.ee.technion.ac.il/~ran/papers/WGZ.

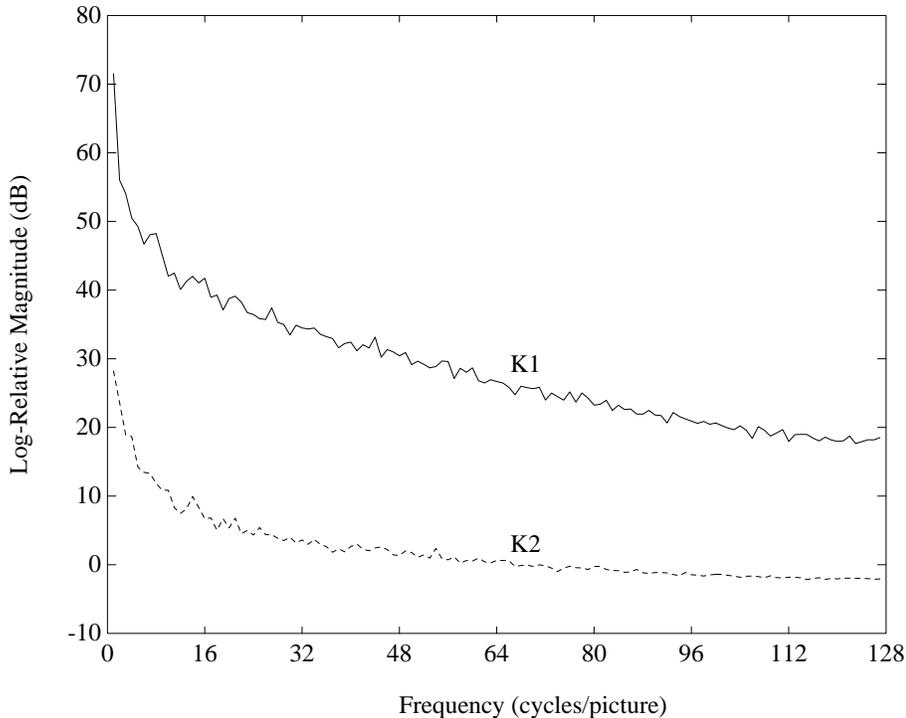


Figure 4: Average logarithmic power spectrum of the K_1 and K_2 planes. K_1 retains a large proportion of the total image energy at high frequencies while K_2 (and similarly K_3) contains significant energy only in the range of low frequencies. The frequency axis units are in cycles/image where each image contains 256×256 pixels.

primal sketch, can be extracted by a Laplacian operator which can also be implemented by center-surround receptive fields found in the retina. Thus traditional enhancement, whereby edges are added to an achromatic image can be understood by the existence of luminous center-surround retinal receptive fields. Neurophysiological evidence points to the existence of similar red-green and blue-yellow chromatic receptive fields [25]. Therefore chromatic edge enhancement should also enhance the visual quality of a color image. It has also been shown that spatial frequency is analyzed separately in the chromatic and luminance channels [25]. We therefore perform Laplacian edge enhancement separately on both the achromatic and each of the chromatic model channels according to the processing scheme of Figure 5 and the processing operator described by equation 5.

Because of the lack of an objective quality metric we define “visual quality” subjectively as a combination of sharpness of edges, vividness and visibility of faint details and general “colorfulness” of a colored image.

L is a standard Laplacian operator (see, for example, [33]). The result of the Laplacian operation is amplified by gain α_i which controls the degree of enhancement in each channel. Without knowledge of the image we choose equal chromatic enhancement, that is $\alpha_2 = \alpha_3$. By considering the relationship between the HVS luminance and chrominance contrast sensitivity functions [39],

axis is relatively small, and since the KLT aims at minimizing the variance along the transformed axes, it is not surprising that the largest contributor to K_1 is V_s , as can be seen in the expression for Φ_h (4). Therefore it seems that the cone response functions have also adapted in the HVS such that when followed by the opponent stage, energy compression will be maximized. This result is in accordance with the role of the HVS as an efficient information transmission system.

Other image orthogonalization techniques which are used in practice, such as YUV (see Table 1), YIQ , and HSI , seem to produce similar results to a straightforward KLT of RGB . To the best of our knowledge, only the proposed transform which incorporates properties of the HVS produces such dramatic decorrelation.

In Figure 4 we present the average power spectral density (PSD) of the achromatic (K_1) and chromatic red-green (K_2) channels. These were obtained by averaging the PSD calculated for characteristic individual scanlines. The behavior of the yellow-blue channel is similar to the red-green channel but much reduced in amplitude. The PSDs of the achromatic and chromatic channels demonstrate that the achromatic channel contains energy in the high spatial frequencies while the chromatic channels have little or no energy in this range. This lack of energy in the range of high frequencies amounts to spatial lowpass filtering which reduces the energy and hence the bandwidth of the chromatic information as compared to the achromatic one. Edges exist at the same spatial location in each color plane [19], and are thus retained in the achromatic channel. This is so because K_1 always consists of an additive combination of V_l , V_m and V_s as explained in Section 2.2. In the chromatic channels, edges and other high spatial frequencies are diminished since spatial subtraction reduces the energy of these features. Hence the well-known fact that chromatic channels have spatial lowpass characteristics.

6 Image Processing in the Context of the Model

A generalized image processing scheme based on the proposed model is shown in Figure 5. An operator or set of operators (Ψ_i) is applied to each model channel (K_i). This result is then added to a weighted version of K_i . In this work, as a simple application, we have chosen chromatic edge enhancement based on neighborhood operators. Hence,

$$\Psi_i = \alpha_i L_8 \tag{5}$$

where L_8 is the standard 3×3 Laplacian operator as defined in [33] and $W_i = 1$. This is detailed further in Section 6.1. This choice of operators and weights results in an enhancement operation commonly used for color image enhancement (of RGB) known as unsharp masking [38].

6.1 Chromatic Edge Enhancement

Although the separation of a color image into an achromatic and two opponent chromatic channels is done by considering only the chromatic content, these channels have a definite spatial structure. We know that spatial information is even more important to the HVS than is chromatic information. Evidence for this is derived from our easy interpretation of black and white images or even of simple line drawings.

To explain the importance of spatial information, Marr [6] suggested that the primal sketch of an image serves as the basis for image understanding by the HVS. Edges, which constitute the

We wanted to determine whether there is a visual difference in using the constant transform as compared with an image matched model prior to enhancement. To this end, we selected an image from the set of natural images, but one that was *not* used in the derivation of Φ_h . We then applied the same enhancement to two different transformed versions of this image. One version was obtained using the specific transform for this particular image. The other was obtained using Φ_h , the transform optimal for the set of natural images. We then performed the enhancement described in Section 6.1 on both images. The results were perceptually indistinguishable.

The implication of this result is that the model can be used to derive constant transforms for different sets of images. These sets may be as diverse as the entire set of all natural images. This general (constant) transform can be computed off-line, thus saving the time consuming process of calculating the image-specific transform Φ which takes about 1.5 minutes on a Sun-4 computer for a color image of spatial resolution of 512×512 . The model can therefore be simplified to two 3×3 linear matrix transformations. This simplified procedure can easily be executed in real time on video signals.

The constant opponent model used by van den Branden Lambrecht [5] and derived in [14] is based on a decorrelation of the cone response functions using a singular value decomposition. This is an arbitrary linear transform whereas our constant transform is based on correlation reduction in the cone response space of natural images which we assume that the HVS is best adapted to.

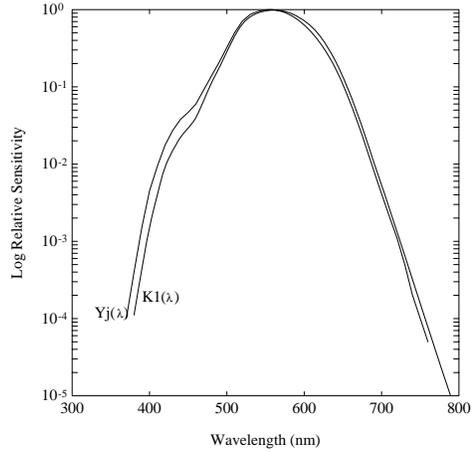
5 Properties of the Model

The KLT redistributes the image energy so that its expected value in the achromatic channel is maximized. Less than 1% of the energy is retained in the red-green channel and much less in the blue-yellow channel. This is illustrated by the results of two natural images, as shown in Table 1; one of a landscape view and one of a ship, transformed by the model. The red, green and blue planes of these images have a near even energy distribution as is characteristic of such natural images [19, 34, 36, 37].

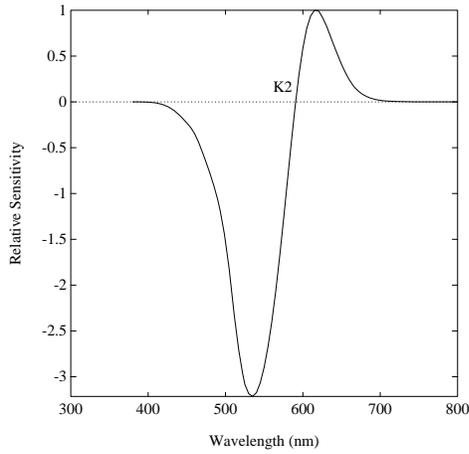
Image	Original image			Uniform space			K space		
	R	G	B	Y	u	v	K_1	K_2	K_3
View	24.0%	35.7%	40.3%	89.6%	6.5%	3.9%	99.95%	0.048%	0.002%
Ship	20.4%	37.6%	42.0%	80.5%	7.1%	12.4%	99.94%	0.059%	0.001%

Table 1: Proportion of total image energy in the original and transformed versions of two color images.

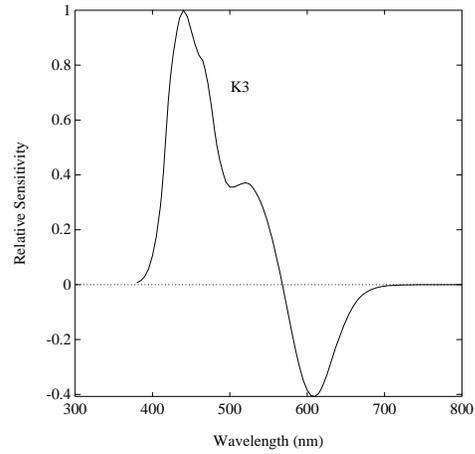
In previous studies [3, 34, 36, 37], where the KLT was used for coding purposes, the *original RGB* values were transformed by the KLT. The energy distribution in the transformed planes in these cases were of the order of $K_1 : K_2 : K_3 = 86 : 13 : 1$. The reason that we have achieved a much higher concentration of energy in the K_1 channel relative to the other two channels is our transformation of the cone tristimuli, V_l , V_m and V_s instead of the original *RGB* values. Looking at Figure 1, we see that the amplitude of $V_s(\lambda)$ relative to $V_l(\lambda)$ and $V_m(\lambda)$ is much smaller. Hence there is already a measure of information compression *prior* to the transformation. Since the variance along the V_s



(a)



(b)



(c)

Figure 3: (a) Spectral sensitivity of the achromatic (K_1) channel compared to Judd's relative luminous efficiency curve ($Y_j(\lambda)$). (b) and (c) depict the spectral sensitivity curves of the red-green (K_2) and blue-yellow (K_3) channels. The K_1 curve has been shifted upwards for comparison and the K_2 and K_3 curves are normalized. (a), (b) and (c) are transformed versions of the curves shown in Figure 1

It is clear that $(T_{RV})^{-1}$ exists, and because the correlation matrix is real and symmetric and therefore has orthogonal eigenvectors, $\Phi^{-1} = \Phi^t$ [35]. Therefore the complete inverse transform always exists.

3 Correspondence of Model to HVS

In this section we verify that the color opponent transform based on redundancy reduction is indeed a good approximation to the HVS opponent transform.

If we assume that primates have evolved such that their visual system is best suited to their natural environment, then the HVS opponent transform should be best adapted to natural daylight images. To derive the HVS transform we selected eight natural daylight images and processed this set using the model of Figure 2 to derive the opponent transform matrix. This resulted in the constant transform matrix Φ_h (see Section 4) where,

$$\Phi_h = \begin{bmatrix} 0.9517 & 0.3070 & 0.0032 \\ 0.3068 & -0.9516 & 0.0193 \\ -0.0089 & 0.0174 & 0.9998 \end{bmatrix} \quad (4)$$

Using Φ_h to transform the Vos-Walraven cone response curves shown in Figure 1 results in the channel response functions shown in Figure 3. In Figure 3 (a) we can see the close agreement between the model predicted K_1 channel and the modified relative luminous efficiency curve derived by Judd [12]. K_1 therefore contains the luminance or black-white information and is called the achromatic channel. The K_2 and K_3 chromatic response functions shown in Figure 3 (b) and Figure 3 (c) respectively, are similar to the red-green and blue-yellow opponent color channels psychophysically estimated [12]. These channels thus contain the opponent chromatic information.

Buchsbaum showed that maximal decorrelation predicts the HVS opponent color transform by assuming that the image that reaches our eye belongs to an ensemble of images which have a broad-band spectral (as opposed to spatial) Fourier frequency power spectrum [17]. Strictly speaking, this is true only for images made up of a combination of monochromatic signals such as those used in visual psychophysics. The constant transform Φ_h which we have found using natural images, not necessarily made up of monochromatic colors, is similar to Buchsbaum's transformation matrix and also yields channel response functions resembling those of the HVS. This result reinforces Buchsbaum's conclusion that the goal of the opponent transform in the retina is to achieve redundancy reduction and information compression and confirms that the HVS is optimally adapted to natural scenes.

4 Constant Opponent Transform

What distinguishes this model from other opponent color models [2, 5, 12–14] is that the opponent transform is dynamically calculated for each image. In order to simplify the model so that it may be used in real time it is desirable to use a constant transform. In Section 3 we showed that processing a set of images with the model indicates that a constant HVS opponent transform is most suitable.

an achromatic and two opponent chromatic channels. Thus the opponent transform can be viewed as a method for removing redundant information arising from the high *chromatic* correlation of natural scenes.

Derrico and Buchsbaum [19] combined methods of spatial and chromatic correlation reduction into a computational model of image *coding* in early vision. The first stage of their model consists of performing an eigenvector transformation on the original red and green planes to reduce chromatic correlation followed by predictive coding on the achromatic channel to reduce spatial correlation. By applying this coding technique to only the red and green planes of a small collection of images they transform an image into two channels which are known to have correlates in visual psychophysics.

In our image *enhancement* method, we have adopted this concept of determining the opponent transform such that chromatic correlation is reduced. Optimal reduction of chromatic correlation and information compression is achieved by performing a multi-spectral Karhunen-Loève transform (KLT) on the cone responses. This reduces chromatic correlation at each spatial location. Thus, the opponent transform matrix, denoted by Φ , consists of the normalized eigenvectors of a chromatic correlation matrix [17,33,34]. Since the cone responses are non-negative, the correlation matrix is real and positive and will always have only one eigenvector whose entries are of like sign [17]. This means that only one of the transformed channels generated by Φ can be an all positive combination. The other two channels must therefore be antagonistic. The all positive channel is labeled K_1 . The two other antagonistic channels are labeled K_2 and K_3 . As we discuss in Section 3, K_1 represents the achromatic or black-white component of the image while K_2 and K_3 represent the red-green and blue-yellow components respectively. K_2 and K_3 are antagonistic since, for example, increasing positive values of K_2 represent increasing values of red while decreasing negative values represent increasing values of green.

Image processing is then performed on these opponent channels and the result is inverse transformed back to the red, green and blue planes for display purposes. A block diagram of this biologically-motivated processing scheme is depicted in Figure 2.

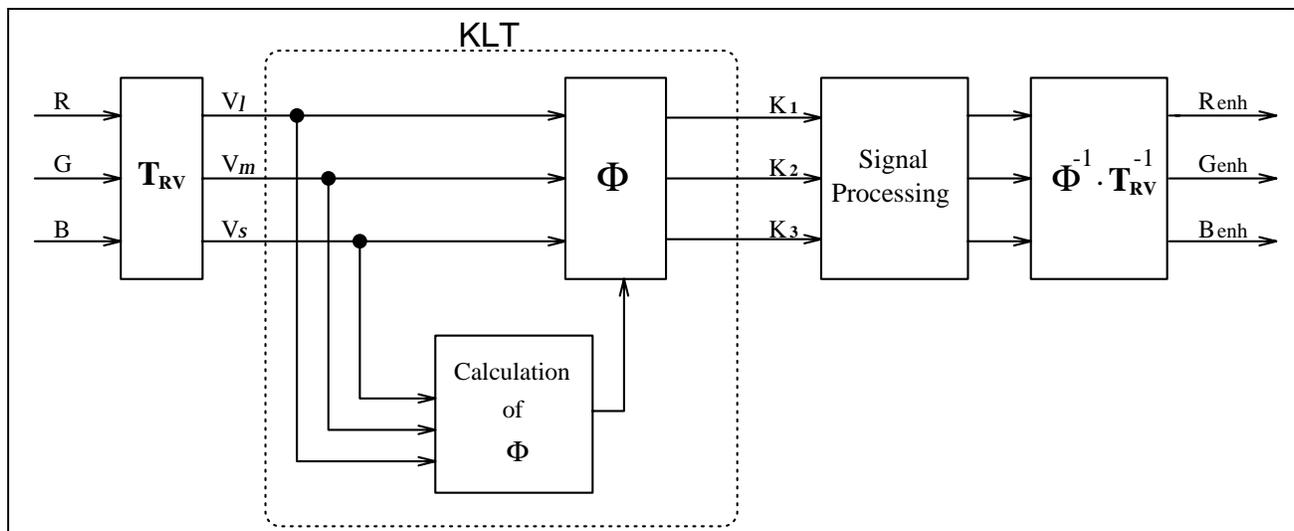


Figure 2: Block diagram of the generalized system for chromatic enhancement

which results in,

$$T_{RV} = \begin{bmatrix} 0.25653 & 0.38702 & 0.79427 \\ 0.42417 & 0.25035 & 0.17708 \\ 3.6682 \times 10^{-8} & 2.4582 \times 10^{-4} & 4.1504 \times 10^{-3} \end{bmatrix} \quad (3)$$

Thus at each spatial location in the image we transform the *RGB* tristimuli to $V_l V_m V_s$ using the color transform, T_{RV} .

2.2 Separation of Achromatic and Chromatic Information

We have investigated various types of color image enhancement techniques including histogram based methods [20] as well as the technique explained in Section 6.1. These methods were applied to colored images transformed in various standard ways such as to the *YIQ*, *YUV*, and *YCrCb* color spaces; orthogonalization (KLT transformation) of the raw *RGB* data; as well as to a hue, intensity and saturation space. Visual observation by a group of eight observers could not distinguish any significant *local chromatic* enhancement in any of these spaces. Therefore we began to consider modeling the transformation of color similarly in the way it appears to be performed in the HVS.

Although we know that radiant energy is mapped into three spectral variables by the cone responses and later into three opponent channels, the transformation from the cone trichromatic responses to the opponent color responses of later retinal stages is still not absolutely clear.

By considering the HVS as an efficient information processing system, some surprising results have emerged. “Efficient” means that redundant information is reduced and energy is compressed such that the channel capacity required to transmit the information at a given level of reliability is minimized. Reasons for considering this approach in the HVS are that natural visual scenes contain both spatial and chromatic regularities, making the retinal image contain much redundancy. In addition, the ratio of photo-receptors to optic nerve fibers is at best 6 : 1 if we only consider cones and about 120 : 1 if we take rods into account as well [25]. Thus if the initial photo-receptor responses were left unaltered, the limited channel capacity of the optic nerve fibers leading to higher stages of the HVS would be unable to cope with this large amount of data. This reasoning has led to the hypothesis that the purpose of image transformation in the early visual system is to reduce correlation and compress information prior to neural transmission [17].

Based on this hypothesis, Srinivasan *et al* [15] accounted for the spatial profile of center-surround retinal ganglion cell receptive fields using a linear predictive filter. They showed that the spatially antagonistic center-surround structure of these receptive fields decorrelate and suppress an image signal. As such they can be viewed as a method for removing redundant information arising from the high *spatial* correlation of natural scenes. Similarly Tsukamoto *et al* [16] demonstrated that the Gaussian profile of retinal ganglion cells is optimal in terms of minimizing spatial correlation.

In terms of *chroma*, apart from the chromatic regularities of natural scenes, additional chromatic correlation is due to the highly overlapping nature of the cone chromatic response functions as can be seen in Figure 1. In a study of chromatic correlation reduction, Buchsbaum and Gottschalk [17] used an eigenvector analysis to investigate the role of opponent type processing in color vision as well as the relation between opponent color transforms and the initial cone responses. They showed that redundancy reduction and information compression is achieved by a transformation to

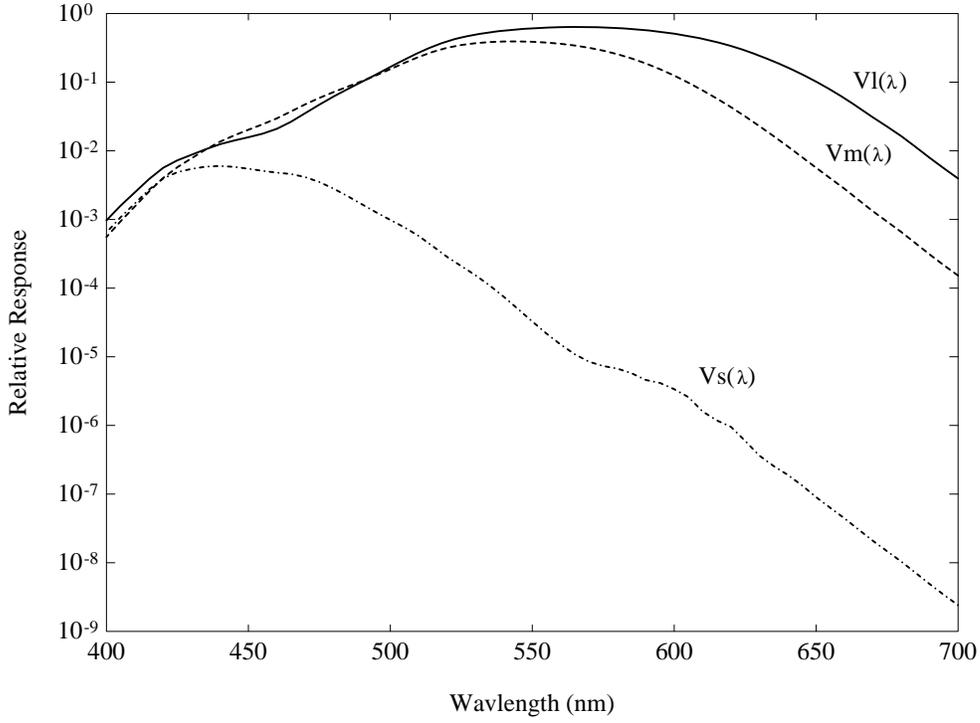


Figure 1: Vos-Walraven cone spectral sensitivity functions $V_l(\lambda)$, $V_m(\lambda)$, and $V_s(\lambda)$.

tristimuli to $V_l V_m V_s$. That is,

$$\begin{bmatrix} V_l \\ V_m \\ V_s \end{bmatrix} = T_{RV} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

The transformation, T_{RX} , from RGB to the CIE XYZ tristimulus values for a standard television monitor is derived as described in [18, 26]. For other monitor systems a color calibration procedure should be followed [27–30]. The transformation, $T_{X'Y'Z'}$, from the modified $X'Y'Z'$ to $V_l V_m V_s$ is well known [12, 18]. The transformation from XYZ to $X'Y'Z'$ which may be implemented using the Vos equations [32], is a non-linear function of the luminance. It is, therefore, constant for the whole image only if Y does not change across the image. Rather than calculate the transformation for each pixel in the color image we have assumed that for image processing purposes, the $X'Y'Z'$ and XYZ values are sufficiently close that this transformation is not necessary. This approach is verified by Guth who has used the same approximation in his latest model of color vision [13] as well as by the results we have obtained. We have adopted this simplification since it accelerates the transformation. This is important if the model is to be used for real-time applications.

With this simplification in mind, using matrix algebra we get,

$$T_{RV} = (T_{VX'})^{-1} \cdot T_{RX} \quad (2)$$

2 The Enhancement Model

The spatio-chromatic enhancement is based on the opponent color theory first proposed by Her- ing [23] which proposes three mechanisms that mediate in color vision: one that accounts for red-green perception, a second for blue-yellow perceptions and a third for black-white or bright- ness discrimination. Physiologically this theory is supported by the existence of red-green and blue-yellow as well as luminance center-surround opponent receptive fields in the retina [12,24,25]. Luminance information is first transduced and processed non-linearly by the three cone types and then, as the signals flow through the retina, they are transformed to the three opponent responses.

The chromatic enhancement processor discussed in the following sections consists of three main building blocks: Cone transformation (T_{RV}) which transforms the image data into a cone response space; opponent transform (Φ) which decorrelates the cone responses and a signal processing stage. This three stage model is illustrated in Figure 2. The processing method is not intrinsic to the model and is chosen according to what is required to be done to the image. For example for edge enhancement this stage could be a simple spatial convolution or for compression it could be a predictive coder. One method suitable for generalized enhancement is separate chromatic and achromatic (luminance) edge enhancement.

2.1 Transformation to Cone Primaries

The input to the system for chromatic enhancement consists of red (R), green (G) and blue (B) planes of a colored image calibrated to the standard color television monitor primaries [18,26]. These planes represent the camera’s response to the viewed scene. These responses are first transformed to a set of responses more closely resembling that of the HVS cones. Note that the image RGB planes may be calibrated to any monitor in order to derive the transformation to CIE tristimulus values [27–30].

There are three cone types involved in human color vision each of which contains a pigment of different spectral absorption characteristics. In the literature there exist a number of cone “fundamental” systems (so called Koenig Fundamentals [12]). The two most widely accepted sets of fundamentals are those of Smith and Pokorny and of Vos and Walraven [12]. These two sets of spectral absorption curves are very similar. The main difference is the relative response of the short wavelength response function $V_s(\lambda)$. Buchsbaum has shown that if we form opponent channels following the considerations of redundancy reduction, then for a large range of relative responses of $V_s(\lambda)$, the resulting channels are independent of the actual response [17]. Since our model is based on opponent channels, the choice between these two sets of curves is not critical and we have arbitrarily chosen to use the Vos-Walraven set. These are illustrated in Figure 1. These curves are termed $V_l(\lambda)$, $V_m(\lambda)$, and $V_s(\lambda)$, where l, m, s stand for the long, medium and short wavelength cone responses respectively (often associated with the Red, Green and Blue colors).

The response of the cones is non-linear mainly for very bright and very dark stimuli [12,24,13]. As such this response mainly affects the perception of the very bright and very dark regions of the image and therefore its main influence is on dynamic range. In this model we do not consider dynamic range enhancement [31] and we therefore assume a linear cone response throughout the stimulus range. Thus we need to find the 3×3 transformation matrix, T_{RV} , that transforms RGB

brightness only and in a sense, treat a colored image like an achromatic one. Another approach to color image enhancement has been to transform the red, green and blue monochromatic planes to the perceptually based hue, saturation and intensity space. In this space techniques such as simple global saturation stretching [8], hue dependent saturation histogram equalization [9] and intensity edge enhancement based on saturation [10] have been attempted. A different method based on reflection models has also been used to restore the “naturalness” at object-background boundaries and is used when the color of an object in an image is changed manually using an image editing program [11].

To overcome the lack of an obvious metric space, we have been investigating an opponent color model of the early HVS and the effects of chromatic edge enhancement within the framework of this model. Previously, opponent color models have been based on a constant transform from cone responses to chromatic channels [1,2,12–14]. This transformation has been derived by fitting model parameters to various sets of psychophysical data. This approach has been found to be useful in cases where the models so derived are used for predicting further psychophysical responses. However, when the visual stimulus is more complex than sinusoidal gratings or colored patches, these models do not necessarily reflect accurately the behavior of the HVS.

A possible clue to the purpose of the spatio-chromatic opponent processing in the visual system has been revealed by considering the HVS as an efficient information processing network. Natural scenes contain both spatial and chromatic redundancies. If these redundancies were retained, the limited bandwidth of the channel that transmits retinal signals to higher stages of the visual system, and processing capabilities within these higher visual stages would be unable to cope with the signals. This implies that already in the retina, before the signal is transmitted to higher visual stages, there is a need for redundancy reduction. This supposition has led to modeling various early HVS processes by mechanisms of redundancy reduction. These early processes include center-surround receptive fields [15], their Gaussian profile [16], as well as the opponent color transform [17–19].

In this paper we describe a model that first transforms the “red”, “green”, and “blue” planes of a digital image into a three-dimensional color space similar to that characteristic of the HVS cone response space of Vos-Walraven [12]. Chromatic correlation is then reduced using a multispectral Karhunen-Loéve transform. This results in three new channels which we show to be an achromatic, or luminance channel, and two opponent chromatic channels, red-green and blue-yellow. Empirical results on a wide variety of images show that the achromatic channel contains over 95% of the total image energy. It is spatially broadband as it contains all spatial frequencies from DC upwards [20]. In contrast the chromatic channels are shown to have most of their energy in the low spatial frequency region and to contain negligibly little energy. In spite of the small amount of energy in the chromatic channels, we show that biologically motivated edge enhancement of the chromatic channels results in enhancement not obtainable by achromatic manipulation alone [21]. This is an unexpected result since due to the very little energy of the chromatic channels they are usually assumed to be less important for imparting the perception of image sharpness and consequently are further decimated by sub-sampling especially for coding purposes [22]. Finally we show, by experiment, that our biologically motivated processing scheme provides color detail enhancement not obtainable using other methods.

Spatio-chromatic image enhancement based on a model of human visual information processing

Stuart G. Wolf

Ran Ginosar

Yehoshua Y. Zeevi

November 13, 1997

Abstract

We have developed a technique for color image enhancement based on a model of the Human Visual System (HVS). A color image represented by RGB is first transformed into a color space based on the HVS cone response characteristics. Subsequently, chromatic correlation reduction and energy compression is realized by using a multispectral Karhunen-Loève transform (KLT) of the cone responses. This yields a color opponent space related to the HVS characteristics. Spatial energy distribution is highly skewed: the chromatic channels contain significantly less relative energy than in standard opponent spaces such as YUV . A constant transform, which closely approximates the input-dependent KLT, has been found. In spite of the little spatial energy in the chromatic channels, chromatic edge enhancement in this space does add significantly to the perception of image detail and enhances its chromatic fidelity, as compared with standard edge enhancement of only the luminance channel. Chromatic edge enhancement is also less sensitive to noise than achromatic edge enhancement. Owing to the simplicity of the transform and enhancement processes, they can be employed for processing real-time video signals.

1 Introduction

A basic limitation in the development of color image enhancement algorithms is the lack of a quantitative measure of image quality. Although some attempts have been made in this direction [1–4], it is still unclear what is the appropriate metric space in which to evaluate color image quality. Recently van den Branden Lambrecht [5] proposed a quality metric for color moving pictures based on an opponent-colors theory and a spatio-temporal vision model. Color image quality metrics however remain unclear and one should therefore consider the design of the human visual system (HVS) as a basis for color image enhancement.

Such an approach has been followed for the general enhancement of black and white images where, for example, the importance of edges has led to various edge and other high frequency enhancement techniques [6]. On the other hand, there has been little investigation into the interaction of the HVS with color for the purpose of the design of color image enhancement methods. One example worth mentioning is Faugeras' homomorphic model for color image processing [1,2] in which the structure of the HVS is exploited to achieve color constancy and improve dynamic range.

In general though, there seems to be an agreement that “we can process each of the three monochrome images separately and combine the results” [7], implying that it is sufficient to enhance