Eurographics Symposium on Rendering 2003
Per Christensen and Daniel Cohen-Or (Editors)

Refinement Criteria Based on *f*-Divergences

Jaume Rigau[†], Miquel Feixas[‡], and Mateu Sbert[§]

Institut d'Informàtica i Aplicacions, Universitat de Girona, Catalonia, Spain

Abstract

In several domains a refinement criterion is often needed to decide whether to go on or to stop sampling a signal. When the sampled values are homogeneous enough, we assume that they represent the signal fairly well and we do not need further refinement, otherwise more samples are required, possibly with adaptive subdivision of the domain. For this purpose, a criterion which is very sensitive to variability is necessary. In this paper we present a family of discrimination measures, the f*-divergences, meeting this requirement. These functions have been well studied and successfully applied to image processing and several areas of engineering. Two applications to global illumination are shown: oracles for hierarchical radiosity and criteria for adaptive refinement in ray-tracing. We obtain significantly better results than with classic criteria, showing that* f*-divergences are worth further investigation in computer graphics.*

Categories and Subject Descriptors (according to ACM CCS): I.3.7 [Three-Dimensional Graphics and Realism]: Color, shading, shadowing, and texture

1. Introduction

When sampling a signal we need a criterion to decide whether to take additional samples, albeit within the original domain or within a hierarchical subdivision. The refinement criteria are mainly based on the encountered homogeneity of the samples. Inhomogeneity should lead to further sampling, possibly with an adaptive subdivision of the domain. Oracles are then built based on these criteria. Examples in computer graphics of this refinement process are hierarchical radiosity[2, 18] and adaptive supersampling in ray-tracing[26, 31].

In this paper, we introduce new refinement criteria based on *f*-divergences. These are a family of convex functions that fulfill very remarkable properties. They were introduced by Csiszár[10] and Ali and Silvey[1] as measures of discrimination or distance between probability distributions. As such, they are perfectly fitted as homogeneity measures, when we consider how distant the distribution of the samples is with respect to its average. They have been successfully used in image processing and several engineering areas [21, 27, 30].

The purpose of this paper is to demonstrate the usefulness of *f*-divergences in computer graphics by applying them in defining new refinement criteria for hierarchical radiosity and adaptive supersampling of a pixel in ray-tracing. We will see how, compared with classic refinement criteria, the *f*-divergences-based ones give significant better results.

This paper is organised as follows. In Section 2, criteria for refinement in hierarchical radiosity and adaptive ray-tracing, and the concept of *f*-divergence are presented. Section 3 describes the application of the refinement criteria based on *f*-divergences to hierarchical radiosity and, in Section 4, to adaptive ray-tracing. Finally, in Section 5 we present our conclusions and future work.

2. Previous Work

In this section, refinement criteria used in hierarchical radiosity and adaptive ray-tracing are reviewed. Also, Jensen's inequality, needed to establish our theoretical framework, and *f*-divergences are shortly introduced.

2.1. Refinement Criteria for Hierarchical Radiosity

The radiosity method uses a finite element approach, discretising the diffuse environment into N_p patches and taking

† jaume.rigau@udg.es
‡ miquel.feixas@udg.es
§ mateu.sbert@udg.es

into account that the radiosities, emissivities and reflectances are constant over the patches. Under these assumptions, the discrete radiosity equation[16] is given by

$$B_i = E_i + \rho_i \sum_{j=1}^{N_p} F_{ij} B_j \quad , \tag{1}$$

where B_i, E_i, and ρ_i, are respectively the radiosity, emissivity, and reflectance of patch i, B_j is the radiosity of patch j, and F_{ij} is the *patch-to-patch form factor*, only dependent on the geometry of the scene. Form factor F_{ij} is defined by

$$F_{ij} = \frac{1}{A_i} \int_{S_i} \int_{S_j} F(x,y) \mathrm{d}A_x \mathrm{d}A_y \quad , \tag{2}$$

where A_i is the area of patch i, S_i and S_j are, respectively, the surfaces of patches i and j, $F(x,y)$ is the *point-to-point form factor*[34] between points $x \in S_i$ and $y \in S_j$, and $\mathrm{d}A_x$ and $\mathrm{d}A_y$ are, respectively, the differential areas at points x and y.

A hierarchical refinement algorithm[18] is used to solve the equation system (1). Since the application of a good refinement criterion is fundamental for its efficiency, many oracles have been proposed in the literature (consult[2, 7, 15]). For the purposes of this paper, two of them, based respectively on kernel smoothness and mutual information, are reviewed.

In Gortler et al.[17], the *variability* of the radiosity kernel, i.e., the point-to-point form factor $F(x,y)$, is taken into account. The refinement criterion based on *kernel smoothness*, when applied to constant approximations, is given by

$$\rho_i \max(F_{ij}^{\mathrm{avg}} - F_{ij}^{\mathrm{min}}, F_{ij}^{\mathrm{max}} - F_{ij}^{\mathrm{avg}}) A_j B_j < \varepsilon \quad , \tag{3}$$

where A_j and B_j are respectively the source element area and the source element radiosity, $F_{ij}^{\mathrm{avg}} = F_{ij}/A_j$ is the average radiosity kernel value, $F_{ij}^{\mathrm{min}} = \min_{x \in S_i, y \in S_j} F(x,y)$ and $F_{ij}^{\mathrm{max}} = \max_{x \in S_i, y \in S_j} F(x,y)$ are the minimum and maximum point-to-point form factors computed with pairs of random points on both elements i and j, and ε is a given threshold.

In Feixas et al.[13, 14], an oracle based on the visibility *discretisation error* between two elements was introduced. This discretisation error is obtained from the difference between continuous and discrete *mutual information* and it can be interpreted as the *loss* of information transfer due to discretisation or as the *maximum potential gain* of information transfer between two elements. Hence, this difference can be considered as the *benefit* to be gained by refining, and consequently is used as a decision criterion. It also represents the *variability* of the radiosity kernel. The oracle based on *mutual information* is given by

$$\rho_i \delta_{ij} B_j < \varepsilon \quad , \tag{4}$$

where

$$\delta_{ij} \approx \frac{A_i A_j}{A_{\mathrm{T}}} \Big(\mathrm{avg}_{1 \le k \le N_s} (F(x_k,y_k) \log F(x_k,y_k)) - \mathrm{avg}_{1 \le k \le N_s} (F(x_k,y_k)) \log(\mathrm{avg}_{1 \le k \le N_s} (F(x_k,y_k))) \Big) \tag{5}$$

is the discretisation error between elements i and j, A_{T} is the total area of the scene, ε is a predefined threshold, and $\mathrm{avg}_{1 \le i \le n}(x_i) = \frac{1}{n} \sum_{i=1}^{n} x_i$. The computation of the point-to-point form factors $F(x_k,y_k)$ is done with N_s random lines (x_k,y_k) joining both elements i and j[13].

2.2. Refinement Criteria for Adaptive Ray-Tracing

Ray-tracing[38] is a point-sampling-based technique for image synthesis. Rays are traced from the eye through a pixel to sample the radiance at the hitpoint in the scene, where radiance is usually computed by a random walk method[35]. Since a finite set of samples is used, some of the information in the scene is lost. Thus, aliasing errors are unavoidable[11].

These errors can be reduced using extra sampling in regions where the sample values vary most. In order to obtain reliable data, the edge of an object, the contour of a shadow, or a high illumination gradient area, need a more intensive treatment than a region with almost uniform illumination. This method of sampling is called *adaptive sampling*[11, 28]: a pixel is first sampled at a relatively low density and, from the initial sample values, a refinement criterion is used to decide whether more sampling is required or not. Finally, all the samples are used to obtain the final pixel colour values[26].

Diverse refinement criteria for adaptive sampling, based on colour intensities and/or scene geometry, can be found to control the sampling rate: Dippé and Wold[11] present an error estimator based on the RMS signal to noise ratio and also consider its variance as a function of the number of samples; Mitchell[26] proposes a contrast measure[6] based on the characteristics of the human eye; Lee et al.[25], Purgathofer[31], and Tamstorf and Jensen[36] develop different methods based on the variance of the samples with their respective confidence intervals. Bolin and Meyer[5] have developed a perceptually-based approach using statistical and vision models.

For the purposes of this paper, we review two commonly used refinement criteria based on the contrast and the variance of the samples. Mitchell[26] uses a contrast measure[6] for each RGB channel defined by

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}} \quad , \tag{6}$$

where $I_{\min}$ and $I_{\max}$ are, respectively, the minimum and maximum light intensities of the channel. Supersampling is done if any contrast is higher than a given threshold. Mitchell proposes RGB threshold values (0.4, 0.3 and 0.6, respectively) based on the relative sensitivity of the visual system. In Glassner[15, pp. 476], this criterion appears weighted by the average colour of the pixel.

The basic idea of variance-based methods[25, 31, 36] is to continue sampling until the confidence level or probability that the true value of luminance L is within a given tolerance d of the estimated value $\widehat{L}$ is $1-\alpha$:

$$\Pr[L \in (\widehat{L}-d, \widehat{L}+d)] = 1-\alpha \ , \tag{7}$$

and this will happen[31] when

$$t_{1-\alpha,n-1}\frac{s}{\sqrt{n}} \leq d \ , \tag{8}$$

where t is the Student distribution and s is the standard deviation of the n samples.

2.3. Jensen's Inequality

A function $f(x)$ is *convex* over an interval $[a,b]$ (the graph of the function lies below any chord) if for every $x_1, x_2 \in [a,b]$ and $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \ . \tag{9}$$

A function is strictly convex if equality holds only if $\lambda = 0$ or $\lambda = 1$. A function $f(x)$ is *concave* (the graph of the function lies above any chord) if $-f(x)$ is convex. For instance, x^2 and $x\log x$ (for $x \geq 0$) are strictly convex functions, and $\log x$ (for $x \geq 0$) is a strictly concave function[9].

A generalization of the above convexity property, called Jensen's inequality, is widely used in mathematics, information theory, and different engineering areas as a *divergence measure*. For example, it has been successfully applied to image registration[19] and DNA segmentation[4].

Jensen's inequality[22]: If f is a convex function on the interval $[a,b]$, then

$$\sum_{i=1}^{n} \lambda_i f(x_i) - f\left(\sum_{i=1}^{n} \lambda_i x_i\right) \geq 0 \ , \tag{10}$$

where $0 \leq \lambda \leq 1$, $\sum_{i=1}^{n} \lambda_i = 1$, and $x_i \in [a,b]$. If f is a concave function, the inequality is reversed.

A very special case of this inequality is when $\lambda_i = \frac{1}{n}$ because then

$$\frac{1}{n}\sum_{i=1}^{n} f(x_i) - f\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \geq 0 \ , \tag{11}$$

i.e., the value of the function at the mean of the x_i is less or equal than the mean of the values of the function at each x_i.

In particular, if f is convex on the range of a random variable X, then

$$E[f(X)] - f(E[X]) \geq 0 \ , \tag{12}$$

where E denotes expectation. Observe that if $f(x) = x^2$, then we obtain the variance: $E(X^2) - (E[X])^2$.

In the Rao's axiomatization of *diversity* measures[32], the concavity condition (the reverse of expression (10)) meets the intuitive requirement that diversity is possibly increased by mixing, i.e., the average diversity between any p, q probability distributions is not greater than that between their average.

Another important inequality can be obtained from Jensen's inequality.

Log-sum inequality: For non-negative numbers, $a_1, a_2, \ldots, a_n$ and $b_1, b_2, \ldots, b_n$,

$$\sum_{i=1}^{n} a_i \log\frac{a_i}{b_i} - \left(\sum_{i=1}^{n} a_i\right)\log\frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \geq 0 \tag{13}$$

with equality if and only if $\frac{a_i}{b_i}$ is constant. We use the convention that $0\log 0 = 0$, $a\log\frac{a}{0} = \infty$ if $a > 0$ and $0\log\frac{0}{0} = 0$. These follow easily from continuity. Note that the conditions in this inequality are much weaker than for Jensen's inequality.

2.4. *f*-divergences

Many different measures quantifying the degree of discrimination between two probability distributions have been studied in the past. They are frequently called *distance* measures, although some of them are not strictly metrics. Let us remember that a metric on a set X is an assignment of a distance $d : X \times X \to \mathbb{R}$ satisfying the following properties[23]:

- *Positivity*: $\forall x, y \in X$, $d(x,y) \geq 0$ and $d(x,y) = 0$ if and only if $x = y$.
- *Symmetry*: $\forall x, y \in X$, $d(x,y) = d(y,x)$.
- *Triangle inequality*: $\forall x, y, z \in X$, $d(x,z) \leq d(x,y) + d(y,z)$.

Next, we review a measure of discrimination between two probability distributions called *f-divergence*. This measure was independently introduced by Csiszár[10] and Ali and Silvey[1]. It has been applied to different areas, such as medical image registration[30] and classification and retrieval[21], among others.

Let $\Omega = \{x_1, x_2, \ldots, x_n\}$ be a set with at least two elements and $\mathcal{P}$ the set of all probability distributions $p = \{p_i | p_i = \Pr(x_i), x_i \in \Omega\}$. Given a convex function $f : [0,\infty) \to \mathbb{R}$ continuous at 0 (i.e. $f(0) = \lim_{x\to 0} f(x)$) and a pair $(p,q) \in \mathcal{P}^2$, then

$$I_f(p,q) = \sum_{i=1}^{n} q_i f\left(\frac{p_i}{q_i}\right) \tag{14}$$

is called the *f-divergence* of the probability distributions p and q.

The following are important properties of the *f*-divergences:

- $I_f(p,q)$ is convex on (p,q), i.e., if (p_1,q_1) and (p_2,q_2) are two pairs of probability density functions, then

$$I_f(\lambda p_1 + (1-\lambda)p_2, \lambda q_1 + (1-\lambda)q_2) \leq \lambda I_f(p_1,q_1) + (1-\lambda) I_f(p_2,q_2) \ . \tag{15}$$

- $I_f(p,q) \geq f(1)$, where the equality holds if $p = q$. If f is strictly convex, the equality holds if and only if $p = q$.
- If $f(1) = 0$ then $I_f(p,q) \geq 0$. In this case, $I_f(p,q)$ fulfills the positivity property of a metric.

Next, we present some of the most important f-divergences[12], called *distances* in the literature. These can be obtained from different convex functions f. Observe that, for all of them, $f(1) = 0$, and thus they fulfill the positivity property. In the following, we take $x > 0$.

- *Kullback-Leibler distance*[24]
  If $f(x) = x \log x$, the Kullback-Leibler distance is given by

$$D(p,q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \quad . \tag{16}$$

- *Chi-square distance*[29]
  If $f(x) = (x-1)^2$, the Chi-square distance is given by

$$\chi^2(p,q) = \sum_{i=1}^{n} \frac{(p_i - q_i)^2}{q_i} \quad . \tag{17}$$

- *Hellinger distance*[20]
  If $f(x) = \frac{1}{2}(1-\sqrt{x})^2$, the Hellinger distance is given by

$$h^2(p,q) = \frac{1}{2} \sum_{i=1}^{n} (\sqrt{p_i} - \sqrt{q_i})^2 \quad . \tag{18}$$

Note that none of the above distance fulfills all the properties of a metric. However, $h(p,q)$, the square root of the Hellinger distance, is a true metric.

3. Application of f-divergences to Radiosity

In this section new oracles based on f-divergences for hierarchical radiosity refinement are introduced.

3.1. f-divergences for Hierarchical Radiosity

The discretisation error (5), seen in Section 2.1, can be written in the following way:

$$\begin{aligned} \delta_{ij} &\approx \frac{A_i A_j}{A_T} \widehat{F} \Big[\mathrm{avg}_{1 \leq k \leq N_s}(p_k \log p_k) \\ &\quad - \mathrm{avg}_{1 \leq k \leq N_s}(p_k) \log(\mathrm{avg}_{1 \leq k \leq N_s}(p_k)) \Big] \\ &= \frac{A_i A_j}{A_T} \widehat{F} \Big[\mathrm{avg}_{1 \leq k \leq N_s}(p_k \log p_k) - \frac{1}{N_s} \log \frac{1}{N_s} \Big] \quad , \end{aligned} \tag{19}$$

where $\widehat{F} = \sum_{k=1}^{N_s} F(x_k, y_k)$, $p_k = \frac{F(x_k,y_k)}{\widehat{F}}$ for all $1 \leq k \leq N_s$, and $\mathrm{avg}_{1 \leq k \leq N_s}(p_k) = \frac{1}{N_s}$.

It is easy to see that the expression between brackets in (19), except for a constant factor $\frac{1}{N_s}$, is the Kullback-Leibler distance between the distributions $p_k = \frac{F(x_k,y_k)}{\widehat{F}}$ and $q_k = \frac{1}{N_s}$. Thus,

$$\delta_{ij} \approx \frac{A_i A_j}{A_T} \frac{1}{N_s} \widehat{F} D(p,q) \quad . \tag{20}$$

This fact suggests that we try other f-divergences in the kernel of the refinement oracle (4). These measures will give us the variability of the distribution $\{\frac{F(x_1,y_1)}{\widehat{F}}, \ldots, \frac{F(x_{N_s},y_{N_s})}{\widehat{F}}\}$ with respect to the uniform distribution $\{\frac{1}{N_s}, \ldots, \frac{1}{N_s}\}$.

Thus, the Kullback-Leibler (16), Chi-square (17), and Hellinger (18) distances have been tested. The Kullback-Leibler-based oracle was already studied in[13, 14] from an information-theoretic perspective.

The oracles used in the test are the following:

- Kullback-Leibler (KL)

$$\rho_i A_i A_j \widehat{F} D(p,q) B_j < \varepsilon \tag{21}$$

- Chi-square (CS)

$$\rho_i A_i A_j \widehat{F} \chi^2(p,q) B_j < \varepsilon \tag{22}$$

- Hellinger (HE)

$$\rho_i A_i A_j \widehat{F} h^2(p,q) B_j < \varepsilon \quad , \tag{23}$$

based all on their respective distances. Observe that the constants $\frac{1}{A_T}$ and $\frac{1}{N_s}$ have been removed.

It is important to note that the expression between brackets in (19) is equal to the first term of Jensen's inequality (11) with $f(x) = x \log x$ and $x = \frac{F(x,y)}{\widehat{F}}$. Moreover, we can also see that this expression is equal to the first term of the log-sum inequality (13), taking $b_i = 1$ and $a_i = \frac{F(x_i,y_i)}{\widehat{F}}$.

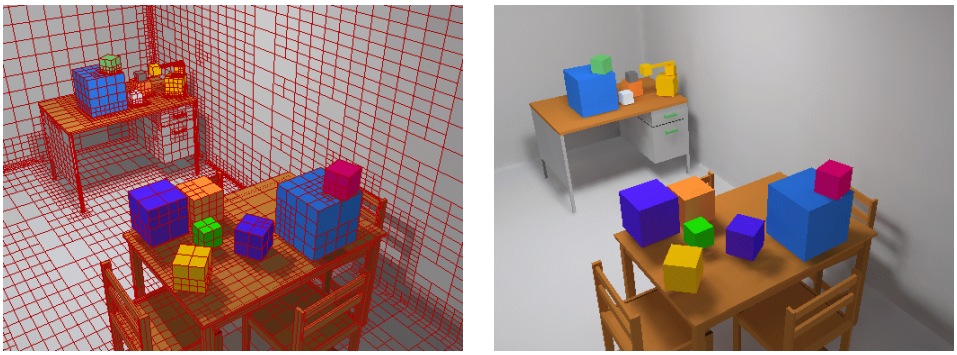
3.2. Empirical Results and Discussion

The kernel-smoothness-based (KS) and f-divergence-based oracles have been implemented on top of the hierarchical Monte Carlo radiosity[3] method of RenderPark[8] software (`www.renderpark.be`). It should be noted that our oracles can be used with any hierarchical radiosity method.

In Fig. 1 we show a general view of the test scene obtained with the KL oracle (21). The left column (i) shows the subdivision obtained, while the right one (ii) corresponds to the Gouraud shaded solution. Each oracle has been evaluated with 10 random lines between the corresponding pair of elements and a total of 2685000 rays have been cast for the radiosity computation. The ε parameter has been tuned so that the grids obtained have approximately 19000 patches in all the methods.

In Fig. 2 we present the results of comparing the KS oracle (3) of Section 2.1 (Fig. 2.(a)) with the f-divergence-based ones (21,22, 23) defined in Section 3.1 (Fig. 2.(b,c,d)) for a closer view of the test scene.

In Fig. 2.(b,c,d) we can see how the f-divergence-based oracles outperform the KS one (Fig. 2.(a)), especially in the much more-defined shadow of the chair and the cubes on the right wall. Observe also the superior quality of the grid created on top of the table, and in the corner between the walls.



(*i*) (*ii*)

Figure 1: *General view of the test scene obtained with the KL-based oracle (21). (i) shows the grid obtained in the refinement process and (ii) shows the Gouraud shaded solution. The oracle has been evaluated with 10 random lines between two elements. A total of 2685000 rays are cast for the radiosity computation, obtaining approximately 19000 patches.*

On the other hand, comparing our three *f*-divergence oracles we conclude that, although they exhibit a similar quality, the KL one is slightly better. For instance, observe that the shadows on the table are more defined. A possible explanation for this better behaviour could be that the KL oracle, unlike the other ones, meets Jensen's inequality (11). This confers a distinctive theoretical advantage on the Kullback-Leibler oracle.

From the above, one could be tempted to use Jensen's inequality alone as a kernel for a refinement oracle. We have experimented with the function $f(x) = x^2$ which when substituted in Jensen's inequality corresponds to the variance. Thus, substituting $F(x_k, y_k) \log F(x_k, y_k)$ by $F(x_k, y_k)^2$ in equation (5), the variance-based oracle is given by

$$\rho_i A_i A_j \widehat{F}^2 V(p,q) B_j < \varepsilon \ , \tag{24}$$

where $V(p,q) = \text{avg}_{1 \le k \le N_s}(p_k^2) - (\frac{1}{N_s})^2$. The results obtained are presented in Fig. 3, showing the inadequacy of this function and, incidentally, of this approach.

4. Application to Ray-Tracing

In this section new refinement criteria based on *f*-divergences for adaptive supersampling in ray-tracing are obtained.

4.1. *f*-divergences for Adaptive Ray-Tracing

The *f*-divergences defined in Section 2.4 will be used to evaluate the inhomogeneity of a set of samples in a region.

The scheme used is the following:

1. A first batch of N_s rays is cast through a pixel and the corresponding luminances $L_{i \in \{1,\ldots,N_s\}}$ are obtained.
2. The *f*-divergences $I_f(p,q)$ are taken between the normalised distribution of the obtained luminances,

$$p_i = \frac{L_i}{\sum_{i=1}^{N_s} L_i} \ , \tag{25}$$

and the uniform distribution $q_i = \frac{1}{N_s}$.

3. The refinement criterion, given by

$$\frac{1}{N_s} \overline{L} I_f(p,q) < \varepsilon \ , \tag{26}$$

is evaluated, where I_f represents the Kullback-Leibler (KL), Chi-square (CS), or Hellinger (HE) distances, $\overline{L}$ is the average luminance

$$\overline{L} = \frac{1}{N_s} \sum_{i=1}^{N_s} L_i \ , \tag{27}$$

and ε is a predefined threshold for the refinement test.

4. Successive batches of N_s rays are cast until the result of the test is positive.

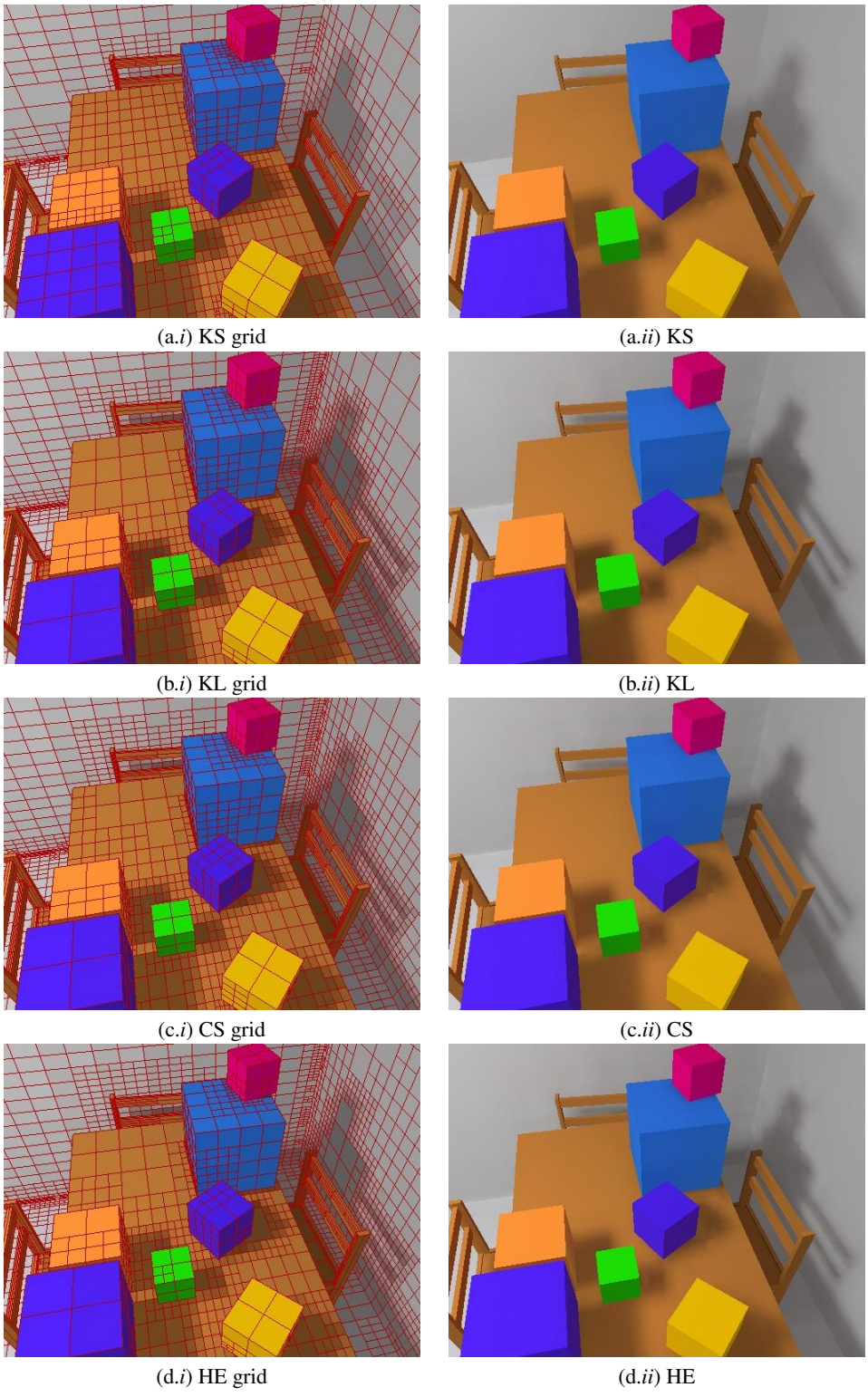
Note that to assign an importance to the distance value $I_f(p,q)$ in (26) we weight it by the average luminance (27), as in Glassner's version of classic contrast[15]. Division by the number of samples N_s in (26) ensures that the refinement process stops.

The new criteria give good visual results, but the error obtained in our tests (see Table 1), although better than in the classic contrast, is higher than with the variance criterion (8). Our next logical step was to try the square root of Hellinger divergence[37], as it is a true metric. The results obtained were very encouraging. By analogy, we then extended the experimentation to the square root of the other divergences. This is not new. For instance, the square root of Kullback-Leibler distance has been used by Yang and Barron[39]. The results also improved the previous ones and were also better than in the variance case.

Thus, the criteria finally used were:

- Square root of Kullback-Leibler distance (SRKL)

$$\frac{1}{N_s} \overline{L} \sqrt{D(p,q)} < \varepsilon \tag{28}$$



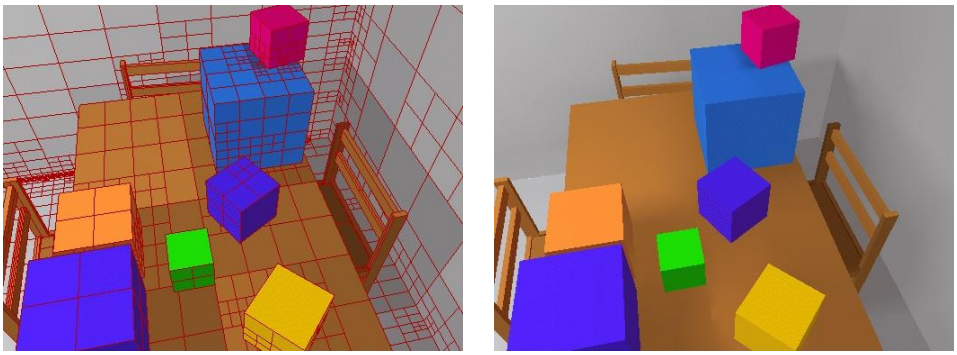
(a.*i*) KS grid (a.*ii*) KS

(b.*i*) KL grid (b.*ii*) KL

(c.*i*) CS grid (c.*ii*) CS

(d.*i*) HE grid (d.*ii*) HE

Figure 2: *A closer view from another camera of test scene for comparison of (a) kernel-smoothness-based (KS) vs.* f*-divergence-based oracles: (b) Kullback-Leibler (KL), (c) Chi-square (CS), and (d) Hellinger (HE). Column (i) shows the grid obtained in the refinement process and column (ii) shows the Gouraud shaded solution. In all the methods, the oracles have been evaluated with 10 random lines between two elements. In each case, a total of 2685000 rays are cast for the radiosity computation, obtaining approximately 19000 patches.*



(*i*) (*ii*)

Figure 3: *A closer view from another camera of test scene using the variance-based oracle (24). (i) shows the grid obtained in the refinement process and (ii) shows the Gouraud shaded solution. The oracle has been evaluated with 10 random lines between two elements. A total of 2685000 rays are cast for the radiosity computation, obtaining approximately 19000 patches. Compare with the results in Fig. 2, obtained with f-divergences.*

- Square root of Chi-square distance (SRCS)

$$\frac{1}{N_s}\overline{L}\sqrt{\chi^2(p,q)} < \varepsilon \tag{29}$$

- Square root of Hellinger distance (SRHE)

$$\frac{1}{N_s}\overline{L}\sqrt{h^2(p,q)} < \varepsilon \ . \tag{30}$$

4.2. Empirical Results and Discussion

In Figures 5 and 6 we present comparative results with different techniques for the test scene of Fig. 4. The following methods are compared:

- *CC*: Classic contrast (6) of the luminance weighted with the respective importance $\overline{L}$.
- *VAR*: Variance (8).
- *SRKL*: Square root of Kullback-Leibler distance (28).
- *SRCS*: Square root of Chi-square distance (29).
- *SRHE*: Square root of Hellinger distance (30).

In all the methods, 8 initial rays are cast in a stratified way (2×4 strata) at each pixel to compute the contrast measures for the refinement decision, and 8 additional rays are successively added until the condition of the criterion is met. In the variance method, we have used $\alpha = 0.1$ and $d = 0.025$. All the images have been obtained with the RenderPark[8]. An implementation of classic path-tracing with next event estimator was used to compute all images. The parameters were tuned so that all four test images were obtained with a similar average number of rays per pixel (60) and a similar computational cost. A constant box filter was used in the reconstruction phase for all the methods.

The resulting images are shown in column (*i*) of Fig. 5 and Fig. 6, with the sampling density maps in column (*ii*) (warm colours correspond to higher sampling rates and cold colours to lower ones). The overall aspect of the images shows that our supersampling scheme performs the best. Observe, for instance, the reduced noise in the shadows cast by the objects. Observe also the detail of the shadow of the sphere reflected on the pyramid.

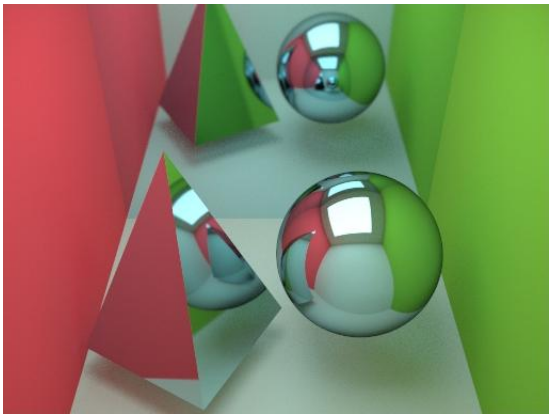
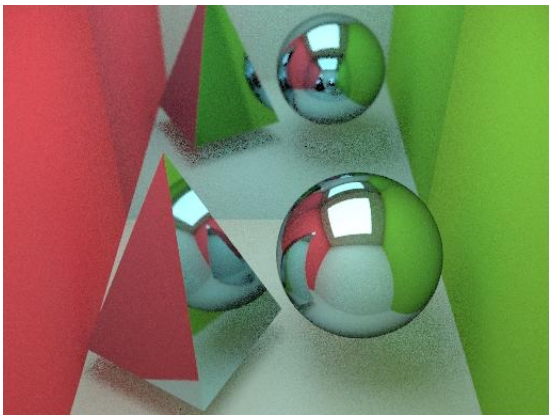


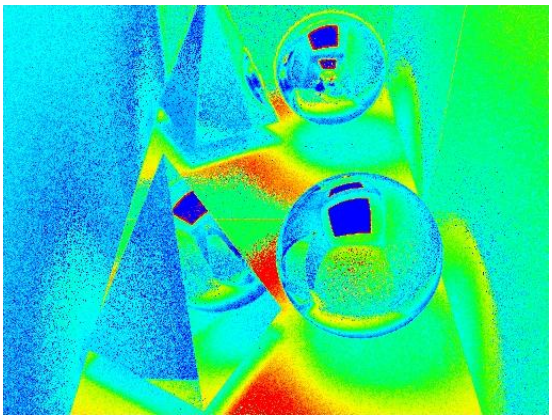
Figure 4: *Reference image for the ray-tracing comparison in Fig. 5 and Fig. 6, obtained with 1000 rays per pixel.*

Comparison of the sampling density maps in Fig. 5.(*ii*) and Fig. 6.(*ii*) shows a better discrimination of complex regions of the scene in the three divergence cases against the classic contrast and variance cases. This explains the better results obtained by our approach. On the other hand, the variance-based approach (Fig. 5.(b)) also performs better than the classic contrast-based method (Fig. 5.(a)). Its sampling map also explains why it performs better. However, it is unable to render the reflected shadows under the mirrored pyramid and sphere with precision.

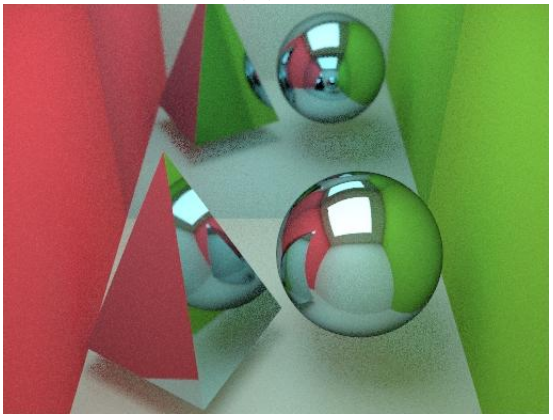
In Table 1, we show the root mean square error (RMSE) of the images obtained with classic (Fig. 5.(*i*)), *f*-divergence,



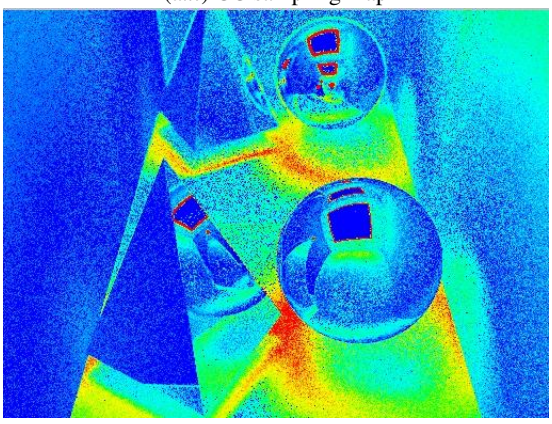
(a.*i*) CC



(a.*ii*) CC sampling map



(b.*i*) VAR



(b.*ii*) VAR sampling map

Figure 5: *Images obtained with an adaptive sampling scheme based on (a) classic contrast (CC) and (b) variance-based (VAR) methods. Column (i) shows the resulting images and (ii) the sampling density map. The average number of rays per pixel is 60 in all the methods, with a similar computation cost. Compare with the images in Fig. 6.*

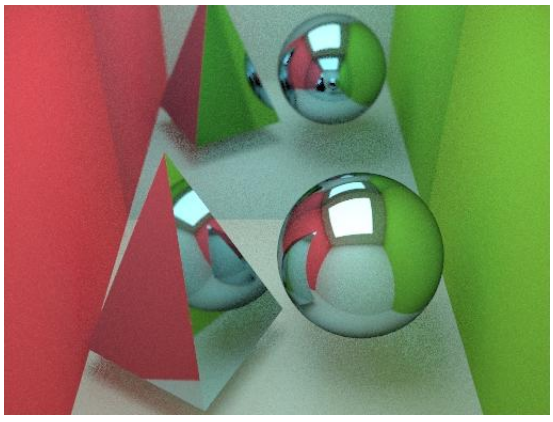
and square root of *f*-divergence (Fig. 6.(*i*)) methods respective to the reference image in Fig. 4. Visual comparison is in concordance with numerical error. The divergence-based criteria used in our experiments (SRKL, SRCS, and SRHE) outperform both classic contrast and variance ones. Finally, the better behaviour of the SRHE criterion could be explained by the fact that it is a true distance.

5. Conclusions and Future Work

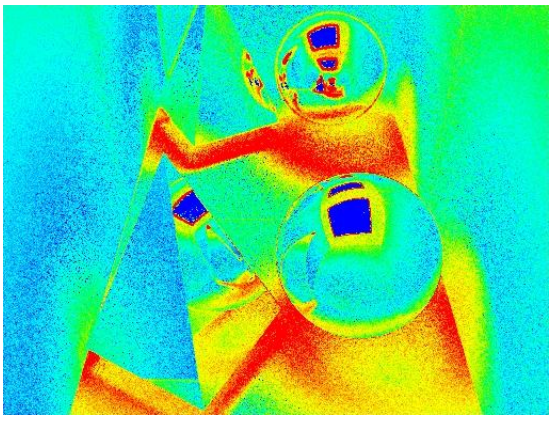
In this paper we have introduced a new family of refinement criteria based on *f*-divergences. These functions have been successfully used as discrimination measures in image processing and several engineering areas. We have applied these criteria to hierarchical radiosity and to adaptive supersampling in ray-tracing. In both areas, our results show the better behaviour of the *f*-divergence-based criteria compared with classic ones. In the hierarchical radiosity algorithm, the Kullback-Leibler criterion gives the best results, while in the

method		*RMS*
Classic	Contrast (CC)	6.157
	Variance (VAR)	5.194
f-divergences	Kullback-Leibler (KL)	5.508
	χ^2 (CS)	5.414
	Hellinger (HE)	5.807
Square root of *f*-divergences	Kullback-Leibler (SRKL)	4.824
	χ^2 (SRCS)	4.772
	Hellinger (SRHE)	4.595

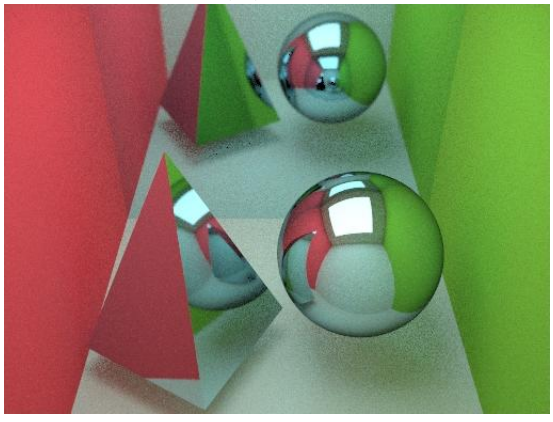
Table 1: *Root Mean Square Error (RMSE) for the different images in Fig. 5 and Fig. 6, with respect to the reference image in Fig. 4.*



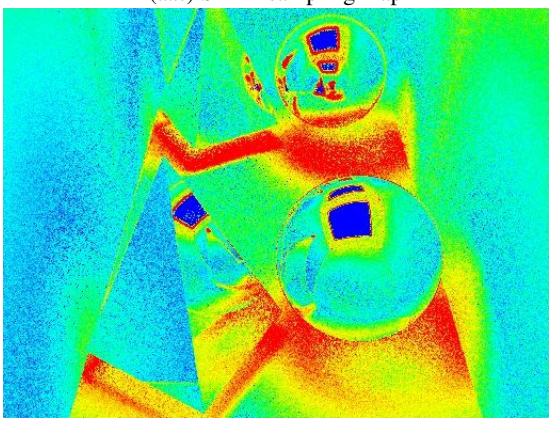
(a.*i*) SRKL



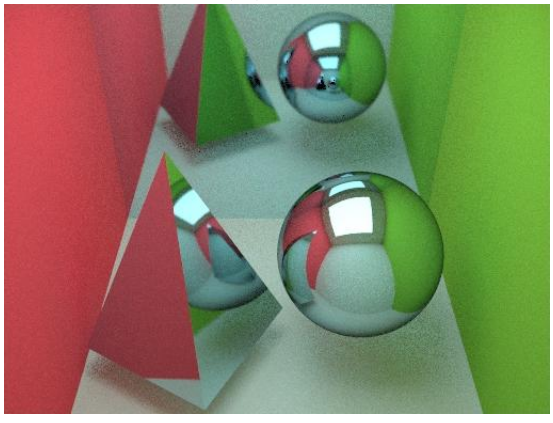
(a.*ii*) SRKL sampling map



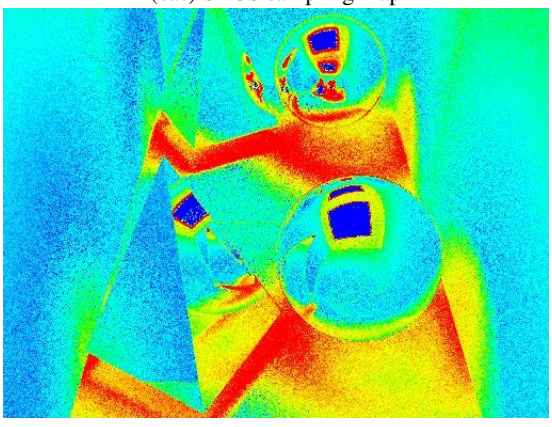
(b.*i*) SRCS



(b.*ii*) SRCS sampling map



(c.*i*) SRHE



(c.*ii*) SRHE sampling map

Figure 6: *Images obtained with an adaptive sampling scheme based on (a) Kullback-Leibler-based approach (SRKL), (b) χ^2-based approach (SRCS), and (c) Hellinger-based approach (SRHE). Column (i) shows the resulting images and (ii) the sampling density map. The average number of rays per pixel is 60 in all the methods, with a similar computation cost. Compare with the images in Fig. 5.*

ray-tracing algorithm the Hellinger-based refinement criterion is the most effective.

Our future work will be addressed towards finding new application areas for the *f*-divergences and investigating other families of divergences based on the Rényi entropy[33]. Also we will analyse the generalisation of the *f*-divergences presented in this paper, which can shed light on the good behaviour of the exponent value $\frac{1}{2}$ used in the ray-tracing case[21]. We will also address the problem of finding automatic criteria for the threshold used in the refinement test. Finally, in adaptive ray-tracing we will investigate why the criteria based on true distances behave better than the ones based on pseudodistances.

Acknowledgements

This project has been funded in part with grant numbers TIC-2001-2416-C03-01 and HU2000-0011 of the Ministry of Science and Technology and the Ministry of Foreign Affairs (Spanish Government), and 2001-SGR-00296 and ACI2002-29 of the Ministry of Universities, Research and the Information Society (Catalan Government). We would like to thank Josep A. Martín-Fernández and other members of the group *Anàlisi Estadística de Dades Composicionals* of the University of Girona for fruitful discussions.

References

1. M.S. Ali and S.D. Silvey. A general class of coefficients of divergence of one distribution from another. *Journal of Royal Statistical Society*, 28(1):131–142, 1966. 1, 3
2. Philippe Bekaert. *Hierarchical and Stochastic Algorithms for Radiosity*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, December 1999. 1, 2
3. Philippe Bekaert, László Neumann, Attila Neumann, Mateu Sbert, and Yves D. Willems. Hierarchical Monte Carlo radiosity. In George Drettakis and Nelson Max, editors, *Rendering Techniques'98 (Proceedings of the 9th Eurographics Workshop on Rendering)*, pages 259–268, New York (NY), USA, June 1998. Springer-Verlag Vienna-New York. Held in Vienna, Austria. 4
4. Pedro Bernaola-Galván, José L. Oliver, and Ramón-Roldán. Decomposition of DNA sequence complexity. *Physical Review Letters*, 83(16):3336–3339, October 1999. 3
5. Mark R. Bolin and Gary W. Meyer. A perceptually based adaptive sampling algorithm. *SIGGRAPH'98 Conference Proceedings*, pages 299–309, July 1998. Held in Orlando (FL), USA. 2
6. Terrence M. Caelli. *Visual Perception: Theory and Practice*. Pergamon Press, Oxford, UK, 1981. 2, 2
7. Michael F. Cohen and John R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Boston (MA), USA, 1993. 2
8. Computer Graphics Research Group. *RenderPark: A Photorealistic Rendering Tool*. Katholieke Universiteit Leuven, Leuven, Belgium, November 2000. 4, 7
9. Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991. 3
10. Imre Csiszár. Eine Informationsheoretische Ungleichung und ihre Anwendungen auf den Beweis der ergodizität von Markoffschen Ketten. *Magyar Tudományos Akadémia Közleményei*, (8):85–108, 1963. 1, 3
11. Mark A.Z. Dippé and Erling H. Wold. Antialiasing through stochastic sampling. *Computer Graphics (Proceedings of SIGGRAPH'85)*, 19(3):69–78, July 1985. Held in San Francisco (CA), USA. 2, 2, 2
12. Sever S. Dragomir. Some inequalities for the csiszár *f*-divergence. In *Inequalities for Csiszár f-Divergence in Information Theory*. Research Group in Mathematical Inequalities and Applications, Victoria University of Technology, Melbourne, Australia, 2000. 4
13. Miquel Feixas. *An Information-Theory Framework for the Study of the Complexity of Visibility and Radiosity in a Scene*. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, Desember 2002. 2, 2, 4
14. Miquel Feixas, Jaume Rigau, Philippe Bekaert, and Mateu Sbert. Information-theoretic oracle based on kernel smoothness for hierarchical radiosity. In *Short Presentations (Eurographics'02)*, pages 325–333, September 2002. Held in Saarbrücken, Germany. 2, 4
15. Andrew S. Glassner. *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995. 2, 2, 5
16. Cindy M. Goral, Kenneth E. Torrance, Donald P. Greenberg, and Bennett Battaile. Modelling the interaction of light between diffuse surfaces. *Computer Graphics (Proceedings of SIGGRAPH'84)*, 18(3):213–222, July 1984. Held in Minneapolis (MN), USA. 2
17. Steven J. Gortler, Peter Schröder, Michael F. Cohen, and Pat Hanrahan. Wavelet radiosity. *Computer Graphics (Proceedings of SIGGRAPH'93)*, 27:221–230, August 1993. Held in Anaheim (CA), USA. 2
18. Pat Hanrahan, David Salzman, and Larry Aupperle. A rapid hierarchical radiosity algorithm. *Computer Graphics (Proceedings of SIGGRAPH'91)*, 25(4):197–206, July 1991. Held in Las Vegas (NV), USA. 1, 2

19. Yun He, A. Ben Hamza, and Hamid Krim. A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing*, 51(5), May 2003. To appear. 3

20. Ernst D. Hellinger. Neue Begründung der Theorie quadratischen Formen von unendlichen vielen Veränderlichen. *Journal für Reine und Angewandte Mathematik*, (136):210–271, 1909. 4

21. Alfred O. Hero, Bing Ma, Olivier Michel, and John Gorman. Alpha-divergence for classification, indexing and retrieval. Technical Report CSPL-328, Communications and Signal Processing Laboratory, Ann Arbor (MI), USA, May 2001. 1, 3, 10

22. J. L. W. V. Jensen. Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, (30):175–193, 1906. 3

23. Samuel Kotz, Norman L. Johnson, and Campbell B. Read, editors. *Encyclopedia of Statistical Sciences*. Wiley Interscience, New York (NY), USA, 1983. 3

24. Solomon Kullback and R.A. Leibler. On information and sufficiency. *Annals of Mathematical Statistics*, (22):76–86, 1951. 4

25. Mark E. Lee, Richard A. Redner, and Samuel P. Uselton. Statiscally optimized sampling for distributed ray tracing. *Computer Graphics (Proceedings of SIGGRAPH'85)*, 19(3):61–67, July 1985. Held in San Francisco (CA), USA. 2, 3

26. Don P. Mitchell. Generating antialiased images at low sampling densities. *Computer Graphics (Proceedings of SIGGRAPH'87)*, 21(4):65–72, July 1987. Held in Anaheim (CA), USA. 1, 2, 2, 2

27. Joseph A. O'Sullivan, Richard E. Blahut, and Donald L. Snyder. Information-theoretic image formation. *IEEE Transactions on Information Theory*, 44(6):2094–2123, October 1998. 1

28. James Painter and Kenneth Sloan. Antialiased ray tracing by adaptive progressive refinement. *Computer Graphics (Proceedings of SIGGRAPH'89)*, 23(3):281–288, July 1989. Held in Boston (MA), USA. 2

29. Karl Pearson. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Psychology Magazine*, 1:157–175, 1900. 4

30. Josien P.W. Pluim. *Mutual Information Based Registration of Medical Images*. PhD thesis, Utrecht University, Utrecht, The Netherlands, 2001. 1, 3

31. Werner Purgathofer. A statistical method for adaptive stochastic sampling. *Eurographics'86: Proceedings of the European Conference and Exhibition*, 11(2):157–162, August 1986. Held in Lisbon, Portugal. 1, 2, 3, 3

32. C.Radhakrishna Rao. *Rao's axiomatization of Diversity Measures (Encyclopedia of Statistical Sciences)*, pages 614–617. Wiley Interscience, New York (NY), USA, 1983. 3

33. Alfréd Rényi. On measures of entropy and information. *Selected Papers of Alfréd Rényi*, 2:565–580, 1976. 10

34. François X. Sillion and Claude Puech. *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1994. 2

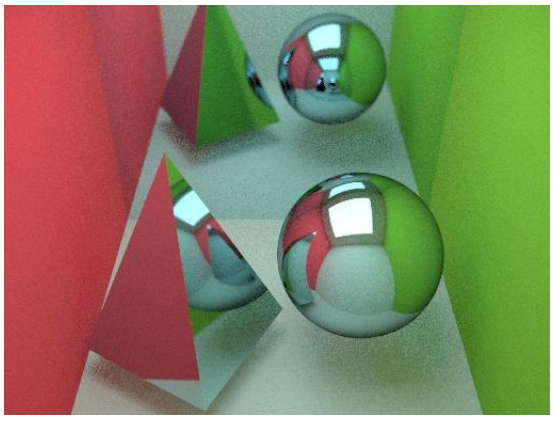
35. László Szirmay-Kalos. *Monte Carlo Methods in Global Illumination*. Institute of Computer Graphics, Vienna University of Technology, Vienna, Austria, 1999. 2

36. Rasmus Tamstorf and Henrik W. Jensen. Adaptive sampling and bias estimation in path tracing. In Julie Dorsey and Philipp Slusallek, editors, *Rendering Techniques'97 (Proceedings of the 8th Eurographics Workshop on Rendering)*, pages 285–295, New York (NY), USA, June 1997. Springer-Verlag Vienna-New York. Held in St. Etienne, France. 2, 3

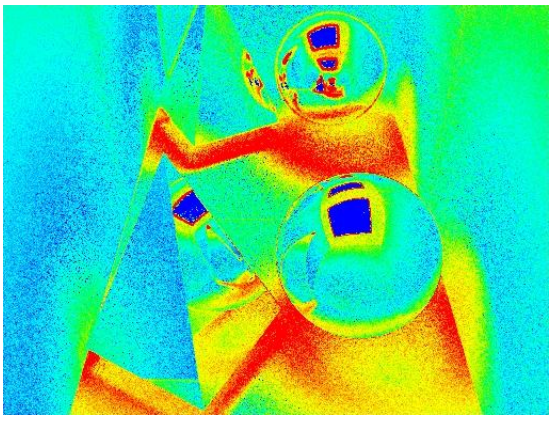
37. Flemming Topsoe. Some inequalities for information divergence and related measures of discrimination. *Research Report Collection*, 2(1):Article 9, 1999. 5

38. Turner Whitted. An improved illumination model for shaded display. *Communications of the ACM*, 23(6):343–349, June 1980. 2

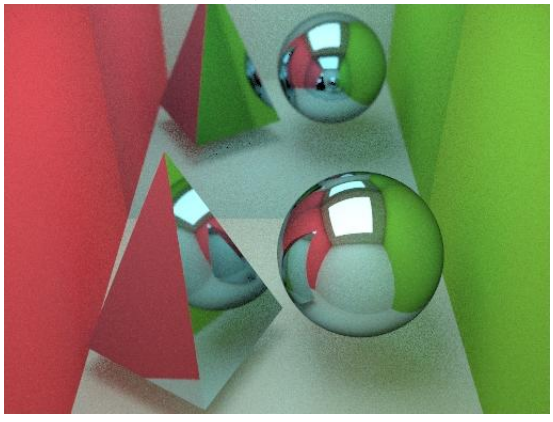
39. Yuhong Yang and Andrew Barron. Information theoretic determination of minimax rates of convergence. *Annals of Statistics*, 27:1546–1599, 1999. 5



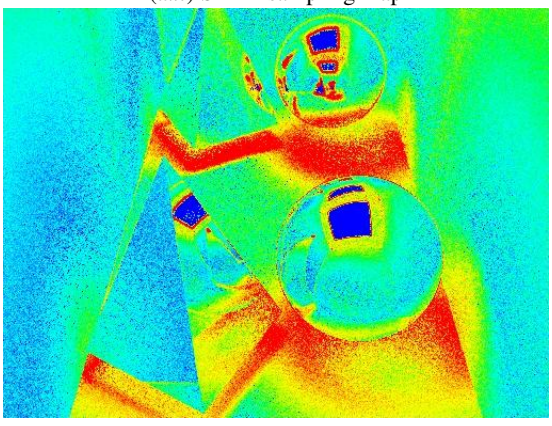
(a.*i*) SRKL



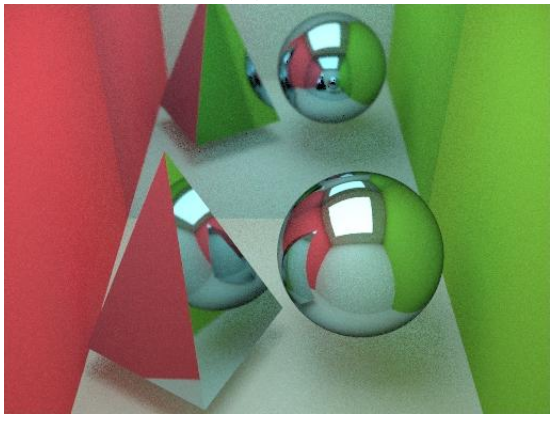
(a.*ii*) SRKL sampling map



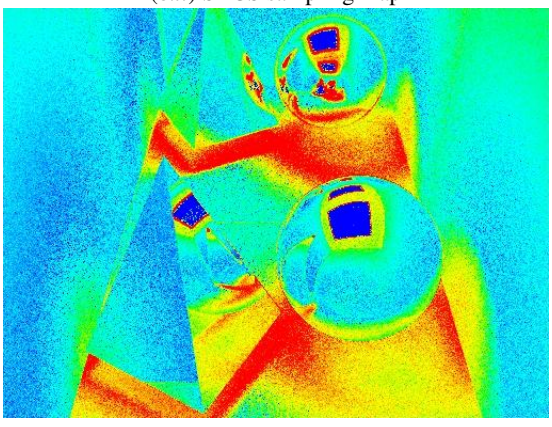
(b.*i*) SRCS



(b.*ii*) SRCS sampling map



(c.*i*) SRHE



(c.*ii*) SRHE sampling map

Figure 6: *Images obtained with an adaptive sampling scheme based on (a) Kullback-Leibler-based approach (SRKL), (b) χ^2-based approach (SRCS), and (c) Hellinger-based approach (SRHE). Column (i) shows the resulting images and (ii) the sampling density map. The average number of rays per pixel is 60 in all the methods, with a similar computation cost. Compare with the images in Fig. 5.*