Vision, Video and Graphics (2005)
E. Trucco, M. Chantler (Editors)

Automatic non-photorealistic rendering through soft-shading removal: a colour-vision approach

A. Olmos and F. A. A. Kingdom

McGill Vision Research, McGill University, 687 Pine Avenue West, Rm. H4-14 Montreal, Quebec, Canada.

Abstract

This paper presents a non-photorealistic rendering algorithm that produces "stylised-style" images by removing the soft shading from the image and by giving objects extra definition through black outlines. The method of shading removal is based on a model of the architecture of the human colour vision system. Some image results are provided and the possible extension of the algorithm using a back-propagation neural network is discussed.

Categories and Subject Descriptors (according to ACM CCS): I.3.0 [Computer Graphics]: Non photorealistic rendering, reflectance map, non-photorealistic rendering, perceptual model, colour vision.

1. Introduction

In many applications a non-photorealistic rendered (NPR) image can have advantages over a photorealistic one. NPR images may convey information more efficiently [DS02, GRG04] by omitting extraneous detail, focusing attention on relevant features, clarifying, simplifying and disambiguating shape. Brennan's [Bre85] research in caricature began as part of a teleconferencing project where the goal was to represent, and transmit over a limited bandwidth, some of the visual nuances present in face-to-face communication. It was discovered that animated caricatures of faces were more acceptable (in this case) than realistic synthesized images of talking heads, because the caricatures made the degree of abstraction in the image more explicit. In the hands of talented artists [DFRS03], abstraction becomes a tool for effective visual communication. Such abstraction results in an image that directs the observer's attention to its most meaningful places and allows an understanding of the structure of an image without conscious effort [Zek99].

2. Background

There are a vast number of NPR methods in the computer graphics literature. They vary in style and target different aspects of visual appearance, but in general they are closely related to conventional artistic techniques [CAFS97, Hae90, Lit97, GGSC98]. Some approaches have involved the automatic generation of facial cartoons, by training a system that determines which, and how, face components (such as the eyes, nose and mouth) should be altered [CLL04, LCXS02]. Other approaches have involved generating a 3D description of a scene in order to outline the edges of objects and then fill-in the surfaces between outlines with colour [Dec96]. In more recent approaches, models of human perception have been applied to develop more accurate NPR representations. DeCarlo et al. [DS02] developed an elegant and interactive system where the meaningful content of the image was identified just by observation. The beauty of the technique lies in the fact that the human-user simply looks at the image for a short period of time and a perceptual model translates the data gathered from an eye-tracker into predictions about which elements of the image representation carry important information. Gooch et al. [GRG04] presented a method for creating black and white image illustrations from photographs of humans. They evaluated the effectiveness of the resulting images through psychophysical studies which assessed the accuracy and speed of both recognition and learning. Their approach used a model of human brightness perception in order to find the significant edges or strokes that could represent the photograph of a human face. Common to all the methods described here is the selection of appropriate or suggestive contours [DS02] as a means to produce an abstract representation of the scene. We will refer to such contours as "significant" edge contours.

2.1. Our Approach

Our method involves removing the soft shading from natural images and adding black outlines to objects. The shading-removal part of the algorithm [OK04] exploits the constraint that in natural scenes chromatic and luminance variations that are co-aligned arise from changes in surface reflectance, whereas near-pure luminance variations arise from shading and shadows [TFA03, FHD02]. The idea in this algorithm is the initial separation of the image into one luminance and two chromatic image planes that correspond to the 'luminance', 'red-green' and 'blue-yellow' channels of the primate visual system. It has been argued that the luminance, red-green, and blue-yellow channels of the primitive visual system are an efficient way of coding the intensive and spectral content of natural images [Wan95]. The algorithm uses the fact that shading should only be present to a significant degree in the luminance image plane, whereas reflectance changes should appear in all three image planes. In the algorithm the chromatic (red-green and blue-yellow) image planes are analysed to provide a map of the changes in surface reflectance, and this map is then used to reconstruct a reflectance image that incorporates both spectral (colour) and intensive (lightness) components. Overall, the idea exploits the theory that colour facilitates object perception and has an important role in scene segmentation [Kin03] and visual memory [Geg03].

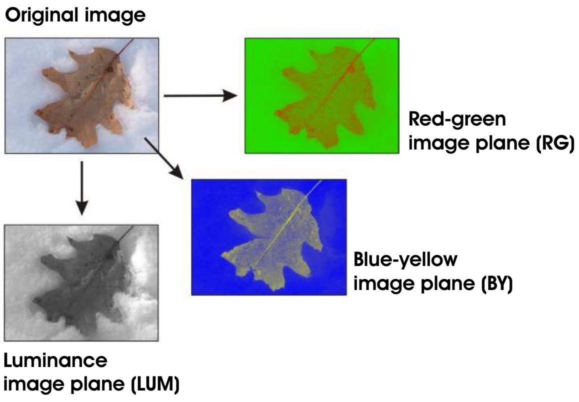

Figure 1: *Modelled responses of the luminance, red-green, and blue-yellow channels of the human visual system to an image. Shading appears in the luminance (LUM) but not in the chromatic planes. A colour version of the images presented in this paper can be found at* `http://ego.psych.mcgill.ca/labs/mvr/Adriana/npr/`.

2.2. The Algorithm

A brief exposition of the algorithm is provided here; full details are provided elsewhere [OK04]. A general overview of the algorithm is given in Figure 2 and described as follows:
1) Starting from an *RGB* image, the images are converted into the *LMS* cone space [RAGS01] (where *L*, *M*, *S* stand for long, middle and short wavelength). This conversion can be achieved by multiplying each *RGB* tristimulus values by a colour space transformation matrix to the *LMS* cone space [RAGS01]. The three post-receptoral channels of the visual system are then computed using the following shadow-removal [PBTM98] pixel-based definitions of the cone inputs:

$$LUM(x,y) = L(x,y) + M(x,y) \quad (1)$$

$$RG(x,y) = \frac{L(x,y) - M(x,y)}{LUM(x,y)} \quad (2)$$

$$BY(x,y) = \frac{S(x,y) - \frac{1}{2}LUM(x,y)}{S(x,y) + \frac{1}{2}LUM(x,y)} \quad (3)$$

where *L*, *M* and *S* are the cone-filtered images and (x,y) pixel coordinates. *LUM*, *RG* and *BY* are respectively the luminance, red-green and blue-yellow image planes. Figure 1 shows the three image planes.

2) Edges are found in each *RG* and *BY* image planes using a Sobel mask [GW02] with a threshold computed as the mean of the gradient magnitude squared. The *RG* and *BY* binary edge maps are then combined using an *OR* operation.

3) In this step the image derivatives of each of the *R*, *G*, *B* image planes are found and classified using the edge map found in the previous step.

4) The classified derivatives are then reintegrated in order to render a colour image without shading. This is achieved using the inverse filtering technique described by Weiss in his study aimed at extracting intrinsic images from image sequences [Wei01]. This process involves finding the pseudo-inverse of an over-constrained system of derivatives. Briefly, if f_x and f_y are the filters used to compute the derivatives in the *x* and *y* directions, and I_x and I_y are the classified reflectance derivatives of the image *I*, the reconstructed image I_r is given by:

$$I_r(x,y) = g * [f_x(-x,-y) * I_x] + (f_y(-x,-y) * I_y) \quad (4)$$

where * denotes convolution, $f_x(-x,-y)$ is a reversed copy of $f_x(x,y)$, and *g* is the solution of:

$$g * [(f_x(-x,-y) * I_x) + (f_y(-x,-y) * I_y)] = \delta. \quad (5)$$

The full colour, reintegrated image is obtained by reintegrating each *R*, *G*, *B* colour plane. The computation can be performed most efficiently using a Fast Fourier Transform. More details about this technique can be found at `http://www.cs.huji.ac.il/~yweiss/`. It is worth mentioning that by simply reintegrating the Luminance (*LUM*) image plane, a gray scale non-photorealistic rendering can be obtained as well.

5) Finally, the edge contours found only in the *RG* image are smoothen and added as black outlines to the rendered objects in the image. This is in accordance with the computer graphics literature [DS02, DFRS03, Dec96] to enhance the cartoon-like appearance. The reason for choosing only the edges in the chromatic *RG* image plane and not the ones in the *BY* image plane, is because the later is more likely to pick up shading contours (for instance, blue shadows due to blue sky-light). We filtered out the small contours (i.e. smaller that 10 pixels). To improve the "stylised-look", small contours (i.e. smaller than 10 pixels) can be filtered out.

The algorithm presented here is similar to the one presented by Olmos and Kingdom [OK04]. The main difference is the goal and in the way the chromatic planes are manipulated. In our previous study [OK04] the goal was to obtain as faithful as possible a representation of both the reflectance and shading maps of natural images. To achieve this, the images were gamma-corrected and the chromatic planes thresholded before further analysis. In the work presented here, we only wanted to find the contour edges without worrying about the fact that some of them might be caused by strong cast shadows and/or strong inter-reflections, because these features enrich a drawing and provide visual feedback about the type of material or object [Dec96]. While it is important to gamma-correct the images for visual display or for psychophysical experimentation, it is arguably not a strong requirement for the conversion of an image from *RGB* to *LMS* space (in the application presented here); this is because the *LMS* axes are not far from the *RGB* axes, failure to gamma-correct the image will produce errors of only about 1 or 3 percent error [RAGS01].

2.3. Results

Figure 3 present some examples of the algorithm applied to various images. The results demonstrate the potential of using a *LUM*, *RG* and *BY* channel decomposition as the basis for the automatic generation of non-photorealistic rendered images. It can be observed in Figure 3B that our algorithm managed to remove the soft shading in the tomato images and the texture detail in the jacket of the person appearing in Figure 3A. Nevertheless, more work would need to be done to remove the content-detail from the background of the image (i.e. Figure 3A) as discussed by DeCarlo et al. [DS02]. On the other hand, as can be seen in Figure 3 D, problems with this algorithm might arise when a significant change in the image is mainly defined in Luminance (the while paw of the soft toy against the white snow; and the brown ribbon against the brown fabric of the soft toy). In order to improve the robustness of the algorithm, future work would involve more sophisticated methods to process the chromatic (*RG* and *BY*) image planes information. One possibility would be to use the two chromatic image planes as inputs to a simple back-propagation network (BPNN) [Hay96] in order to find the significant edge contours. Following this approach, the

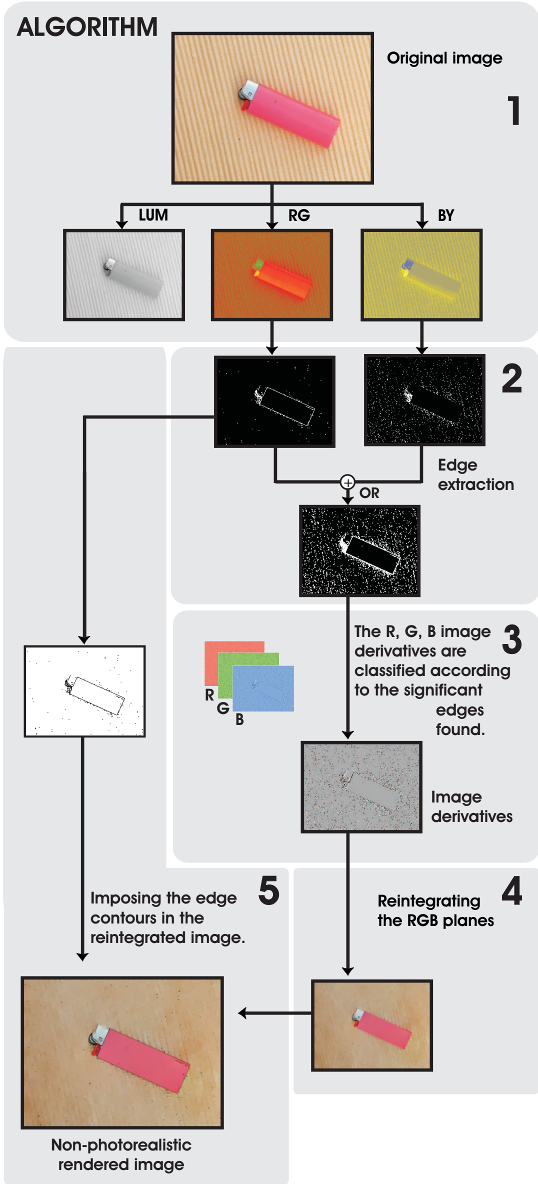

Figure 2: *Flow diagram of the algorithm. 1) computation of the LUM, RG and BY image planes; 2) the edges at each chromatic plant (RG and BY) are computed and combined; 3) the image derivatives are classified; 4) the classified image derivatives are reintegrated and 5) the contour found in both chromatic plane (RG and BY) imposed to finalise the non-photorealistic rendering. The final result can be better appreciate in Figure 3. A colour version of this flow chart can be found at* `http://ego.psych.mcgill.ca/labs/mvr/Adriana/npr/`.

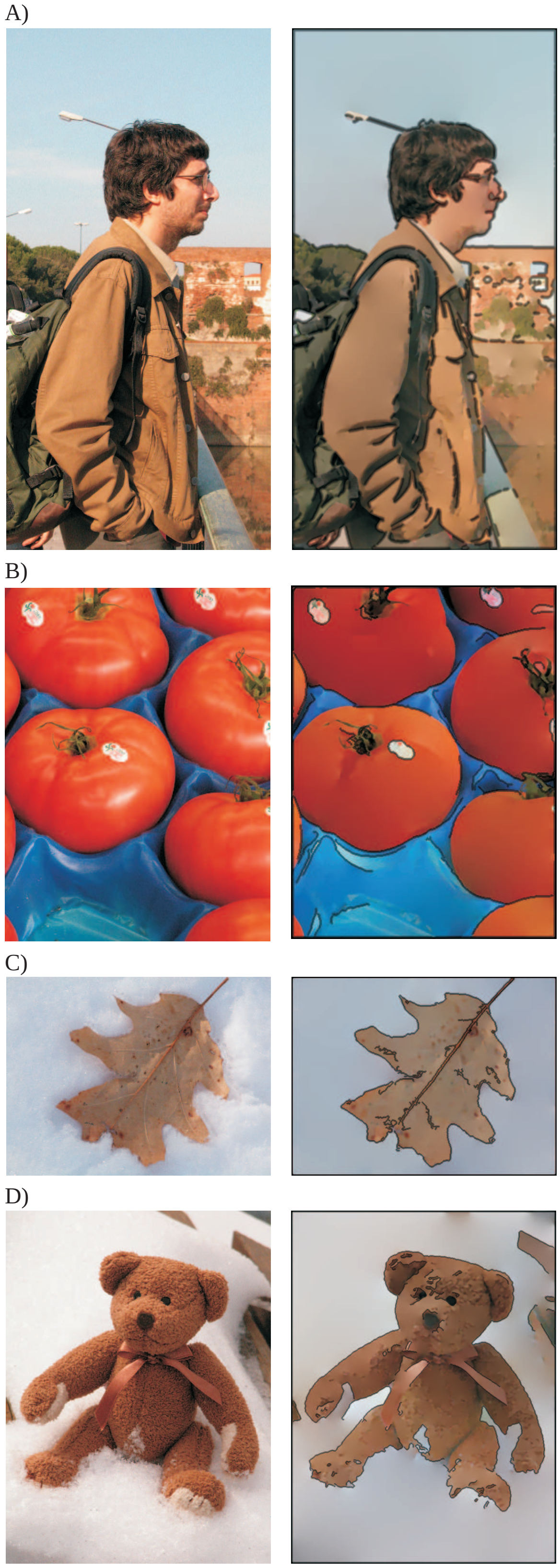

Figure 3: *Image examples of the non-photorealistic rendering algorithm (left) based on a model of the architecture of the human colour vision system. Original images (right) taken from the McGill Colour Vision Database. A colour version of these results can be found at* `http://ego.psych.mcgill.ca/labs/mvr/Adriana/npr/`.

training data could be just a few manually generated cartoon strokes of a natural scene. A quick method for improving the "stylised-look" of the images presented here could be by drawing the outline contours as pencil strokes [Sou02].

2.4. Conclusions

The algorithm and the results presented in this paper represent a potential alternative method for the automatic rendering of non-photorealistic images. The interesting aspect of the algorithm resides in its method of decomposing the image into the modelled responses of the luminance and chromatic channels of the human visual system, as the basis for the removal of soft shading for NPR. We stress at the outset that we make no claims regarding the superiority of our algorithm compared to its predecessors. Our aim here is to explore the feasibility of using a colour perceptual model (related to the three post-receptoral mechanisms of the human visual system) in non-photorealistic rendering, in this case the colour-opponent channels of primate vision, as these channels are likely to play an important role in facilitating object and scene segmentation. Problems with this algorithm might arise when a significant change in the image is mainly defined in Luminance.

2.5. Acknowledgement

We would like to thank the following persons: Nilima Nigam from Mcgill University for her advice on the reconstruction of boundaries; Mark Drew from Simon Fraser University and William Freeman from Massachusetts Institute of Technology for helpful comments on the computation of reflectance and shading maps. Research supported by the Canadian Institute of Health Research grant MOP-11554 given to Fred Kingdom.

References

[Bre85] Brennan S. E.: Caricature generator: The dynammic exaggeration of faces by computer. In *Leonard* (1985), vol. 18, pp. 170–178.

[CAFS97] Curtis C. J., Anderson S. E., Fleischer K. W., Salesin D. H.: Computer-generated watercolor. In *SIGGRAPH 97 Conference Proceedings* (1997), pp. 421–430.

[CLL04] Chiang P.-Y., Liao W.-H., Li T.-Y.: Automatic caricature generation by analysing facial features. In *Proceedings of the Asian conference on Computer Vision, Korea* (2004).

[Dec96] Decaudin P.: *Cartoon Looking Rendering of 3D-Scenes*. INRIA Research Report No. 29219, 1996.

[DFRS03] DeCarlo D., Finkelstein A., Rusinkiewicz S., Santella A.: Suggestive contours for conveying shape. In *SIGGRAPH 2003 Conference Proceedings* (2003), pp. 848–855.

[DS02] DeCarlo D., Santella A.: Stylization and abstraction of photographs. In *SIGGRAPH 2002 Conference Proceedings* (2002), pp. 769–776.

[FHD02] Finlayson G. D., Hordley S. D., Drew M. S.: Removing shadows from images. In *European Conference on Computer Vision, ECCV'02 Proceedings* (2002), vol. 4, pp. 823–836.

[Geg03] Gegenfurtner K. R.: Cortical mechanisms of colour vision. In *Neuroscience: Nature Reviews* (2003), vol. 4, pp. 563–572.

[GGSC98] Gooch A., Gooch B., Shirley P., Cohen E.: A non-photorealistic lighting model for automatic technical illustration. In *SIGGRAPH 1998 Conference Proceedings* (1998), pp. 447–452.

[GRG04] Gooch B., Reinhard E., Gooch A.: Human facial illustrations: Creation and psychophysical evaluation. In *ACM Transactions on Graphics* (2004), pp. 27–44.

[GW02] Gonzales R. C., Woods R. E.: *Digital Image Processing*. 2nd ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

[Hae90] Haeberli P.: Paint by numbers: Abstract image representation. In *SIGGRAPH 90 Conference Proceedings* (1990), pp. 207–214.

[Hay96] Haykin S.: *Neural Networks - a Comprehensibe Foundation*. Prentice Hall, New Jersey, 1996.

[Kin03] Kingdom F. A. A.: Colour brings relief to human vision. In *Nature Neuroscience* (2003), vol. 6, pp. 641–644.

[LCXS02] Liang L., Chen H., Xu Y.-Q., Shum H.-Y.: Example-based caricature generation with exaggeration. In *Conference Proceedings on the Pacific Graphics and Applications* (2002), pp. 386–393.

[Lit97] Litwinowicz P.: Processing images and video for an impressionist effect. In *SIGGRAPH 97 Conference Proceesings* (1997), pp. 151–158.

[OK04] Olmos A., Kingdom F. A. A.: Biologically inspired recovery of shading and reflectance maps in a single image. In *Perception* (2004), vol. 33, pp. 2463–1473.

[PBTM98] Parraga C. A., Brelstaf G., Troscianko T., Moorehead I. R.: Colour and luminance information in natural scenes. In *Journal of the Optical Society of America A* (1998), vol. 15, pp. 563–569.

[RAGS01] Reinhard E., Ashikjmin B., Gooch B., Shirley P.: Color transfer between images. *IEEE Computer and Graphics: Applied Perception 1*, 5 (2001), 34–41.

[Sou02] Sousa M. C.: Observational models of graphite pencil materials. *Computer Graphics Forum 19* (2002), 27–49.

[TFA03] Tappen M., Freeman W., Adelson E.: Recovering intrinsic images from a single image. *Advances in Neural Information Processing Systems, NIPS 15* (2003).

[Wan95] Wandel B. A.: *Fundations of Vision, Chapter 9*. Sinauer: Sunderland, Massachusetts, 1995.

[Wei01] Weiss Y.: Deriving intrinsic images from image sequences. In *Proceedings of the 8th ICCV* (2001), pp. 68–76.

[Zek99] Zeki S.: *Inner Vision: An Exploration of Art and Brain*. Oxford University Press, 1999.