Quality Image Metrics for Synthetic Images Based on Perceptual Color Differences

Stephane Albin, Gilles Rougeron, Bernard Péroche, and Alain Trémeau

Abstract—Due to the improvement of image rendering processes, and the increasing importance of quantitative comparisons among synthetic color images, it is essential to define perceptually based metrics which enable to objectively assess the visual quality of digital simulations. In response to this need, this paper proposes a new methodology for the determination of an objective image quality metric, and gives an answer to this problem through three metrics. This methodology is based on the LLAB color space for perception of color in complex images, a recent modification of the CIELab1976 color space. The first metric proposed is a pixel by pixel metric which introduces a local distance map between two images. The second metric associates, to a pair of images, a global value. Finally, the third metric uses a recursive subdivision of the images to obtain an adaptive distance map, rougher but less expensive to compute than the first method.

Index Terms—Colorimetry, difference metric, image quality, rendering, synthetic images, visual perception.

I. INTRODUCTION

THE computation of realistic images is composed of two main steps. The first one consists of physically based calculations, where the flow of energy is modeled as accurately as possible. This step gives the distribution of light at each point in the scene. The second step is a display process, where the results of the previous computation are transformed to be presented on a display device. This is a perceptually based step, where the objective is to satisfy the observer.

It is clear that the final stage of the rendering process is reached when the image is viewed and judged for suitability by a human observer. Thus, there is a great need to evaluate visual simulations, in particular for the following problems [1].

- To validate simulations against measurements. This problem appears, for example, in domains such as lighting calculations for indoor or outdoor architecture, street lamps design, vehicle lights design, etc.

- To compare the results of simulation methods. This happens, for example, if different rendering algorithms are used with the same scene (cf. [2]). It may also be to evaluate the importance of the various parameters of one such rendering algorithm and to facilitate the choice of its parameters.

- To guide progressive image synthesis calculations more efficiently, for example with a radiosity algorithm [3] or with a ray tracing one [4]. In this case, the idea is to accurately compute the features of the rendering solution that are perceptually important. A first step in this direction appeared in [5] or in [6].

In all the cases briefly described just above, perceptually based image metrics are needed. Any error metric based on radiometric comparisons cannot guarantee that additional errors will not be introduced during the display process. Differences in luminance values may in fact be undetectable after the display transform has been performed. This fact is all the more important because display devices currently in use such as monitors or head-mounted displays are far from perfect, with a reduced color gamut, and a limited dynamic range.

Tools coming from digital image processing are not necessarily well adapted, because they only deal with RGB pictures whose origin is always unknown. In particular, the mean squared error or the root mean squared error, often used in this domain, are not adequate measures. Comparisons are only based on corresponding pixels and do not include any knowledge about the visual human system and the underlying features in the picture.

In this paper, we have investigated three ways to define perceptually based metrics in computer graphics by using the LLAB color space. The first way is a local pixel by pixel metric between images which allows to define a distance map. The second way associates with a pair of images a global distance value whose purpose is to define a kind of metric between the two images. The last way uses a recursive subdivision of the images related to the value defined in the second way to define a distance map which is rougher than in the first way, but less computationally expensive.

The objective of this study is not to suggest a new metric which would take a maximum of visual characteristics into account, but to develop a metric which enables to assess the accuracy (and if possible the efficiency) of the result of a rendering algorithm in terms of visual aspect. The usefulness of such an approach, rather than another one more commonly used, is linked to the images studied. In our case, images can be described as a set of shiny or matte surfaces, smooth or slightly rough surfaces, interacting with light, shape and shadow. For example, whereas most of image quality models take into account the contrast sensitivity function (CSF) of the visual human system to adjust contrast values according

S. Albin and G. Rougeron are with the Laboratoire d'Images de Synthèse de Saint-Étienne (LISSE), Ecole des Mines de Saint-Étienne, 42023 Saint-Étienne, France.

B. Péroche is with the Laboratoire d'Images de Synthèse de Saint-Étienne (LISSE), Ecole des Mines de Saint-Étienne, 42023 Saint-Étienne, France. He is also with the Laboratoire d'Informatique Graphique, Image et Modélization (LIGIM), Université Claude Bernard—Lyon 1, 69 622 Villeurbanne Cedex, France (e-mail: bperoche@ligim.univ-lyon1.fr).

A. Trémeau is with the Laboratoire d'Informatique Graphique et d'Ingénierie de la Vision (LIGIV), Université Jean Monnet—Saint-Étienne, 42007 Saint-Étienne, France (e-mail: tremeau@vision.univ-st-etienne.fr).

Publisher Item Identifier 10.1109/TIP.2002.802544.

to spatial frequency, our metrics does not: in our study, only contours generate high spatial frequencies, and these contours are useless in our image quality metric as they do not vary from one rendering process to another one. One of the assumptions used in this paper is that color contrast attributes between objects in a scene play an important role in image quality visual assessment, because the visual appearance of an object depends, not only of its own characteristics, but also of the characteristics of objects surrounding it [7]–[9].

The remainder of the paper is organized as follows. In Section II, we remind some previous works on the problem of the evaluation of a metric between images. In Section III, we describe structures which allow to compute the color contrast of an object from its surrounding elements. Section IV introduces some tools which will be fundamental for our algorithm. The algorithm is described in Section V, and some results are discussed in Section VI. Finally, a conclusion and some further developments are given in Section VII.

II. PREVIOUS WORK

In the domain of digital image processing, a numerous literature deals with the problem of the evaluation of a metric between images ([10] and [11] for still monochrome pictures, [12]). Several techniques have been developed to give a quantified answer to themes standard in this field, like image compression, color quantization or search in image databases, for example [13]–[16].

In computer graphics, only few works deal with this problem, even if the definition of a metric turned out to be very useful. In [17], the authors conducted experimental verification between the simulation of a scene by a rendering algorithm such as radiosity and a model of the same scene. In a first step, a radiometric comparison is presented. A radiosity algorithm is used to generate synthetic computer images. An experimental apparatus allows to make measurements of radiant energy flux density. These measures are then compared with the predictions of the radiosity method. In a second step, perceptual comparisons are made. Color science methods are used to create a color television image from the output of the radiosity method. This picture is then compared by a group of experimental subjects against a real model as seen through the back of a view camera. This allowed to check the psycho-visual validity of the simulation.

Paper [18] is the first one which tried to introduce a metric between images in computer graphics. The authors used three algorithms stemming from digital image coding literature, and which are based on Fourier transform and filtering. Rushmeier *et al.* made comparisons on luminance images, not on displayed ones. These images were coming, on one hand, from a CCD based system put in front of a given scene, and, on the other hand, from simulations of the same scene obtained by several rendering algorithms (flat shading, Radiance ray tracing with varying levels of quality specified). For appreciating the performance of a metric algorithm, Rushmeier *et al.* introduced five rather pragmatic evaluation criteria.

In [19], a wavelet based perceptual metric is proposed, whose purpose is very similar to ours. The proposed metric has the ability to measure variations in images at specific locations,

orientations and scales. It has also the ability to discern intensity functions of varying degree of smoothness in the image. The method begins by an orthogonal wavelet transform. A detection of coherent structures follows that allows to detect significant structures in an image. Then, the coefficients of the wavelet transform are modulated with a contrast sensitivity function which measures the response of the human vision system to different frequencies. Finally, the images are compared in the mean square sense.

In [6], the perceptually based visual differences predictor developed by Daly [20] is used to monitor the perceived quality of two rendering algorithms used in computer graphics: progressive radiosity and a Monte Carlo algorithm.

More recently, a number of accurate and efficient metrics based on human perception have been developed for realistic image synthesis [21], [22], such as the visual difference metric proposed by Bolin *et al.* [23], or the visual discrimination metric proposed by Lubin *et al.* [24].

As the visual human system has a varying sensitivity to error that is based upon the viewing context, these metrics used quality descriptors that take into account high background illumination levels, luminance and chrominance of objects, high spatial frequency, and high contrast features (visual masking). An important feature of some of these metrics is that they handle luminance-dependent features and spatially-dependent features independently, such as in [21]. The metric proposed by Bolin *et al.* in [23] is simpler than those introduced by Lubin *et al.* [24]; moreover, it uses a Haar wavelet basis for the cortical transform and a less severe spatial pooling operation. One advantage of these metrics is that they deal with color images.

Other works of general purpose, concerning image quality metrics based on visual models, have also been published in recent years [25], [26]. We must also mention some work done by Watson who has taken into account viewing distance effects, contrast masking effects, and other cognitive effects linked to early stages of the human visual system [27], [28].

III. PRINCIPLES SUBJACENT TO A LOCAL ANALYSIS

Perceived differences between synthetic images are due to highlight smoothing effects, jaggedness effects on illuminated surfaces, ghost figures from object shadows, and other effects less emphasized. They reflect the degree of correspondence of images displayed to *memorized reality* (see experiments done in [29]). That is to say the perceptual quality of an image depends closely to its naturalness. This basic assumption may be followed by a second one according to which background illumination level, and luminance and chrominance contrast attributes between objects in a scene, play an important role in image quality visual assessment: the visual appearance of an object depends not only of its own characteristics, but also of the characteristics of objects surrounding it [7]–[9]. It is therefore essential to

- isolate each object of the scene from elements which surround it;
- evaluate the color contrast between any object to elements which surround it.

In order to reach this objective, we have to use

- an initial segmentation step, which can be easily performed when geometrical features are available to strengthen color features; otherwise it is necessary to use more sophisticated segmentation process [30];
- an adjacency graph construction step, which can be easily achieved when adjacency relationships are described by a linear model [31], [32].

In order to record all adjacency relationships between all the objects in a scene, i.e., all color contrasts between regions, we have considered two similar structures: the region adjacency graph (RAG) and its line-graph, which can be illustrated by Fig. 1 and defined by the following principle (cf. Fig. 2 for an example).

The RAG associates a vertex with each region and an edge with each pair of adjacent regions. At each vertex v_i corresponds therefore a region R_i and two color values $\vec{\mu_i}$ and σ_i representative of the color distribution of this region. At each edge e_{ij} corresponds a pair of adjacent regions (R_i, R_j) and a color distance $d(R_i, R_j)$ which differentiates colorimetrically these two regions. The associated line-graph (LG) is defined as follows: its vertices are the edges of the RAG and its edges the adjacency relations between the edges of the RAG (i.e., two vertices v_i and v_j of LG are connected if the edges of the RAG, represented by v_i and v_j in LG, are adjacent).

Thanks to these structures, it will be possible, from a theoretical point of view, to compute the color contrast of each object from its surrounding elements. Nevertheless, as there is no order relation between colors (from a colorimetric point of view), we are limited to compute only color contrasts between pairs of adjacent objects.

Let us recall that our purpose is to quantify perceived differences between color images computed by a rendering algorithm. Consequently, for these images, it is more accurate, in terms of perceived quality, to use local color attributes than color attributes between neighbor regions, because the observer is more sensible to color contrasts inside objects than to color contrasts between objects. That is to say, the local rendering of objects is the main perceived attribute of quality of synthetic images. In order to analyze the local rendering of each image area, we will propose in Section IV-A to use a neighborhood operator based on the principle of focus of attention. Then, to analyze the difference of local rendering between two images, we will propose in Section IV-B to use the LLAB color distance based on the principle of color appearance measurement.

IV. Some Tools

In this section, we shall present three concepts used by our algorithm: the visual field, the *LLAB* color space, and a refinement process of computation of image differences.

A. Visual Field

The visual field is subdivided into two areas (see Fig. 3):

- the *focus*, which is a visual field of 2° associated with the foveal vision [33];
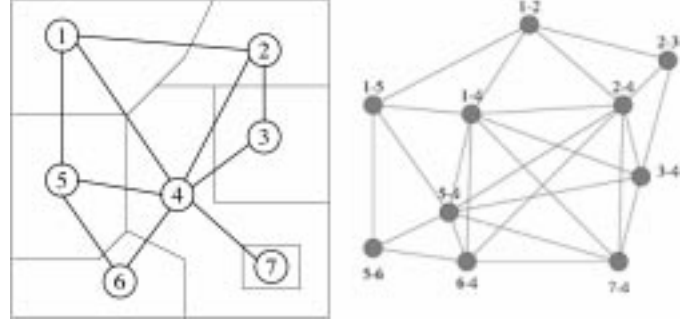

Fig. 1. (a) Region adjacency graph and (b) line-graph associated. Adjacency relationships between regions. For example, if we consider adjacent regions (R_1 and R_4) and (R_4 and R_7), we can see than the edge e_{14} relies effectively the two vertices v_1 and v_4, meanwhile than the edge e_{47} relies two vertices v_4 and v_7. Consequently, these two edge are adjacent relatively to the vertex v_4.
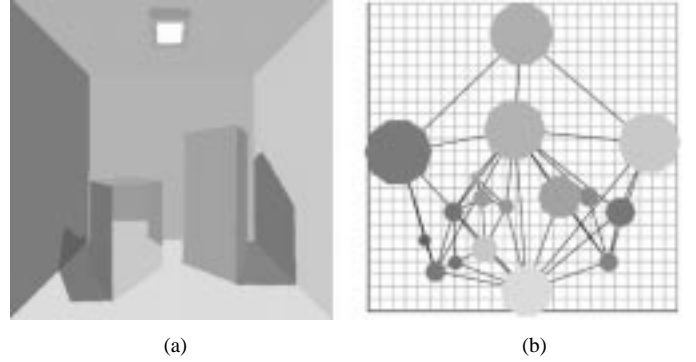

(a)   (b)

Fig. 2. Color contrasts between pairs of adjacent objects of the "Cornell-box." (a) Image Cornell-box segmented and (b) adjacency relationships between regions.
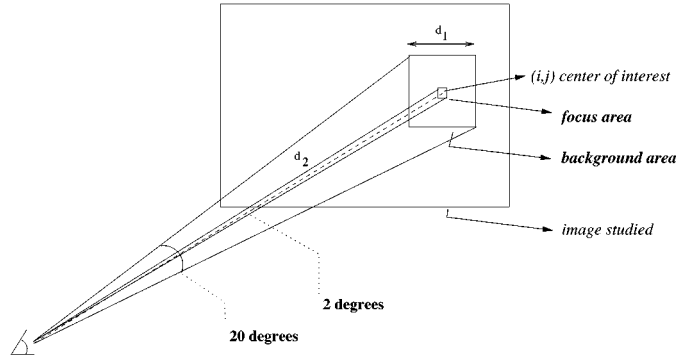

Fig. 3. Subdivision of the visual field in two rectangular masks centered on the (i, j) pixel location.

- the *background*, which corresponds to a field of view of 20° [34].

Two masks are associated with each pixel (i, j) of the picture [35].

- A rectangular area $f(i, j)$, of aperture 2°, located around pixel (i, j), with nl_2 pixels on the left, nr_2 pixels on the right, nt_2 pixels on the top and nb_2 pixels on the bottom. This area corresponds to the focus.
- A rectangular area $b(i, j)$, of aperture 20°, located around pixel (i, j), with, respectively, $nl_{20}, nr_{20}, nt_{20}$, and nb_{20} pixels on the left, on the right, on the top, and on the bottom. This area corresponds to the background.

To compute these masks, the following formulas are used:

$$nl = i - \frac{1}{2}\left(\frac{2\,i\,n_{pr}\tan\frac{\alpha_h}{2} - n_{pr}^2\tan\frac{\alpha_p}{2}}{n_{pr}\tan\frac{\alpha_h}{2} + 2\,i\tan\frac{\alpha_h}{2}^2\tan\frac{\alpha_p}{2}}\right)$$

$$nr = \frac{1}{2}\left(\frac{2\,j\,n_{pr}\tan\frac{\alpha_h}{2} + n_{pr}^2\tan\frac{\alpha_p}{2}}{n_{pr}\tan\frac{\alpha_h}{2} - 2\,j\tan\frac{\alpha_h}{2}^2\tan\frac{\alpha_p}{2}}\right) - i$$

$$nt = j - \frac{1}{2}\left(\frac{2\,i\,n_{pc}\tan\frac{\alpha_v}{2} - n_{pc}^2\tan\frac{\alpha_p}{2}}{n_{pc}\tan\frac{\alpha_v}{2} + 2\,i\tan\frac{\alpha_v}{2}^2\tan\frac{\alpha_p}{2}}\right)$$

$$nb = \frac{1}{2}\left(\frac{2\,j\,n_{pc}\tan\frac{\alpha_v}{2} + n_{pc}^2\tan\frac{\alpha_p}{2}}{n_{pc}\tan\frac{\alpha_v}{2} - 2\,j\tan\frac{\alpha_v}{2}^2\tan\frac{\alpha_p}{2}}\right) - j \quad (1)$$

where

- i and j are the coordinates of the pixel in relation to the center of the screen;
- α_p is the field of view of the area to be computed ($\alpha_p = 2$ or $20°$);
- α_h and α_v are the horizontal and vertical apertures;
- n_{pr} and n_{pc} are the number of pixels per row or per column.

Visual acuity is roughly decreasing in $1/\theta$ [10], [36], where θ is the angle in relation to the visual axis. This relation is estimated by the following weight distribution (cf. Fig. 4):

- a constant weight equal to one is given to each pixel belonging to the focus area;
- a weight linearly decreasing from 1 to 0 is given to pixels located respectively at the center of the background and at its boundary.

We shall denote by w_f the sum of the weights associated to the focus, and by w_b the sum of the weights associated to the background. On the borders of the picture, the masks are clipped and the values of w_f and w_b are updated accordingly.

B. A Color Distance Computed From the LLAB Space

The *LLAB* space [37], [38], derived from the *CIE Lab* 1976 color space, has been settled [39]:

- to give a precise prediction of color appearance between a pair of complex images under different viewing conditions [9], [40];
- to provide a uniform color space for color gamut mapping and color difference evaluation;
- to include single mathematical equations and to easily derive its reverse model.

The use of the *LLAB* space needs two steps. The first one is a BFD chromatic adaptation transform [41]. It is used to convert tristimulus values $X_sY_sZ_s$ of the surface under any source illuminant to tristimulus values $X_rY_rZ_r$ under the reference illuminant $D6500$.

The second step is a computation, modified with regard to the color space *CIELab1976* color space, of the following perceived attributes: lightness (L_L), redness–greenness (A_L), yellowness–blueness (B_L), colorfulness (C_L), hue angle (h_L) and hue composition (H_L).

Let us remind that the computation of these attributes is related to the following experiment: a colored surface with a vision angle of $2°$ is surrounded by a uniform achromatic
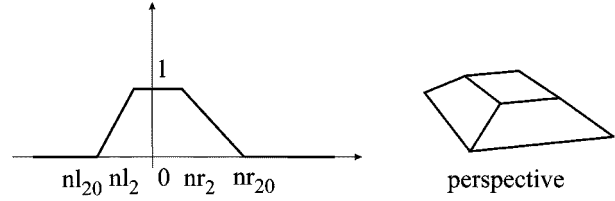

Fig. 4. Weight distribution $w(i, j)$ associated with pixel (i, j) location.

background. Let us denote by L the luminance (expressed in $lum\,m^{-2}\,sr^{-1}$) of the reference source (here $D6500$), and by Y_b the lightness of the achromatic background surrounding the colored surface.

Three parameters: F_S, F_L, and F_C are used to perform the computations. Their values, given in Table I, are dependent upon the ratio $Yb/100$. Let $X_N = 95.05$, $Y_N = 100.0$ and $Z_N = 108.88$ be the tristimulus values of reference white $D6500$. The various appearance attributes are obtained by the following formulas:

$$L_L = 116\,f(Y)^z - 16$$
$$A = 500\,(f(X) - f(Y))$$
$$B = 200\,(f(Y) - f(Z))$$

where $z = 1 + F_L\sqrt{Yb/100}$.
If X_r/X_N, Y_r/Y_N or $Z_r/Z_N > 0.008\,856$,

$$f(X) = (X_r/X_N)^{1/F_S}$$
$$f(Y) = (Y_r/X_N)^{1/F_S}$$
$$f(Z) = (Z_r/X_N)^{1/F_S}.$$

If X_r/X_N, Y_r/Y_N or $Z_r/Z_N \le 0.008\,856$,

$$f(X) = \left(0.008\,856^{1/F_S} - \frac{16}{116}\right)\Big/0.008\,85 * X_r/X_N + \frac{16}{116}$$

$$f(Y) = \left(0.008\,856^{1/F_S} - \frac{16}{116}\right)\Big/0.008\,85 * Y_r/Y_N + \frac{16}{116}$$

$$f(Z) = \left(0.008\,856^{1/F_S} - \frac{16}{116}\right)\Big/0.008\,85 * Z_r/Z_N + \frac{16}{116}$$

$$C_L = (4.907 + 0.162C + 10.92 \\ \cdot \ln(0.638 + 0.072\,16C))F_C S_C$$

where

$$C = \sqrt{A^2 + B^2},$$
$$S_C = 1.0 + 0.47\log L - 0.057(\log L)^2$$
$$h_L = \arctan 2(b/a)$$

where $\arctan 2$ is the function which returns an angle between 0 and $360°$

$$A_L = C_L\cos(h_L)$$
$$B_L = C_L\sin(h_L)$$
$$H_L = H_{L1} + (H_{L2} - H_{L1})(h_L - h_{L1})/(h_{L2} - h_{L1})$$

where h_{L1} and h_{L2} are the hue angles having nearest lower and higher values of h_L respectively.

TABLE I
VALUES OF PARAMETERS F_S, F_L, AND F_C

	F_S	F_L	F_C
Average surround $(20 \leq Yb/100)$	3.0	1.0	1.00
Dim surround $(1 \leq Yb/100 \leq 20)$	3.5	1.0	1.15
Dark surround $(Yb/100 \leq 1)$	4.2	1.0	0.95

A notion of colorimetric distance is defined as follows. Let us suppose we deal with two colored surfaces, each one being surrounded by an achromatic background, and hit by any two illuminants. Tristimulus values $X_r Y_r Z_r$ are first computed, in order to adapt the lighting conditions to what should be given under reference illuminant $D6500$. Then, for each surface, lightness (L_L), chromaticity (C_L) and hue angle (h_L) are computed. The color difference is then:

$$\Delta E_L = \sqrt{(\Delta L_L)^2 + (\Delta C_L)^2 + (\Delta H_L)^2} \qquad (2)$$

where:

$$\Delta L_L = L_2 - L_1$$
$$\Delta C_L = C_2 - C_1$$
$$\Delta H_L = 2\sqrt{C_2 C_1} \sin((h_2 - h_1)/2).$$

C. Refinement of the Study Area by Segmentation

As explained in the previous section, the computation of the various appearance attributes associated with the *LLAB* color space corresponds to a situation where a uniform target placed in front of a uniform achromatic background is shown to an observer. But, study areas defined in Section IV-A contain a part of the scene where it is possible to find some objects with various appearances (for example, cf. Fig. 5). In order to get closer to the experimental situation, we shall suppose that, in the focus area, the visual attention of the observer leads him to bring out the object aimed at the direction of the pixel from the rest of the scene, which then makes up the background.

It is thus necessary that the rendering algorithm used to produce the picture gives to the metric algorithm a segmentation of the image. In our case, the ray tracing algorithm used generates, besides a computed image, a file containing for each pixel the number of the object and/or the number of the face hit by the primary ray (but such a result can also be obtained with the item buffer technique).

With this information, we may refine the study area. For that purpose, we subdivide it into two areas (for example, cf. Fig. 6):

- the *target,* composed of all the pixels belonging to the focus and having the same object or face number than the central pixel;
- the *surround,* composed of all the pixels belonging to the background plus all the pixels from the focus which have not been put in the target.

At the end of this segmentation step, the values of w_f and w_b are updated in order to represent, respectively, the sum of the weights of the target and the sum of the weights of the surround.
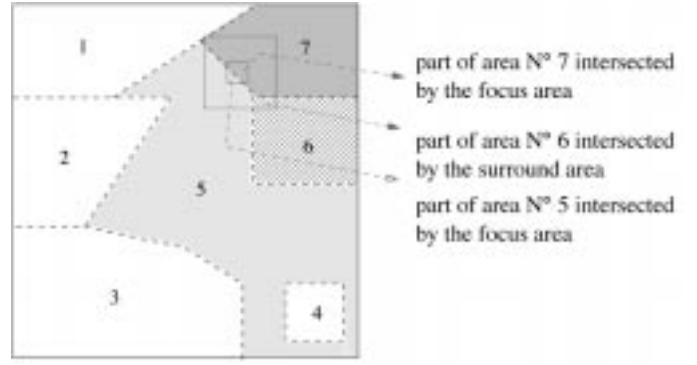

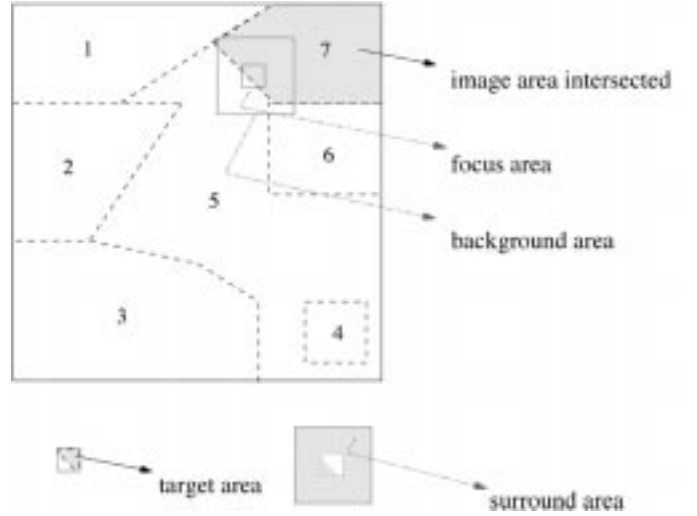Fig. 5. Example of study areas partly intersected by a focus area or by its background.

Fig. 6. Example of a study area defined from a focus area subdivised into a target area and its surround.

V. OUR METHOD

The metric computation programs that we propose take as input two computed images, represented by two files img1.lum and img2.lum. In these files, each pixel is associated to three floating numbers. These numbers are supposed to be CIE 1931 RGB tristimulus values. This last color space has been defined at the end of an equalization experiment brought with the following monochromatic primaries: 700.0, 546.1, and 435.8 nm [42].

Thus, there is a preprocessing step to first obtain XYZ tristimulus values.

A. RGB to XYZ Conversion

The matrix transformation used is

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.49 & 0.31 & 0.20 \\ 0.17697 & 0.81240 & 0.01063 \\ 0.00 & 0.01 & 0.99 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (3)$$

Let us note that the rendering computation, leading to the images that are going to be compared, assigns to the light source tristimulus values $R = 1$, $G = 1$ and $B = 1$. Obviously, after transformation, we also have $X = 1$, $Y = 1$ and $Z = 1$. We shall suppose, to avoid a chromatic adaptation computation, that

the source located in the scene is of type $D6500$. By the inverse transform, the tristimulus values of this source must be: $R = 0.841$, $G = 1.089$ and $B = 1.033$. As this source is the only one present in our test scenes, it is sufficient to respectively multiply RGB tristimulus values coming from files $.lum$ by these last three numbers, and then to apply them the transform matrix.

So, the light source supposed to send out a power of 100 W is given a luminous power $L = 6.8310^5 \, lum \, m^{-2} sr^{-1}$.

B. Computation of a Distance Map Based on Visual Characteristics

The purpose of our method is to compare the sensations of two virtual observers supposed to be present in the scene. We make the hypothesis that their visual attention is equally attracted in all directions.

We proceed as follows, for each pixel of images 1 and 2.

1) The sizes of the study areas associated with vision angles of 2° and 20° are computed with equation (1).
2) Weights are attached to each pixel according they belong to the focus or to the background.
3) The sum of weights w_f and w_b are computed.
4) The focus is refined by separating the target from the surround.
5) Values of w_f and w_b are updated.

   We may notice that the results coming from these first five steps are identical for the two images. Then, for each pixel of each image:

6) Y_b is computed as the weighted average of coordinates $Y(i, j)$ of pixels belonging to the background area

$$Y_b = \frac{1}{w_b} \sum_{(i, j) \in \text{background}} w(i, j) . 100 . \frac{Y(i, j)}{L}.$$

7) Depending on the value of ratio $Yb/100$, values of F_S, F_L, and F_C are computed.
8) For each pixel of the target area, coordinates L_L, C_L and h_L are computed.
9) The pixel by pixel error ΔE_L between images 1 and 2 can therefore been computed, for each pixel of the target, by formula (2);[1]
10) The mean error is then given to the central pixel of the target, i.e.,

$$dist(i, j) = \frac{1}{w_f} \sum_{(i, j) \in \text{target}} \Delta E_L.$$

As output, our program provides an image of distances, as a file $.dist$ which contains, for each pixel, a real number corresponding to the computed distance. To allow the visualization of areas with low or high errors, we define a displayable image of distances. From files $.dist$, false grayscale images are built. With two thresholds defined by the user and called $imperceptibility\_threshold$ and $acceptability\_threshold$, gray levels are assigned as follows:

---

[1]Whatever the value of these errors, these will be all the more perceptible to the observer that they appear in the focus area, as the acuity is maximal in the focus area unlike to the background area; that is the reason why errors are computed only in the target area.

- If $dist(i, j) < imperceptibility\_threshold$, then Red = Green = Blue = 0;
- If $imperceptibility\_threshold \leq dist(i, j) \leq acceptability\_threshold$, then $0 \leq$ Red = Green = Blue ≤ 255, with a linear interpolation;
- If $dist(i, j) > acceptability\_threshold$, Red = Green = Blue = 255.

C. Computation of a Distance Map Based on a Color Dispersion Measure

Different descriptors can be used to measure the degree of homogeneity of an image area [7], [9]:

- either to measure its spatiocolor dispersion (i.e., the local color contrast). To measure locally this dispersion, we propose to compute the *standard deviation of colors* $\sigma_f(i, j)$, of each image area $f(i, j)$ centered around pixel (i, j), defined by (see [43])

$$\sigma_f^2(i, j) = \frac{1}{Card \, f(i, j)} \sum_{(i', j') \in f(i, j)} \|\vec{c}(i', j') - \vec{\mu}_f(i, j)\|^2$$

where $\vec{c}(i, j)$ is the color components vector of pixel (i, j) and

$$\vec{\mu}_f(i, j) = \frac{1}{Card \, f(i, j)} \sum_{(i', j') \in f(i, j)} \vec{c}(i', j');$$

- either to measure its color dispersion according to its principal component, especially the standard deviation of the most representative color gamut of the area under study (cf. [44]). To locally measure this dispersion, we propose first to compute the KARHUNEN LŒVE transformation corresponding to the image area considered [45], [46], then to select the "most significant" color feature among features given by this transformation, i.e., the feature which gives the highest eigenvalue of the covariance matrix of colors distributions in this area, and lastly to compute the standard deviation of this feature.

Two measures can be used to measure, locally, the difference of homogeneity between image areas in relation to the principal color feature. They are

- the FISHER distance, defined locally by

$$D_{Fish}(i, j) = 1 - \frac{d_{Fish}(i, j)}{\max_{(i', j') \in f(i, j)} d_{Fish}(i', j')}$$

if $\varsigma^1(i, j)^2 \leq \varepsilon$ and $\varsigma^2(i, j)^2 \leq \varepsilon$

$$d_{Fish}(i, j) = |\mu^1(i, j) - \mu^2(i, j)|$$

otherwise

$$d_{Fish}(i, j) = Card \, f(i, j) \left(\frac{|\mu^1(i, j) - \mu^2(i, j)|^2}{\varsigma^1(i, j)^2 + \varsigma^2(i, j)^2} \right)$$

where μ^1 and ς^1 are respectively the mean and the standard deviation of the most significant color feature of image area $f(i, j)$ centered around pixel (i, j), computed for image 1. Likewise, μ^2 and ς^2 are computed for the second image to be compared to the first one.

- the normalized mean squared error measure [47], defined locally by

$$NMSE(i, j) = 1 - \frac{\sum\limits_{(i', j') \in f(i, j)} \left| \xi^1(i', j') - \xi^2(i', j') \right|^2}{\sum\limits_{(i', j') \in f(i, j)} \xi^1(i', j')^2}$$

where $\xi^1(i', j')$ is the color coordinate of pixel $(i', j') \in f(i, j)$ defined, for image 1, according to the most significant color feature of image area $f(i, j)$ centered around pixel (i, j). Likewise, $\xi^2(i', j')$ for image 2.

D. Computation of a Global Distance

Of course, we may compute the mean distance from the file *.dist* computed in Section V-B. But, as we shall see in the results, the computation time is then very high.

Thus, we propose another solution, based on a Monte Carlo process. The method is as follows: we randomly choose n pixels (i, j) and for each such pixel, we compute $dist(i, j)$ as defined in Section V-B. Finally, the global distance d between the two images is defined as $d = (1/n) \sum_n dist(i, j)$.

From our experience, a value n about 1000 or 2000 gives a global distance with an error less than 2% in relation to a complete computation.

E. Adaptive Computation of a Distance Map

The purpose of this method is to obtain a distance map showing more or less rough areas of the pictures with low or high errors.

Let α be a real number in [0, 100] and k be a real number. We suppose that the pictures have a $2^n \times 2^n$ pixels size to make the calculations easier:

1) by the method explained in Section V-D, a value d is computed for the pair of pictures;
2) if at most $(1 - \alpha)$ percent values $d(i, j)$ do not belong to interval $[d - k, d + k]$, then subdivide the pictures in four equal subimages and go to 1;
3) otherwise, assign value d to the area.

At the end of this algorithm, a quadtree data structure has been defined, where each leaf is assigned a value d_i. We may use exactly the same process than in Section V-B to obtain a displayable image of distances. Of course, the obtained distance map is rougher than that obtained in Section V-B. But, as we shall see in the Section VI, this method is less computationally expensive, without degradation of the computation of the global distance.

VI. RESULTS

A. Data Base

For our tests, we use a lot of 512×512 pixels size pictures showing a standard scene in computer graphics: the "Cornell_box." In Fig. 9, three of them are shown:

- *Cornell_amb,* computed with a standard ray tracer with an ambient term (computation time: 3 min, 27 s);

- *Cornell_diff,* computed with a ray tracer taking global illumination into account with an evaluation of the diffuse component [48] (computation time: 4 min 9 s);
- *Cornell_mc,* computed with a Monte Carlo method with 256 samples per hemisphere (computation time: 14 h, 5 min, 13 s).

The last picture will be considered as our reference. Thus, we shall compute an image of distances between Cornell_amb and Cornell_mc, and then between Cornell_diff and Cornell_mc.

Whatever the ray tracer method used, we can consider that it is computationally intensive in regards to computations required by image quality metrics. Whatever the image quality metrics used, we can consider that they have more or less the same efficiency in terms of computing time. On the other hand, they have not the same efficiency to quantify such or such visual characteristic. That is the reason why the features used in this study have been selected according to their ability to assess the accuracy of the result of a rendering algorithm in terms of visual aspect.

B. Analysis of Results

We shall give some results on the computation of distance maps (cf. Sections V-B and V-C), of global distances (cf. Section V-D), and of adaptive distance maps (cf. Section V-E).

1) Distance Maps: The computation time to obtain the distance file is about 35 min, on a 200 Mhz MIPS R10 K processor. This computation time is very high because the algorithm is inherently sequential and for a 512×512 pixels size picture, the size of the background mask is around 100×100 pixels near the center of the picture. Consequently, the first distance map (see Section V-B) is quite more computationally intensive than the second one (see Section V-C).

The two distance maps, associated with the two couples of images of Fig. 7, have been first computed from the distance map presented in Section V-B, and next displayed, on Fig. 8, with the following parameters: imperceptibility_threshold = 2.5 and acceptability_threshold = 6.

Let us notice that these thresholds have been chosen at the end of an experimental study conducted in our laboratory by around 20 people [35]. Let us also notice that in order to obtain smoother images

1) A linear interpolation has been applied to F_s between 4.2 and 3.0, for $1 \leq Yb/100 \leq 20$.
2) Instead of assigning to the central pixel the mean of the errors, we took the median value. This has the effect of erasing the residual noise coming from antialiasing.

The four distance maps (cf. Figs. 9 and 10), associated with the two couples of images of Fig. 7, have been first computed from the distance map presented in Section V-C.

For each image of Fig. 7, we have displayed [cf. Fig. 9(c), (e), and (g)], its standard deviation, computed locally. This display uses a black and white scale from the smallest value to the highest one.[2]

For each image of Fig. 7, we have displayed [cf. Fig. 9(d), (f) and (h)], the projection on its principal axis, computed locally.

---

[2]A gamma correction has been applied to local variances values to enhance the most noticeable local contrasts.
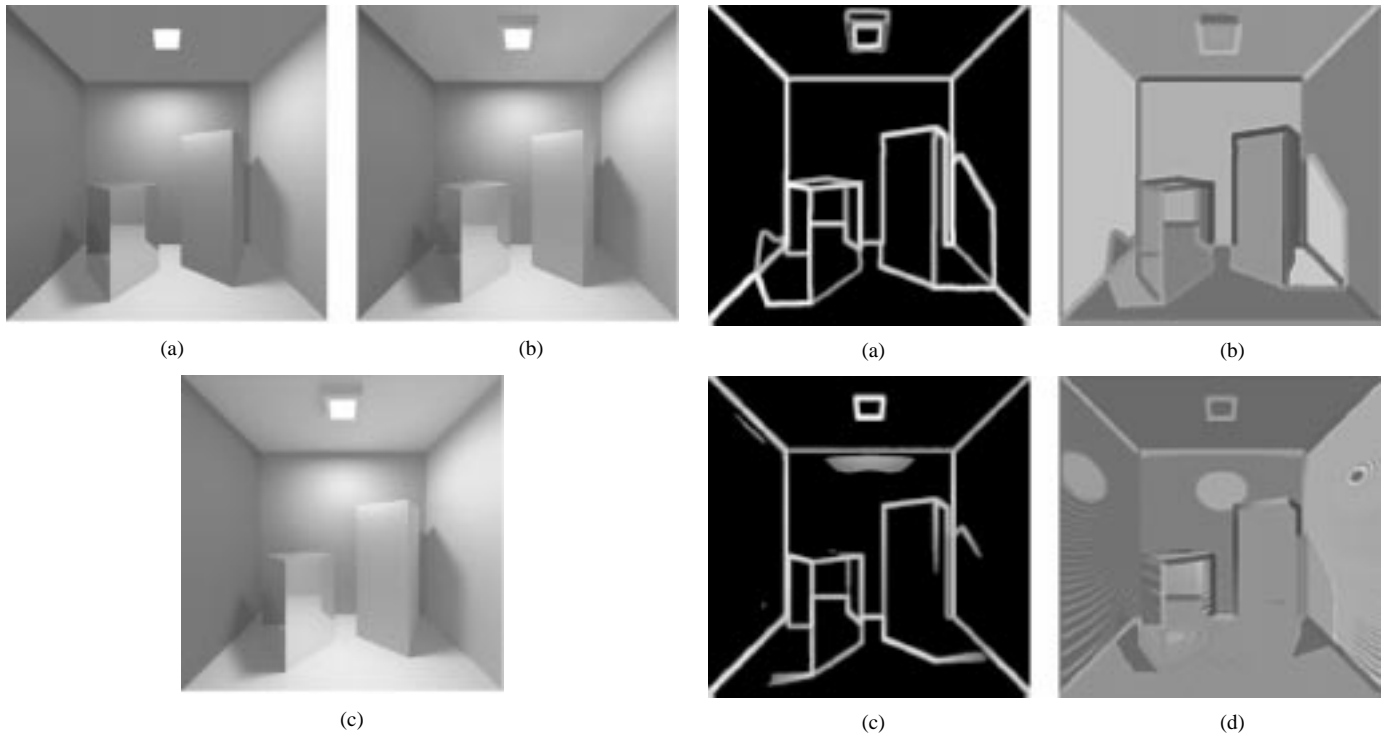
Fig. 7.   Test images. (a) cornell_amb, (b) cornell_diff, and (c) cornell_mc.
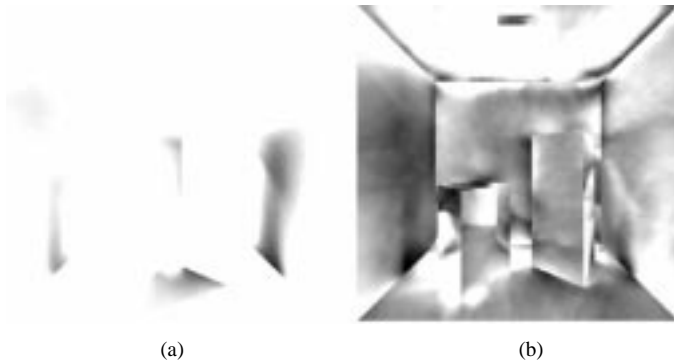

Fig. 8.   Images of distances. (a) Amb/MC and (b) Diff/MC.

This display uses a RGB color scale, such as each pixel is associated with a color which represents the axis direction of the most significant color feature computed in the area centered at this pixel.

In regards to these images, we may notice the following.

- All the images studied (cf. Fig. 7) seem to have, locally, the same color dispersion. Let us notice nevertheless some differences at the top of the walls, due to highlight effects, and some differences at the bottom of the boxes, due to shadows effects.
- All the images studied (cf. Fig. 7) present, in fact, locally, a color rendering slightly different. For example, look at the wall located at the left side of the scene. The change of direction, computed from the principal color feature, underlies a jaggedness effect which differs from one image to the other. This effect is due to a local illumination effect which modifies linearly the lightness of the surface under consideration (eventually its color
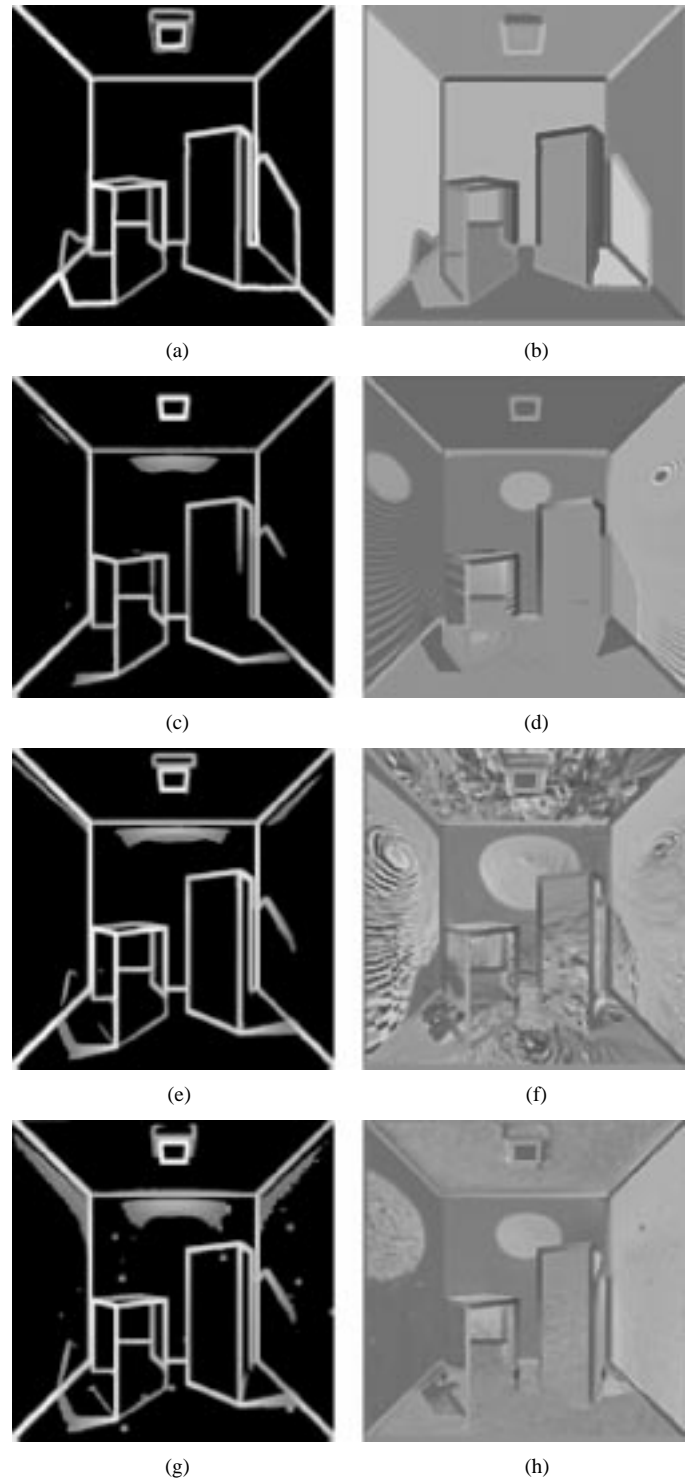

Fig. 9.   For each image, we have at left: its standard deviation computed from local areas of size 10×10 and at right: its projection on its principal axis computed from local areas of size 10×10. (a) Cornell segmented, (b) Cornell segmented, (c) cornell_amb, (d) cornell_amb, (e) cornell_diff, (f) cornell_diff, (g) cornell_mc, and (h) cornell_mc.

also). It is important to notice that it is the discretization of this highlight effect which introduces this jaggedness effect. This effect modifies only the color rendering of the main color attribute of the surface under study; that is the reason why it appears only on this component and not on the set of color components.

Next, for each couple of images of Fig. 7, the locally computed FISHER distance has been displayed [cf. Fig. 10(a) and (c)]. This display uses a black and white scale for which black pixels (value near zero) represent image elements for which the difference is maximal and white pixels (value near 1) represent image elements for which the difference is minimal.

Let us notice that, for most of surface areas, the FISHER distance corresponds in fact to the color distance because these areas are quite homogeneous. The smaller the parameter value ε is, the more the distance takes into account the color dispersion of study areas, to weight the color distance compared to standard deviations computed for each area according to the main color feature. As we may see on Fig. 10, this measure is relevant of perceived differences between images (look after highlight effects on walls and shadow effects linked to boxes); nevertheless, it induces a noisy effect which is difficult to isolate from significant elements.

For each couple of images of Fig. 7, the locally computed NMSE distance has been displayed [cf. Fig. 10(b) and (d)], with the same grayscale than in the previous figures.

Let us notice that the second measure is less sensitive to high color contrasts than those defined by edge elements, and more sensitive to sloped color contrasts than those obtained from jaggedness illumination effects on homogeneous surfaces. As for the FISHER distance, this measure induces a noisy effect which is difficult to isolate from significant elements. Consequently, even though these measures seem to reflect closely perceived differences between images, we are faced to the problem of dissociating elements linked to perceptible differences from noisy elements, in order to compute a global value really significant of noticeable differences. This problem can be overcame by masking the most homogeneous areas before computing images differences. Nevertheless, let us notice that, in our study, the noisy effect which has been noticed in our measures comes, in part, from image *cornell_mc* itself, as we can see in Fig. 9(h).

These latter measures do not require to establish a segmentation between image areas because their field of interest is limited by definition to the internal part of the surfaces of the scene, so theses measures are useless to describe the edges of adjacent surfaces. They are therefore complementary to the color distance previously introduced in Section IV-B for which the field of analysis covers both the internal part of surfaces and the edges of adjacent surfaces. That is the reason why it is sometimes intersecting to prevail this latter measure over the two other ones, even if these former measures are more widely known and used.

2) Global Distance: The distance between pictures Cornell_diff and Cornell_mc computed with our Monte Carlo method is 3.40. The true distance computed with all the pixels is 3.37. The computation time needed by the Monte Carlo method is around 16 s.

3) Adaptive Distance Maps: We may see on Fig. 11 the adaptive distance map associated with two images of Fig. 9, with the same parameters than in Section VI-B-1. In this case, the computation time is around 8 min (to be compared to the 35 min needed to compute a distance map). The global distance computed pixel by pixel is the same (3.37) than the one computed with the help of the distance map.
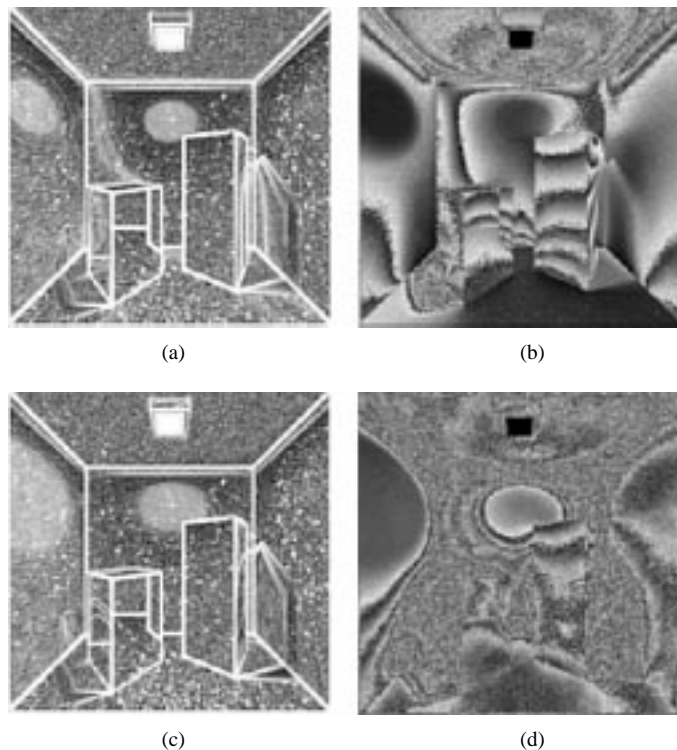

Fig. 10. (a) MC/Amb, (b) MC/Diff, (c) MC/Amb, and (d) MC/Diff. For each of the images, we have at left: the Fischer distance computed locally according to a mask of size 10×10 and to a $\varepsilon = 10 - 4$ and at right: the NMSE measure computed locally according to a mask of size 10×10.

Fig. 11. An image of adaptive distance.

VII. CONCLUSION AND FURTHER DEVELOPMENTS

In this paper, we have proposed three algorithms to compute a perceptual metric between colored images, specific to computer graphics. The tests made show the relevance of this tool. For all the scenes computed with rendering algorithms with increasing quality, the algorithms provide sound results in relation with our expectations.

However, this work is only a first attempt on the way to define a perceptual metric for computer graphics. In particular, nothing proves that the only mean distance computed on the whole set of pixels defines a mathematical distance. A new parameter (or perhaps several), more judicious, could be developed by using spatial and statistical distributions of the distances. By way of validation, we shall attempt to satisfy the five criteria of [18] or at least the three criteria given in [19]. Finally, it could be particularly useful to apply this kind of tool during a rendering

computation, in order to dynamically and efficiently guide the running of the algorithm (cf. [5] and [6]).

For the future, we could set up to a finer segmentation of the study area, by taking shadows and highlights into account, for example.

Some psycho-visual experiments should be driven near both a group of experimental subjects and a set of images, in order to help us to measure the quality of the metric and that of the computed images. In particular, we should like to be able to quantify the five-grade quality scale (excellent, good, fair, poor, bad) and impairment scale (imperceptible, perceptible but not annoying, slightly annoying, annoying, very annoying) quoted in [11] with respect to the above thresholds. The experimental procedure that could be used would be based on the objective picture quality scale (PQS) proposed by Miyahara *et al.* in [49]. It would also use some principles of the experimental procedure we have developed in [40].

REFERENCES

[1] G. Rougeron and B. Péroche, "Color fidelity in computer graphics: A survey," *Comput. Graph. Forum*, vol. 17, pp. 1–13, 1998.

[2] A. J. Willmott and P. S. Heckbert, "An empirical comparison of progressive and wavelet radiosity," in *Proc. 8th Eurographics Workshop on Rendering*, Saint Etienne, France, June 1997, pp. 175–186.

[3] M. Cohen, S. Chen, J. Wallace, and D. Greenberg, "A progressive refinement approach to fast radiosity image generation," *Comput. Graph.*, vol. 22, no. 4, pp. 75–84, 1988.

[4] J.-L. Maillot, L. Carraro, and B. Péroche, "Progressive ray tracing," in *Proc. 3rd Eurographics Workshop on Rendering*, Bristol, U.K., May 1992, pp. 9–19.

[5] S. Gibson and R. J. Hubbold, "Perceptually-driven radiosity," *Comput. Graph. Forum*, vol. 16, no. 2, pp. 129–141, 1997.

[6] K. Myszkowski, "The visible differences predictor: Applications to global illumination problems," in *Proc. 9th Eurographics Workshop on Rendering*, Vienna, Austria, June 1998, pp. 223–236.

[7] A. Trémeau, "Color contrast parameters for analysing image differences," in *Proc., Color Image Science 2000 Conf.*. Derby, U.K.: Univ. Derby, Apr. 2000, pp. 11–23.

[8] E. Favier, P. Colantoni, and A. Tremeau, "Color contrast parameters using edges data and sharpness data," in *Int. Conf. Color in Graphics and Image Processing 2000*, Saint-Etienne, France, Oct. 2000, pp. 231–237.

[9] A. Trémeau and P. Colantoni, "Color contrast parameters using emergence features and edges features," in *Color Image Science: Exploiting Digital Media*, L. W. MacDonald, Ed. New York: Wiley.

[10] F. X. Lukas and Z. L. Budrikis, "Picture quality prediction based on a visual model," *IEEE Trans. Commun.*, vol. COM-30, pp. 1679–1692, July 1982.

[11] S. Comes and B. Macq, "Human visual quality criterion," *Proc. SPIE*, vol. 1360, pp. 2–13, 1990.

[12] A. Trémeau, M. Calonnier, and B. Laget, "Color quantization error in terms of perceived image quality," in *Proc. IEEE Conf. Acoustics, Speech, Signal Processing*, vol. 5-I, 1994, pp. 93–96.

[13] P. Cosman, R. Gray, and R. Olshen, "Evaluating quality of compressed medical images: SNR, subjective rating, and diagnostic accuracy," *Proc. IEEE*, vol. 82, pp. 912–932, June 1994.

[14] C. Jacobs, A. Finkelstein, and D. Salesin, "Fast multiresolution image querying," *Comput. Graph.*, vol. 29, pp. 277–286, 1995.

[15] A. Watson, A. Gale, A. Joshua, and A. Ahumada, "Visual thresholds for wavelet quantization errors," *Proc. SPIE*, vol. 2657, pp. 697–700, 1996.

[16] A. Trémeau, E. Dinet, and E. Favier, "Measurement and display of color image differences based on visual attention," *J. Imag. Sci. Technol.*, vol. 40, no. 6, pp. 522–534, Nov. 1996. IS&T/SID.

[17] G. W. Meyer, H. E. Rushmeier, M. F. Cohen, D. P. Greenberg, and K. E. Torrance, "An experimental evaluation of computer graphics imagery," *ACM Trans. Graph.*, vol. 5, no. 1, pp. 30–50, Jan. 1986.

[18] H. Rushmeier, G. Ward, C. Piatko, P. Sanders, and B. Rust, "Comparing real and synthetic images: Some ideas about metrics," in *Proc. 6th Eurographics Workshop on Rendering*, Dublin, Ireland, June 1995, pp. 213–222.

[19] A. Gaddipatti, R. Machiraju, and R. Yagel, "Steering image generation with wavelet based perceptual metric," *Comput. Graph. Forum*, vol. 16, no. 3, pp. 241–251, 1997.

[20] S. Daly, *Digital Image and Human Vision*. Cambridge, MA: MIT Press, 1993.

[21] M. Ramasubramanian, S. N. Pattanaik, and D. P. Greenberg, "A perceptually based physical error metric for realistic image synthesis," in *SIGGRAPH 1999, Computer Graphics Proceedings*, A. Rockwood, Ed. Reading, MA: Addison-Wesley, 1999, pp. 73–82.

[22] J. Prikryl and W. Purgathofer, "State of the art of perceptually driven radiosity," in *Proc. Eurographics'98*, vol. 17, 1998.

[23] M. R. Bolin and G. W. Meyer, "Visual difference metric for realistic image synthesis," *Proc. SPIE*, vol. 3644, pp. 106–120, 1999.

[24] J. Lubin, "A visual discrimination model for imaging system design and evaluation," in *Vision Models for Target Detection and Recognition*. Singapore: World Scientific, 1995, pp. 245–283.

[25] T. Carney, A. Stanley, and W. Christopher *et al.*, "Development of an image/threshold database for designing and testing human vision models," *Proc. SPIE*, vol. 3644, pp. 542–551, 1999.

[26] S. Winkler, "Vision models and quality metrics for image processing applications," Ph.D. dissertation, Ecole Polytech. Fédérale de Lausanne, France, Dec. 21, 2000.

[27] A. B. Watson, J. A. Solomon, A. J. Ahumada, and A. Gale, "Discrete cosine transform (dct) basis function visibility: Effects of viewing distance and contrast masking," *Proc. SPIE*, vol. 2179, pp. 99–108, 1994.

[28] B. Watson, A. Friedman, and A. McGaffey, "Using naming time to evaluate quality predictors for model simplification," in *ACM CHI 2000 Conf.*, 2000, pp. 113–120.

[29] E. A. Fedorovskaya, H. de Ridder, and F. J. J. Blommaert, "Chroma variations and perceived quality of color images of natural scenes," *Color Res. Applicat.*, vol. 22, no. 2, pp. 96–110, Apr. 1997.

[30] G. Healey, "Segmenting images using normalized color," *IEEE Trans. Syst., Man, Cybern.*, vol. 22, pp. 64–73, Jan./Feb. 1992.

[31] X. Shen and M. Spann, "Segmentation of 2d and 3d images through a hierarchical clustering based on region modeling," in *IEEE Proc. ICIP'97*, vol. 3, 1997, pp. 50–53.

[32] A. Trémeau and P. Colantoni, "Regions adjacency graph applied to color image segmentation," *IEEE Trans. Image Processing*, vol. 9, pp. 735–744, Apr. 2000.

[33] M. D. Levine, *Vision in Man and Machine*. New York: McGraw-Hill, 1985.

[34] M. H. Pirenne and R. Crouzy, *L'Oeil et la Vision*. Paris, France: Gauthier-Villars Editeur, 1972.

[35] G. Rougeron, "Problèmes liés à la couleur en synthèse d'images," Ph.D. dissertation, Ecole des mines de Saint-Etienne, Univ. Jean Monnet, France, Jan. 27, 1998.

[36] P. J. Burt, "Attention mechanism for vision in a dynamic world," in *Proc. ICPR*, 1988, pp. 977–987.

[37] M. R. Luo, M. C. Lo, and W. G. Kuo, "The LLAB(1:C) color model," *Color Res. Applicat.*, vol. 21, pp. 412–429, 1996.

[38] M. R. Luo, "The LLAB model for color appearance and color difference evaluation," *Proc. SPIE*, vol. 2658, pp. 261–269, 1996.

[39] M. D. Fairchild, *Color Appearance Models*, 1998.

[40] A. Trémeau and C. Charrier, "Influence of chromatic changes on perception of color image quality," *Color Res. Applicat.*, vol. 25, no. 3, pp. 200–213, June 2000.

[41] K. M. Lam, "Metamerism and color constancy," Ph.D. dissertation, Univ. Bradford, U.K., 1985.

[42] G. Wyszecki and W. S. Stiles, *Color Science: Concepts & Methods, Quantitative Data and Formulae*, 2nd ed. New York: Wiley, 1982.

[43] A. Trémeau, P. Colantoni, and B. Laget, "On color segmentation guided by the cooccurrence matrix," in *Proc. OSA Annu. Conf. Optics Imaging Information Age*, Rochester, NY, Oct. 1996, pp. 30–38.

[44] P. Colantoni, A. Trémeau, and B. Laget, "Comparison of color median filters," in *Int. Symp. Electronic Image Capture Publishing*, Zürich, Switzerland, May 1998, pp. 186–195.

[45] Y. I. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," *Comput. Graph. Image Process.*, vol. 13, pp. 222–241, 1980.

[46] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," Tech. Rep., 1997.

[47] A. J. Bardos and S. J. Sangwine, "Selective vector median filtering of color images," in *Proc. IPA'97*, vol. 443, July 1997, pp. 708–711.

[48] J. Zaninetti and B. Péroche, "A vector approach for global illumination in ray tracing," in *Proc. Eurographics'98*, vol. 17, 1998.

[49] M. Miyahara, K. Kotani, and V. R. Algazi, "Objective picture quality scale (PQS) for image coding," *IEEE Trans. Commun.*, vol. 46, pp. 1215–1226, Sept. 1998.

Bernard Péroche is a Professor of computer science with the Université Claude Bernard Lyon 1, France. He is the Head of LIGIM, a research laboratory working on computer graphics, vision and digital processing. His research interests include modeling and rendering in computer graphics. He is currently mainly focused on rendering, with a particular emphasis on global illumination with ray tracing and on perceptually-based rendering algorithms.

Stephane Albin received a degree in computer graphics in 1998 from the Ecole Nationale Superieure des Mines de Saint-Etienne, France, where he is pursuing currently the Ph.D. degree.

His research interests include color, perception and psycho-visual problems, light sources, directionally dependent phenomena, and all rendering techniques.

Gilles Rougeron graduated from the Ecole Nationale Superieure de l'Electronique et de ses Applications, France, in 1993. He received the Ph.D. degree in computer graphics from Saint Etienne University in 1998. His thesis, "Problems related to color in image synthesis," was concerned with spectral rendering, color fidelity, and color image distances topics.

In 1999, he had a one year postdoctoral position with the Centre Scientifique et Technique du Bâtiment, Grenoble, France, where he worked on the development of a radio wave propagation simulation software. Since then, he has been a Research and Development Software Engineer with Abvent, a French firm specialized in computer graphics.

Alain Trémeau is a Professor of color imaging with the Université Jean Monnet Saint-Etienne. He is the head of LIGIV (http://www.ligiv.org), a research laboratory working on computer graphics, vision engineering and color imaging science. He is currently mainly focused on mathematical imaging and color science with reference to human vision and perception. His research covers quantization, dithering, image quality, segmentation, and texture analysis. He also works in color metric with regard to color appearance and rendering measurements. He has written numerous papers on computational color imaging and processing. He is at the Head of the French Color Imaging Group which co-organized different international conferences on color imaging, such as CGIV'2000 and CGIP'2002.