A Perceptual Adaptive Image Metric for Computer Graphics

Jean Philippe Farrugia
LIRIS
Université Claude Bernard Lyon I
Bâtiment Nautibus
8, boulevard Niels Bohr
69622 Villeurbanne Cedex
France
jpfarrug@bat710.univ-lyon1.fr

Stéphane Albin
LISSE
École Nationale Supérieure
des Mines de Saint-Étienne
158, cours Fauriel
42023 Saint-Étienne
France
stephane.albin@emse.fr

Bernard Péroche
LIRIS
Université Claude Bernard Lyon I
Bâtiment Nautibus
8, boulevard Niels Bohr
69622 Villeurbanne Cedex
France
bperoche@liris.cnrs.fr

ABSTRACT

This paper presents two points: a new simple color vision model and an adaptive way to compute an image metric based on a vision model. Metrics are very useful in computer graphics. Applications include perceptually-based rendering or image comparison for photorealism. Usual vision model-based metrics make an expensive use of memory and cpu resources, mainly for two reasons. First, the vision model is a pipeline of non linear functions applying on a multi-scale decomposition of the image. Second, the model is computed for every single pixel of the picture. In this paper, we designed a very simple mono-scale vision model taking into account many perceptual issues like masking effects and adaptation. We also propose an adaptive approach of distance computation : the image plane sample scheme is designed to be denser when distance variation is greater. This method is usable with any vision model and only uses two parameters, making it very easy to configure. By combining it with our simple vision model, it computes a difference map interactively for 512x512 pictures.

Keywords
Image metric, Perception, Vision model, Rendering.

1 INTRODUCTION

Perceptual issues is a fast growing investigation field in computer graphics. In fact, taking into account visual perception in image generation appears to be a natural step, since these images are viewed by human observers. Here are examples of some advantages computer graphics researchers may benefit from visual system knowledge:

- High dynamic range restitution (tone mapping problems);
- Interactive detail decimation on complex geometric models;
- Perceptually based rendering;
- Photorealistic results validation.

To be efficient, an image metric relies on a computational model of the visual human system. This model consists of a linear sequence of psychometric functions. The main drawback of this pipeline approach is the expensive cost of such a vision model: it usually takes several minutes to compute the desired information, mainly because every pixel is sent through the processing pipeline of the model. In this paper, we propose first a very simple model of the visual human system and then an adaptive method rendering a visual difference map by processing only a small percentage of the pixels. The initial sample map is then refined where needed. The refinement criteria is based on the homogeneity of distance values.

This paper will be organized into six sections including this introduction. The second section will briefly present prior work on perceptual issues in computer graphics. The third section is a description of our simple vision model and the fourth section is devoted to our adaptive algorithm to compute an image metric. The fifth section shows some results obtained with this method. Finally, we discuss future improvements in the sixth section.

WSCG proceedings
WSCG'2004, February 2-6, 2004, Plzen, Czech Republic.

2 PREVIOUS WORK ON VISION MODELS

There is a great variety of vision models and image metrics designed in the image processing field (cf. [Ahu93], [FZvdBLS97]), but very few of them were reused in computer graphics. In this section, we will only present the most commonly used in image synthesis. Daly's Visual Differences Predictor [Dal93] was one of the first models tested with synthetic images. In [RWP$^+$95], the authors used a metric inspired by existing compression algorithms to compare radiosity pictures. The algorithm in [Lub95] is an attempt to simulate the visual pathway from the retina to the mammalian cortex. The authors of [PFFG98] improved Lubin's model to take color perception into account.

Opposed to the above models is Ahumada and Beard's Simple Vision Model [AB98]. This model is low cost because it does not use any multi-scale image decomposition and it performs well with luminance images, so we chose it to be the basis of our work on a simple vision model.

3 OUR SIMPLE VISION MODEL

Our model has to be simple enough to be computed interactively. We chose the simple model of Ahumada and Beard [AB98]; we modified it to take color pictures into account and improved its frequency treatment by replacing the double gaussian filtering with a discrete cosine transform. We also added a tone reproduction operator. The steps of this model are depicted in the next section.

3.1 Tone Reproduction

The initial picture is made of three float channels representing luminance in $Cd.m^{-2}$. To take visual adaptation into account, we chose to apply a tone mapping operator to this picture first. We thought that placing tone mapping at the beginning of the process was quite natural, as it allows to take the fact that pictures are visualized on a display device into account. Another advantage is that any tone mapping operator may be used to enhance the abilities of the model. We decided to choose the fast and widely used tone mapping operator by Ward [War94]. After this treatment, we obtain a classical RGB picture I which can be processed via the next steps.

3.2 Color Conversion of Ahumada and Beard's Model

Ahumada and Beard's vision model [AB98] can only handle monochromatic pictures. We had to modify it so that it can manage color pictures. We chose the AC1C2 color space introduced by Meyer [Mey86].

3.3 Image CSF Filtering

We thought that having a multi-scale model would subsequently affect computation time, so we designed a single scale frequency treatment. We achieved our CSF filtering by using a fast cosine transform to obtain frequency content and by multiplying these frequencies by the corresponding coefficient in our bidimensionnal contrast sensitivity functions; after an inverse DCT processing, we obtain the local contrast Co by dividing the initial image I by the result of the DCT filtering F.

$$Co(p) = \frac{I(p)}{F(p)}$$

3.4 Local Contrast Energy

The local contrast energy is obtained from the local contrasts Co via the following formula:

$$E(p) = Co(p)^2 \times F_e(p)$$

where F_e is a low pass Gaussian filter with the same characteristics as in Ahumada and Beard's model.

3.5 Contrast Masking

Contrast masking describes the mutual interaction of picture contrasts between each others. It is obtained via the following transducer:

$$V(p) = \frac{Co(p)}{(1 + (g_e.E(p))^2)^{0.5}} \qquad (1)$$

This step was previously named "local contrast gain" in Ahumada and Beard's model. g_e is the same gain parameter.

3.6 Computing the distance map from this model

The vision models computes visual contrast maps. Now we need to convert the two contrast maps from the pictures we would like to compare in an understandable distance map. As Ahumada and Beard do, we use a Minkowski metric on the output of the model to obtain our distance values.

$$d' = 10.5((V_1(x,y) - V_2(x,y))^4)^{0.25} \qquad (2)$$

This metric is computed on the three color channels: in formula 2, V_1 and V_2 are vectors. A greyscale picture, called the difference map in the sequel, may be obtained by mapping the previous data from $[0, 1]$ to $[0, 255]$.

4 OUR ADAPTIVE ALGORITHM

Our method intends to reduce computing time by processing only a fraction of the picture's pixels. We use an enhancement of the method by Albin*et al*, described in [ARPT02]. To have the vision model computable on a single pixel, we decided to apply the first step of the vision model on the entire image in a pretreatment phase, then compute the following steps on individual pixels.

The image plane is subdivided with a quad tree. For each cell of this quad tree, we will shoot a fixed number N of samples depending on the size of the cell. The distance is computed for each sample and the mean value is affected to the cell. Then we have to decide if this cell should be subdivided or not. We use the homogeneity test described in [ARPT02] as a subdivision criteria. We compute the following expression called "the homegeneity criteria":

$$\frac{\#\{x \in [x - \epsilon; x + \epsilon]\}}{|X|}$$

where ϵ is the threshold distance from the mean value inside the cell and $|X|$ is the number of pixels in the cell.

If this quantity is under a fixed homogeneity percentage, then the cell will be subdivided. The algorithm stops when no cell needs to be subdivided anymore.

5 RESULTS

5.1 Test scene

We decided to test our metric on computer generated pictures, as our metric is designed to integrate a rendering algorithm.

The test scene is a checkerboard with four spheres lying on it. The pictures on the left side are computed by casting 64 rays per pixel, the right ones by casting only 8 rays per pixel.The noise of the projected shadow on the ground is more or less visible depending on the texture of the floor. We test this feature with two different textures, a rough one (figure 1) and a smooth one (figure 2).

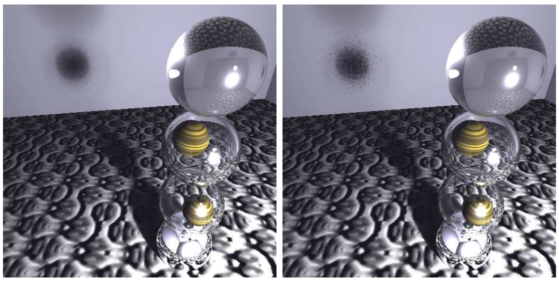

Figure 1: Test 1.a

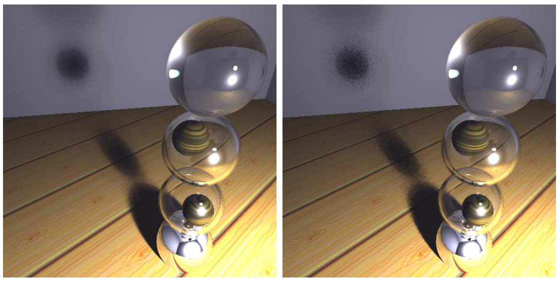

Figure 2: Test 1.b

5.2 Distances Maps and Computation Times

The tables on Figure 3 and 4 give computation times for the test scene. These times are obtained without any function tabulation.

Scene	Time	Mean distance value
Test 1.a	0.54s	0.11
Test 1.b	0.52s	0.15

Figure 3: Results with our model

Scene	Time	Mean distance value
Test 1.a	9.79s	3.5
Test 1.b	9.75s	5.32

Figure 4: Results with Pattanaik's model

5.2.1 The Floor with Spheres Scene

The results on figures 5 and 6 show, for both models, that the mean distance value is higher for the smooth texture scene.

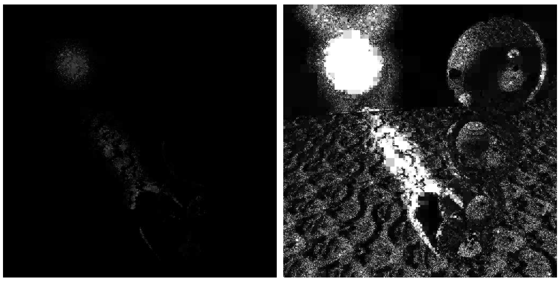

Figure 5: First test

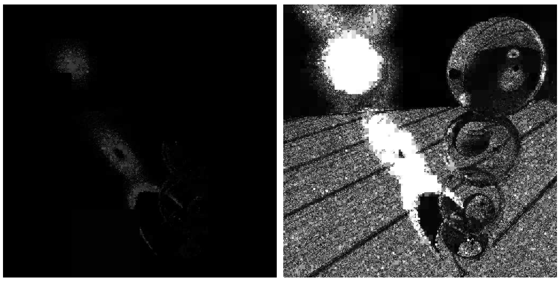

Figure 6: Second test

6 CONCLUSION AND FUTURE WORK

In this paper, we presented a new adaptive perceptually-based image metric. This method is faster than most other image metrics but the results remain significantly accurate for computer graphics, especially for progressive image rendering where only an area information is needed. This precision may easily be tuned by adjusting only two empiric parameters. This method is generic, we used our simplified vision model and Pattanaik *et al*'s one, but the adaptive computation may be used with any other model.

There is much more to do with this work. First, it is obvious that we may not totally validate the model only with the only test scene shown in this paper. We would like to do some more tests with a more important number of human subjects. Second, we would like to design a vision model with a decomposition transform which could be computed adaptively. Therefore, we would not have to compute this decomposition for every pixel before the adaptive process. Finally, this model is ideally adapted to be included in a rendering algorithm: it gives a local information on difference perception, and could guide a progressive rendering to obtain pictures much faster than with empiric criterias.

7 REFERENCES

[AB98] Albert Ahumada and Bettina L. Beard. A simple vision model for inhomogeneous image quality assessment. *SID Digest of Technical Papers*, 29, 1998.

[Ahu93] Albert J. Ahumada. Computational image quality metrics: A review. *Society for Information Display International Symposium Digest of Technical Papers*, 24:305–308, 1993.

[ARPT02] S. Albin, G. Rougeron, B. Peroche, and A. Tremeau. Quality image metrics for synthetic images based on perceptual color differences. *IEEE Transactions on Image Processing*, 11(9):961–971, September 2002.

[Dal93] Scott Daly. The visual differences predictor : an algorithm for the assessment of image fidelity. In A. B. Watson, editor, *Digital Images and Human Vision*, pages 179–206, Cambridge, 1993. MIT Press.

[FZvdBLS97] Joyce E. Farrell, Xuemei Zhang, Christian J. van den Branden Lambrecht, and D. Amnon Silverstein. Image quality metrics based on single and multi-channel models of visual processing. In *Proceedings of the 42nd IEEE Computer Society International Conference (COMPCON'97)*, 1997.

[Lub95] J. Lubin. A visual discrimination model for imaging system design and evaluation. *Vision Models for Target Detection and Recognition*, pages 245–283, 1995.

[Mey86] Gary W. Meyer. Wavelength selection for synthetic image generation. *Computer Graphics and Image Processing*, II:39–44, 1986.

[PFFG98] Sumanta N. Pattanaik, James A. Ferwerda, Mark D. Fairchild, and Donald P. Greenberg. A multiscale model of adaptation and spatial vision for realistic image display. In *Computer Graphics (SIGGRAPH 98 Conference Proceedings)*, volume 32, pages 287–298, July 1998.

[RWP+95] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust. Comparing real and synthetic images : some ideas about metrics. In *6th Eurographics Workshop on Rendering, Dublin*, pages 213–222, June 1995.

[War94] Greg Ward. A contrast-based scalefactor for luminance display. In Paul Heckbert, editor, *Graphics Gems IV*, pages 415–421. Academic Press, Boston, 1994.