Applications of Information Theory to Computer Graphics Part I: Introduction

Mateu Sbert*, Miquel Feixas*, Jaume Rigau*, Ivan Viola** and Miguel Chover***

* University of Girona, Spain
** University of Bergen, Norway
*** Jaume I University, Spain

Abstract
We present different applications of information theory to computer graphics, based on the use of the measures of entropy, mutual information, f-divergences and generalized entropies. The application areas are hierarchical radiosity, adaptive ray-tracing, selection of best viewpoints, object and scene exploration, mesh saliency, mesh simplification and scientific visualization. We also give some hints on information-theoretic applications to object recognition and image processing.

Categories and Subject Descriptors (according to ACM CCS): I.3.3 [Computing Methodologies]: Computer Graphics

1. Presentation

These are the notes for the Eurographics 2007 tutorial *Applications of Information theory to Computer Graphics*. After an introduction to the basics concepts of Information Theory, different applications are presented to the fields of radiosity, adaptive ray-tracing, selection of best viewpoints, object and scene exploration, mesh saliency, mesh simplification and scientific visualization. These applications are based on the use of the measures of entropy, mutual information, f-divergences and generalized entropies. Other applications, such as image-based rendering, object recognition, image processing and light positioning, do not appear in the current version of these notes.

This document is organized in the following way. After this introductory part, second part deals with the basics of Information Theory (IT). The concept of information channel is introduced, and the quantities of entropy and mutual information are defined together with important relationships such as the Jensen-Shannon inequality. Finally the information bottleneck method is explained.

Third part deals with applications to radiosity. Radiosity is a viewpoint independent global illumination technique that discretizes the scene into small polygons or patches to solve a transport system of equations. The way the scene is discretized is critical for the efficiency of the result. We define

first a scene information channel, which allows us to study the interchange of information between the patches. From the study of this channel several refinement oracles, i.e., criteria for subdividing the geometry, are obtained, aimed at maximizing the transport of information. Both classic and generalized information-theoretic quantities are used for this.

Fourth part is about adaptive ray-tracing. This technique is aimed at tracing more rays only where they are needed. For instance, smoothly illuminated regions of the scene with low variation do not need as much effort as rapidly varying illumination or also geometric discontinuities. The information theory quantities will be used again to define adaptive refinement oracles. New oracles are also defined for radiosity and adaptive ray-tracing using the following f-divergences: Kullback-Leibler, Chi-square, and Hellinger distances.

In the fifth part we define a viewpoint information channel between the points of view around an object and the polygons of the object. Several quantities associated to this channel, such as mutual information and entropy, are interpreted in terms of viewpoint quality measures. Viewpoint similarity and stability are defined, as well as methods for the selection of best n-views and for exploring the object. Mesh saliency is interpreted in terms of the viewpoint channel and polygonal mutual information as an ambient occlusion quantity. Impor-

tance is introduced into the scheme and saliency is used as importance to guide the viewpoint selection.

Sixth part deals with view selection in scientific visualization. The problem and context of the visualization of volumetric data sets is presented, together with the different viewpoint quality measures used. The framework presented in part sixth is then applied to this context. Guided navigation using higher-level semantics is also studied.

Finally, seventh part is about viewpoint-driven simplification. The several simplification algorithms are based on the variations perceived in image space, measured in our case through information-theoretic metrics: entropy, Kullback-Leibler distance and mutual information. These techniques are shown to give a better simplified mesh than object-based approaches, although at the cost of an increased processing time.

2. Authors

- Mateu Sbert
  University of Girona, Spain
  e-mail: mateu@ima.udg.edu
  Mateu Sbert is a professor in Computer Science at the University of Girona, Spain. He received a M.Sc. in Theoretical Physics (1977) at the university of Valencia, a M.Sc. in Mathematics (Statistics and Operations Research, 1983) at U.N.E.D. university (Madrid) and a Ph.D. in Computer Science at the U.P.C.(Universitat Politècnica de Catalunya, 1997, Best PhD Award). Mateu Sbert's research interests include the application of Monte Carlo, Integral Geometry and Information Theory techniques to Radiosity, Global Illumination and Image Based Rendering. He has authored or co-authored more than 100 papers in his areas of research, participated in several Eurographics tutorials and served as a member of program committee in Spanish and international conferences. He has participated in several European and Spanish research projects, leading the VIth European Framework Gametools project. Mateu Sbert coorganized the Dagsthul Seminars number 01242, Stochastic Methods in Rendering and number 06221, Computational Aesthetics in Graphics, Visualization and Imaging.
  http://ima.udg.edu/~mateu/
- Miquel Feixas
  University of Girona, Spain
  e-mail: feixas@ima.udg.edu
  http://ima.udg.edu/~feixas/
  Miquel Feixas is an associate professor in Computer Science at the University of Girona, Spain. He received a M.Sc. in Theoretical Physics (1979) at the UAB (Universitat Autònoma de Barcelona) and a Ph.D. in Computer Science at the UPC (Universitat Politècnica de Catalunya, 2002). His research is focused on the application of Information Theory techniques to Radiosity, Global Illumination, Viewpoint Selection and Image Processing. He

has co-authored several papers in his area of research. He acted as a reviewer for conferences in the field of computer graphics and image processing. He has participated in Spanish research projects and joint actions with several European universities.
- Jaume Rigau
  University of Girona, Spain
  e-mail: rigau@ima.udg.edu
  http://ima.udg.edu/~rigau/
  Jaume Rigau is an associate professor in Computer Science at the University of Girona, Spain. He received a M.Sc. in Computer Science (1993) at the UPC (Universitat Politècnica de Catalunya) and a Ph.D. by the UPC (2006). His research is focused on the application of Information-Theory techniques to Radiosity, Global Illumination and Image Processing. He has co-authored several papers in his area of research. He has participated in several Spanish research projects and joint actions with several European universities.
- Ivan Viola
  University of Bergen, Norway
  e-mail: ivan@ii.uib.no
  http://www.ii.uib.no/~ivan/
  Ivan Viola is a Post Doctor research associate at University of Bergen, Norway. He was formerly associated with Vienna University of Technology, Austria, where he received M.Sc. in 2002 and Ph.D. in 2005. His research is focused on development of novel methods for automatically generating expressive visualizations of complex data. Viola co-authored several scientific works published on international conferences such as IEEE Visualization, EuroVis, and Vision Modeling and Visualization and acted as a reviewer and program committee member for conferences in the field of computer graphics and visualization. Recently he co-organized series of tutorials on Illustrative Visualization.
- Miguel Chover
  Jaume I University, Spain
  e-mail: chover@uji.es
  http://www3.uji.es/~chover/
  Miguel Chover is an associate professor at the Universitat Jaume I de Castelló, Spain. His research focused on interactive computer graphics, computer games and Web3D. His current work includes level of detail modelling, simplification algorithms, rendering natural phenomena and texturing techniques. He received his MS degree in Computer Science in 1992 and a PhD in Computer Science in 1996 from the Universidad Politécnica de Valencia, Spain. He is member of Eurographics.

3. Syllabus

1. Introduction (5 min)
   Speaker: Mateu Sbert
2. Information Theory Basics (35 min)
   Speaker: Miquel Feixas

- Information channel: entropy and mutual information
- Important inequalities
- Information bottleneck method

3. Refinement Criteria for Radiosity (20 min)
   Speaker: Jaume Rigau

   - Scene information channel
   - Refinement criteria for hierarchical radiosity
   - Mutual-information-based oracle
   - f-divergence-based oracles

4. Adaptive Refinement for Ray-tracing (15 min)
   Speaker: Jaume Rigau

   - Refinement criteria for ray-tracing
   - Entropy-based refinement criteria
   - f-divergence-based refinement criteria

5. Viewpoint Selection and Mesh Saliency (30 min)
   Speaker: Mateu Sbert

   - Viewpoint information-theoretic measures
   - Viewpoint information channel: mutual information, similarity and stability
   - Selection of best views and object exploration
   - Polygonal mutual information and information-theoretic ambient occlusion
   - Mesh saliency
   - Importance-driven viewpoint selection

6. View Selection in Scientific Visualization (30 min)
   Speaker: Ivan Viola

   - View Selection for Volumes and Iso-Surfaces
   - Importance-Driven Focus of Attention
   - Guided Navigation using Higher-Level Semantics

7. Viewpoint-driven Simplification (30 min)
   Speaker: Miguel Chover

   - Recent work on simplification
   - Information-theoretic metrics: entropy, Kullback-Leibler distance and mutual information
   - Simplification algorithms

8. Other Applications (15 min)
   Speaker: Miquel Feixas and Mateu Sbert

   - Image processing: registration and segmentation
   - Object recognition: shape descriptors

Aknowledgements

Applications of Information Theory to Computer Graphics
Part II: Information Theory Basics

Miquel Feixas, Jaume Rigau, and Mateu Sbert

University of Girona, Spain

1. Introduction

In 1948, Claude Shannon published "A mathematical theory of communication" [Sha48] which marks the beginning of information theory. In this paper, he defined measures such as entropy and mutual information (called rate of transmission), and introduced the fundamental laws of data compression and transmission.

Information theory deals with the transmission, storage and processing of information and is used in fields such as physics, computer science, mathematics, statistics, economics, biology, linguistics, neurology, learning, etc. It is applied successfully in areas such as medical image processing, computer vision, robot motion and computer graphics.

In this part, we present some basic concepts of information theory. A very good reference is the text by Cover and Thomas [CT91]. Other references are Blahut [Bla87] and Lubbe [vdL97].

2. Entropy

In [Sha48], after representing a discrete information source as a Markov process, Shannon asks himself: "Can we define a quantity which will measure, in some sense, how much *information* is "produced" by such a process, or better, at what rate information is produced?".

His answer is: "Suppose we have a *set of possible* events whose *probabilities of occurrence* are p_1, p_2, ..., p_n. These probabilities are known but that is all we know concerning which event will occur. Can we find a measure of how much "choice" is involved in the selection of the event or of how uncertain we are of the outcome?

If there is such a measure, say $H(p_1, p_2, \ldots, p_n)$, it is reasonable to require of it the following properties:

1. H would be continuous in the p_i.
2. If all the p_i are equal, $p_i = \frac{1}{n}$, then H should be a monotonic increasing function of n. With equally likely events

there is more choice, or uncertainty, when there are more possible events.
3. If a choice is broken down into two successive choices, the original H should be the weighted sum of the individual values of H. The meaning of this is illustrated in Figure 1.
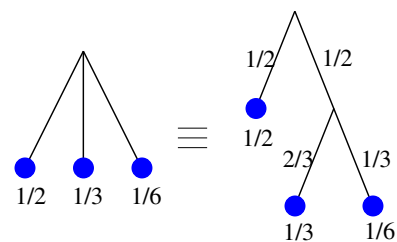

Figure 1: *Grouping property of the entropy.*

On the left, we have three possibilities $p_1 = \frac{1}{2}$, $p_2 = \frac{1}{3}$, $p_3 = \frac{1}{6}$. On the right, we first choose between two possibilities each with probability $\frac{1}{2}$, and if the second occurs, we make another choice with probabilities $\frac{2}{3}$, $\frac{1}{3}$. The final results have the same probabilities as before. We require, in this special case, that $H(\frac{1}{2}, \frac{1}{3}, \frac{1}{6}) = H(\frac{1}{2}, \frac{1}{2}) + \frac{1}{2}H(\frac{2}{3}, \frac{1}{3})$. The coefficient $\frac{1}{2}$ is because this second choice only occurs half the time."

After these requirements, he introduces the following *theorem*: "The only H satisfying the three above assumptions is of the form:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \qquad (1)$$

where K is a positive constant". When $K = 1$ and the logarithm is $\log_2$, information is measured in bits.

Shannon calls this quantity *entropy*, as "the form of H will be recognized as that of entropy as defined in certain formulations of statistical mechanics where p_i is the probability of

a system being in cell *i* of its phase space". There are other axiomatic formulations which involve the same definition of entropy [CT91].

The Shannon entropy is the classical measure of *information*, where information is simply *the outcome of a selection among a finite number of possibilities*. Entropy also measures *uncertainty* or *ignorance*.

Thus, the *Shannon entropy* $H(X)$ of a discrete random variable X with values in the set $S = \{x_1, x_2, \ldots, x_n\}$ is defined as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \qquad (2)$$

where $n = |\mathcal{X}|$, $p_i = p(x_i) = Pr[X = x_i]$ for $i \in \{1, \ldots, n\}$, the logarithms are taken in base 2 (entropy is expressed in bits), and we use the convention that $0 \log 0 = 0$, which is justified by continuity. We can use interchangeably the notation $H(X)$ or $H(p)$ for the entropy, where p is the probability distribution $\{p_1, p_2, \ldots, p_n\}$, also represented by p_i. As $-\log p_i$ represents the *information* associated with the result x_i, the entropy gives us the *average information* or *uncertainty* of a random variable. Information and uncertainty are opposite. Uncertainty is considered before the event, information after. So, information reduces uncertainty. Note that the entropy depends only on the probabilities.

Some other relevant properties [Sha48] of the entropy are

1. $0 \leq H(X) \leq \log n$

   - $H(X) = 0$ if and only if all the probabilities except one are zero, this one having the unit value, i.e., when we are certain of the outcome.
   - $H(X) = \log n$ when all the probabilities are equal. This is the most uncertain situation.

2. If we equalize the probabilities, entropy increases.

If we consider another random variable Y with probability distribution q_i corresponding to values in the set $\mathcal{Y} = \{y_1, y_2, \ldots, y_m\}$, the *joint entropy* of X and Y is defined as

$$H(X,Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log p_{ij} \qquad (3)$$

where $m = |S'|$ and $p_{ij} = P(x_i, y_j) = Pr[X = x_i, Y = y_j]$ is the joint probability.

When $n = 2$, the *binary* entropy (Figure 2) is given by

$$H(X) = -p \log p - (1-p) \log (1-p) \qquad (4)$$

where $p = \{p, 1-p\}$.

Also, the *conditional entropy* is defined as

$$H(X|Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j} \qquad (5)$$

where $p_{i|j} = p_{x_i|y_j} = Pr[X = x_i|Y = y_j]$ is the conditional probability.
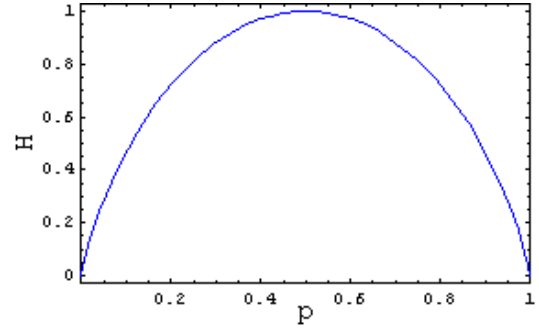

Figure 2: *Binary entropy.*

The Bayes theorem expresses the relation between the different probabilities:

$$p_{ij} = p_i p_{j|i} = q_j p_{i|j} \qquad (6)$$

If X and Y are *independent*, then $p_{ij} = p_i q_j$.

The conditional entropy can be thought of in terms of a *channel* whose input is the random variable X and whose output is the random variable Y. $H(X|Y)$ corresponds to the uncertainty in the channel input from the receiver's point of view, and vice versa for $H(Y|X)$. Note that in general $H(X|Y) \neq H(Y|X)$.

The following properties are also met:

1. $H(X,Y) \leq H(X) + H(Y)$
2. $H(X,Y) = H(X) + H(Y|X) = H(Y) + H(X|Y)$
3. $H(X) \geq H(X|Y) \geq 0$

3. Mutual Information

The *mutual information* between two random variables X and Y is defined as

$$\begin{aligned}
I(X,Y) &= H(X) - H(X|Y) \\
&= H(Y) - H(Y|X) \\
&= -\sum_{i=1}^{n} p_i \log p_i + \sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j} \qquad (7)
\end{aligned}$$

Mutual information represents the amount of information that one random variable, the output of the channel, gives (or contains) about a second random variable, the input of the channel, and vice versa, i.e., how much the knowledge of X decreases the uncertainty of Y and vice versa. Therefore, $I(X,Y)$ is a measure of the shared information between X and Y.

Mutual information $I(X,Y)$ has the following properties:

1. $I(X,Y) \geq 0$ with equality if, and only if, X and Y are independent.

2. $I(X,Y) = I(Y,X)$
3. $I(X,Y) = H(X) + H(Y) - H(X,Y)$
4. $I(X,Y) \leq H(X)$

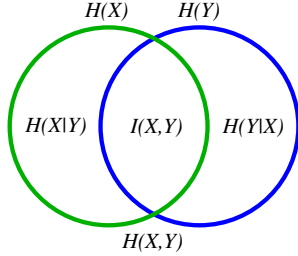The relationship between all the above measures can be expressed by the Venn diagram, as shown in Figure 3.

Figure 3: *Venn diagram of a discrete channel.*

The *relative entropy* or *Kullback-Leibler distance* between two probability distributions $p = \{p_i\}$ and $q = \{q_i\}$, that are defined over the set S, is defined as

$$D_{KL}(p\|q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \qquad (8)$$

where, from continuity, we use the convention that $0\log 0 = 0$, $p_i \log \frac{p_i}{0} = \infty$ if $a > 0$ and $0 \log \frac{0}{0} = 0$.

The relative entropy is "a measure of the inefficiency of assuming that the distribution is q when the true distribution is p" [CT91].

The relative entropy satisfies the *information inequality* $D_{KL}(p\|q) \geq 0$, with equality only if $p = q$. The relative entropy is also called *discrimination* and it is not strictly a distance, since it is not symmetric and does not satisfy the triangle inequality. Moreover, we have to emphasize that the mutual information can be expressed as

$$I(X,Y) = D_{KL}(\{p_{ij}\}\|\{p_i q_j\}) \qquad (9)$$

4. Entropy Rate of a Markov Chain

The joint entropy of a collection of n random variables is given by

$$H(X_1,\ldots,X_n) =$$
$$H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1},\ldots,X_1) \qquad (10)$$

The *entropy rate* or *entropy density* of a stochastic process $\{X_i\}$ is defined by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} H(X_1,X_2,\ldots,X_n) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1},\ldots,X_1) \end{aligned} \qquad (11)$$

representing the *average information content* per output symbol [†] [CT91]. It is the "uncertainty associated with a

---

[†] At least, h exists for all stationary stochastic processes.

given symbol if all the preceding symbols are known" and can be viewed as "the intrinsic *unpredictability*" or "the irreducible *randomness*" associated with the chain [FC98].

In particular, a Markov chain can be considered as a chain of random variables complying with

$$H(X_n|X_1,X_2,\ldots,X_{n-1}) = H(X_n|X_{n-1}) \qquad (12)$$

An important result is the following theorem: For a stationary Markov chain, with stationary distribution w_i, the entropy rate or information content is given by

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} H(X_1,X_2,\ldots,X_n) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1}) \\ &= H(X_2|X_1) = -\sum_{i=1}^{n} w_i \sum_{j=1}^{n} P_{ij} \log P_{ij} \end{aligned} \qquad (13)$$

where w_i is the equilibrium distribution and P_{ij} is the transition probability from state i to state j.

Finally, the *excess entropy* or *effective measure complexity* [CP83, Gra86, Sha84, SG86] of an infinite chain is defined by

$$E = \lim_{n\to\infty} (H(X_1,X_2,\ldots,X_n) - nh) \qquad (14)$$

where h is the entropy rate of the chain and n is the length of this chain. The excess entropy can be interpreted as the mutual information between two semi-infinite halves of the chain. "Another way of viewing this, is that excess entropy is the *cost of amnesia* – the excess entropy measures how much more random the system would become if we suddenly forgot all information about the left half of the string" [Fel97].

5. Important Inequalities

Some of the above properties can be deduced from the following inequalities [CT91].

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 $0 \leq \lambda \leq 1$,

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \qquad (15)$$

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, $x \log x$ for $x \geq 0$ is a strictly convex function, and $\log x$ for $x \geq 0$ is a strictly concave function [CT91].

Jensen's inequality: If f is convex on the range of a random variable X, then

$$f(E[X]) \leq E[f(X)] \qquad (16)$$

where E denotes expectation. Moreover, if $f(x)$ is strictly

convex, the equality implies that $X = E[X]$ with probability 1, i.e., X is a deterministic random variable with $Pr[X = x_0] = 1$ for some x_0.

One of the most important consequences of Jensen's inequality is the information inequality $D_{KL}(p\|q) \geq 0$. Other previous properties can also be derived from this inequality.

Observe that if $f(x) = x^2$ (convex function), then $E[X^2] - (E[X])^2 \geq 0$. So, the variance is invariably positive.

If f is given by the Shannon entropy, which is a concave function, we obtain the *Jensen-Shannon inequality* [BR82]:

$$JS(\pi_1, \pi_2, \ldots, \pi_N; p_1, p_2, \ldots, p_N)$$
$$\equiv H\left(\sum_{i=1}^{N} \pi_i p_i\right) - \sum_{i=1}^{N} \pi_i H(p_i) \geq 0, \qquad (17)$$

where $JS(\pi_1, \pi_2, \ldots, \pi_N; p_1, p_2, \ldots, p_N)$ is the *Jensen-Shannon divergence* of probability distributions $p_1, p_2, \ldots, p_N$ with prior probabilities or weights $\pi_1, \pi_2, \ldots, \pi_N$, fulfilling $\sum_{i=1}^{N} \pi_i = 1$. The JS-divergence measures how 'far' are the probabilities p_i from their likely joint source $\sum_{i=1}^{N} \pi_i p_i$ and equals zero if and only if all the p_i are equal. It is important to note that the JS-divergence is identical to $I(X, Y)$ when $\pi_i = p(x_i)$ and $p_i = p(Y|x_i)$ for each $x_i \in \mathcal{X}$, where $p(X) = \{p(x_i)\}$ is the input distribution, $p(Y|x_i) = \{p(y_1|x_i), p(y_2|x_i), \ldots, p(y_M|x_i)\}$, $N = |\mathcal{X}|$, and $M = |\mathcal{Y}|$ [BR82, ST00]. This notation is followed in some parts of the tutorial.

The log-sum inequality

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

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \qquad (18)$$

with equality if and only if $\frac{a_i}{b_i} = constant$.

Note that the conditions in this inequality are much weaker than for Jensen's inequality.

From this inequality, certain results can be derived:

1. $D_{KL}(p\|q)$ is convex in the pair (p, q)
2. $H(X)$ is a concave function of p
3. If X and Y have the joint pdf $p(x, y) = p(x)p(y|x)$, then $I(X, Y)$ is a concave function of $p(x)$ for fixed $p(y|x)$ and a convex function of $p(y|x)$ for fixed $p(x)$.

Data processing inequality

Data processing inequality: If $X \to Y \to Z$ is a Markov chain, then

$$I(X, Y) \geq I(X, Z) \qquad (19)$$

This result demonstrates that no processing of Y, deterministic or random, can increase the information that Y contains about X.

Fano's inequality

Suppose we have two correlated random variables X and Y and we wish to measure the probability of error in guessing X from the knowledge of Y. Fano's inequality gives us a tight lower bound on this error probability in terms of the conditional entropy $H(X|Y)$ [CT91, FM94]. As $H(X|Y)$ is zero if and only if X is a function of Y, we can estimate X from Y with zero probability of error if and only if $H(X|Y) = 0$. Intuitively, we expect to be able to estimate X with a low probability of error if and only if $H(X|Y)$ is small [CT91].

If X and Y have the joint pdf $p(x, y) = p(x)p(y|x)$, from Y we calculate a function $g(Y) = \widehat{X}$ which is an estimate of X. Observe that $X \to Y \to \widehat{X}$ is a Markov chain. The probability of error is defined by

$$P_e = Pr[\widehat{X} \neq X] \qquad (20)$$

Fano's inequality:

$$H(P_e) + P_e \log n \geq H(X|Y) \qquad (21)$$

where $H(P_e)$ is the binary entropy from $\{P_e, 1 - P_e\}$.

This inequality can be weakened to

$$P_e \geq \frac{H(X|Y) - 1}{\log n} \qquad (22)$$

Thus, Fano's inequality bounds the probability that $\widehat{X} \neq X$.

6. Entropy and Coding

Other ways of interpreting the Shannon entropy are possible:

- As we have seen in section 2, $-\log p_i$ represents the *information* associated with the result x_i. But $-\log p_i$ can also be interpreted as the *surprise* associated with the outcome x_i. If p_i is small, the surprise is large; if p_i is large, the surprise is small. Thus, the entropy

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i$$

is the expectation value of the surprise [Fel97].
- Entropy is also related to the difficulty in guessing the outcome of a random variable. Thus, it can be seen [CT91, Fel97] that

$$H(X) \leq \overline{questions} < H(X) + 1 \qquad (23)$$

where $\overline{questions}$ is the average minimum number of binary questions to determine X. This idea agrees with the interpretation of entropy as a measure of uncertainty and also with the next interpretation.
- A fundamental result of information theory is the Shannon source coding theorem, which deals with the encoding of an object in order to store or transmit it efficiently [CT91, Fel97]. "Data compression can be achieved by assigning short descriptions to the most frequent outcomes

of the data source and necessarily longer descriptions to the less frequent outcomes" [CT91]. For instance, Huffman instantaneous coding [‡] is optimal and fulfils the following theorems:

– Similarly to (23), we have

$$H(X) \leq \overline{\ell} < H(X) + 1 \qquad (24)$$

where $\overline{\ell}$ is the expected length of the optimal binary code for X.

– If we encode n identically distributed random variables X with a binary code, the Shannon source coding theorem can be enunciated in the following way:

$$H(X) \leq \overline{\ell_n} < H(X) + \frac{1}{n} \qquad (25)$$

where $\overline{\ell_n}$ is the expected codeword length per unit symbol. Thus, by using large block lengths, we can achieve an expected codelength per symbol arbitrarily close to the entropy [CT91].

– For a stationary stochastic process, we have

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \overline{\ell_n} < \frac{H(X_1, X_2, \ldots, X_n)}{n} + 1 \quad (26)$$

and thus, by definition of entropy rate h (11),

$$\lim_{n \to \infty} \overline{\ell_n} \to h \qquad (27)$$

Thus, the entropy rate is the expected number of bits per symbol required to describe the stochastic process.

In conclusion, the entropy of a random variable is a measure of the amount of information required on average to describe it.

7. Continuous Channel

In this section, entropy and mutual information are defined for continuous sources of information. For a continuous source X, messages are taken from a continuous set S.

The entropy of a discrete set of probabilities p has been defined (2) as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \qquad (28)$$

Similarly, the *continuous entropy* of a continuous random variable X with a probability density function $p(x)$ is defined by

$$H^c(X) = -\int_S p(x) \log p(x) dx \qquad (29)$$

In the same way, for two continuous random variables X and Y, the continuous conditional entropy is defined as

$$H^c(X|Y) = -\int_S \int_S p(x,y) \log p(x|y) dx dy \qquad (30)$$

[‡] A code is called a prefix or instantaneous code if no codeword is a prefix of any other codeword.

and the continuous mutual information is defined as

$$I^c(X,Y) = \int_S \int_S p(x,y) \log \frac{p(x,y)}{p(x)p(y)} dx dy \qquad (31)$$

where $p(x|y)$ and $p(x,y)$ are, respectively, the conditional density function and the joint density function associated with X and Y.

If we divide the range of the continuous random variable X into n bins of length Δ, and we consider its discretised version X^Δ (see [CT91]), it can be seen that the entropy of a continuous random variable does not equal the entropy of the discretised random variable in the limit of a finer discretisation [Sha48, CT91, Fel97]:

$$\lim_{\Delta \to 0} H(X^\Delta) = H^c(X) - \log \Delta \qquad (32)$$

On the other hand, the mutual information between two continuous random variables X and Y is the limit of the mutual information between their discretised versions. Thus, when the number of bins tends to infinity:

$$\lim_{\Delta \to 0} I(X^\Delta, Y^\Delta) = I^c(X,Y) \qquad (33)$$

In addition, Kolmogorov [Kol56] and Pinsker [Pin60] defined mutual information as $I(X,Y) = \sup_{P,Q} I([X]_P, [Y]_Q)$, where the supremum is over all finite partitions P and Q. From this definition, two important properties can be deduced: *the continuous mutual information is the least upper bound for the discrete mutual information* and *refinement can never decrease the discrete mutual information* [Gra90]. This last property can also be deduced from the data processing inequality (19) [Gra90].

References

[Bla87] BLAHUT R. E.: *Principles and Practice of Information Theory*. Addison-Wesley, Reading (MA), USA, 1987.

[BR82] BURBEA J., RAO C. R.: On the convexity of some divergence measures based on entropy functions. *IEEE Transactions on Information Theory 28*, 3 (May 1982), 489–495.

[CP83] CRUTCHFIELD J. P., PACKARD N.: Symbolic dynamics of noisy chaos. *Physica 7D* (1983), 201–223.

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[FC98] FELDMAN D. P., CRUTCHFIELD J. P.: *Discovering Noncritical Organization: Statistical Mechanical, Information Theoretic and Computational Views of Patterns in One-Dimensional Spin Systems*. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998.

[Fel97]   FELDMAN D. P.: A brief introduction to: Information theory, excess entropy and computational mechanics, 1997.

[FM94]   FEDER M., MERHAV N.: Relations between entropy and error probability. *IEEE Transactions on Information Theory 40*, 1 (January 1994), 259–266.

[Gra86]   GRASSBERGER P.: Toward a quantitative theory of self-generated complexity. *International Journal of Theoretical Physics 25*, 9 (1986), 907–938.

[Gra90]   GRAY R. M.: *Entropy and Information Theory*. Springer-Verlag Vienna-New York, New York (NY), USA, 1990.

[Kol56]   KOLMOGOROV A. N.: On the shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory 2* (1956), 102–108.

[Pin60]   PINSKER M. S.: *Information and Stability of Random Variables and Processes*. Izdatel'stvo Akademii Nauk SSSR, Moscow, Russia, 1960. Translated by A. Feinstein, 1964.

[SG86]   SZÉPFALUSY P., GYÖRGYI G.: Entropy decay as a measure of stochasticity in chaotic systems. *Physical Review A 33*, 4 (1986), 2852.

[Sha48]   SHANNON C. E.: A mathematical theory of communication. *The Bell System Technical Journal 27* (July, October 1948), 379–423, 623–656.

[Sha84]   SHAW R.: *The Dripping Faucet as a Model Chaotic System*. Aerial Press, Santa Cruz (CA), USA, 1984.

[ST00]   SLONIM N., TISHBY N.: Document clustering using word clusters via the information bottleneck method. In *Proceedings of the 23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval* (2000), ACM Press, pp. 208–215. Held in Athens, Greece.

[vdL97]   VAN DER LUBBE J. C.: *Information Theory*. Cambridge University Press, Cambridge, UK, 1997.

Applications of Information Theory to Computer Graphics Part III-IV: Refinement Criteria Based on f-Divergences

Jaume Rigau, Miquel Feixas, and Mateu Sbert

University of Girona, Spain

1. Introduction

Just as we have seen in the previous part, 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 homogeneity encountered in the samples. Heterogeneity should lead to further sampling, possibly with an adaptive subdivision of the domain. Oracles are then built based on these criteria.

In this part, we introduce new refinement criteria based on *f-divergence*s. The introduction of these measures is motivated by the observation that the mutual information-based oracle (see Part III) can be rewritten as an *f*-divergence. *f*-Divergences are a family of convex functions that possess very remarkable properties. They were introduced by Csiszár [Csi63] and Ali and Silvey [AS66] as measures of discrimination or distance between probability distributions and have been successfully used in image processing and several engineering areas [OBS98, LeG99, HMMG01, Plu01].

Our purpose is to demonstrate the usefulness of *f*-divergences in computer graphics by applying them to defining new refinement criteria for the techniques of the previous chapters: hierarchical radiosity (§4) and adaptive sampling in ray-tracing (§5). We consider that some divergences are perfectly fitted as homogeneity measures, when we consider how distant the distribution of the samples is with respect to the uniform distribution. We will see how, compared with classic refinement criteria, the *f*-divergence-based ones give significantly better results. Previously, we give a brief introduction to divergence measures (§2) and in particular to the *f*-divergences (§3).

2. Divergence Measures

In this section, we establish the semantics of divergence measures (§2.1) and we present three specific types of them (§2.2).

2.1. Concept

What does "divergence" mean? One brief definition for divergence is *"a deviation from a course or standard"* [Mer05]. In general, the difference in shades of meaning between words such as difference, dissimilarity, distance, and divergence are so subtle that we end up considering them practically synonyms in every day language. In a statistical context, the objective is to measure the level of separation between two elements of a sampling. Depending on the properties that make up the measure, it can be qualified in one sense or another [Gow85, BB97, Mar01].

Let X be a nonempty set and $d : X^2 \to \mathbb{R}$ a function. Then, d is a measure of

Difference If it fulfils

- *Symmetry*: $d(x,y) = d(y,x) \quad \forall x,y \in X$
- *Minimum difference*: $d(x,y) \geq d(x,x) \quad \forall x,y \in X$

Dissimilarity If it is a difference measure which fulfils

- *Non-negativity*: $d(x,y) \geq 0 \quad \forall x,y \in X$
- *Self-similarity*: $d(x,x) = 0 \quad \forall x \in X$

Distance Also called a *metric*, if it is a dissimilarity measure which fulfils ($\forall x,y,z \in X$)

- *Defined*: $d(x,y) = 0 \Rightarrow x = y$
- *Triangle inequality*: $d(x,y) + d(y,z) \geq d(x,z)$

When the objective of the measure is to reflect the discrepancy or difference between two probability distributions, it is called divergence (without loss of generality, divergences are limited to discrete probability distributions) [Bur83]:

Definition 1 Let $\mathcal{X}$ be a countable observation space with $n > 1$ elements and $\mathcal{P}$ the set of all the possible probability distributions of $\mathcal{X}$. Then, $D : \mathcal{P}^2 \to \mathbb{R}^+ \cup \{0\}$ is a divergence if, for all $(p,q) \in \mathcal{P}^2$, it fulfils:

- $D(\{p_1, \ldots, p_n\}, \{q_1, \ldots, q_n\})$ is a continuous function of its $2 \cdot n$ variables.
- $D(p,q)$ is invariant under the permutations of the pairs (p_i, q_i) for $i \in \{1, \ldots, n\}$.

- $D(\{p_1,\ldots,p_n,0\},\{q_1,\ldots,q_n,0\}) = D(p,q)$.
- $D(p,q) \geq 0$.
- $D(p,q) = 0 \Leftrightarrow p = q$.

Note that if a divergence were symmetric it would be equivalent to a defined dissimilarity which only lacks the triangle inequality to attain the category of metric.

2.2. Divergence Classes

With the previous definition of divergence, it is possible to obtain a large set of divergence measures $D(p,q)$. In general, the problem consists in discerning the suitable measures for every specific case. Convexity is a desirable property. A generalisation of it, Jensen's inequality (see Part II), is widely used in mathematics, information theory, and different engineering areas as a kernel of divergence measures. For example, it has been successfully applied to image registration [HHK03] and DNA segmentation [BOR99].

We now see particular examples of divergences (following Pardo [PV03]). From the perspective of information theory, the importance of the information divergence, or Kullback-Leibler distance, is objectively accepted. In mathematical statistics, the same role is played by the chi-square divergence. Also, in convex analysis, arithmetic and geometric means are used in the arithmetic-geometric divergence. In the probability theory, the Vasershtein-Ornstein divergence plays an important role together with L_α-norm divergence, which is also very useful in statistics and other mathematical areas (see Table 1).

These examples and many others are special cases of a kind of divergence measure which obeys the scheme (in this context, we keep the usual functional notation for the probability distributions).

$$D(p,q) = \sum_{x \in \mathcal{X}} \varphi(p(x), q(x)) \qquad (1)$$

for a given real function $\varphi(u,v)$ of positive variables u,v. This function is assumed to be extended to $[0,\infty) \times [0,\infty)$ taking:

$$\varphi(0,0) = 0 \quad \varphi(0,v) = \lim_{u \to 0^+} \varphi(u,v) \quad \varphi(u,0) = \lim_{v \to 0^+} \varphi(u,v) \qquad (2)$$

for $u,v > 0$ where the limits may be infinite.

This scheme has been introduced in information theory for some classes of functions $\varphi(u,v)$ where $(u,v) \in [0,1]^2$. An important case is the kind of divergences that can be generated with the help of a convex function $f : (0,\infty) \to (-\infty,\infty)$, extended to $[0,\infty] \to (-\infty,\infty]$ by the continuity rules:

$$f(0) = \lim_{x \to 0^+} f(x) \qquad f(\infty) = \lim_{x \to \infty} f(x). \qquad (3)$$

Let $\mathcal{F}$ be the set of these functions f which are twice differentiable with continuous derivatives satisfying $f(1) = 0$ and $f''(1) > 0$. Using $f \in \mathcal{F}$, the following divergences are defined in accordance with (1):

Csiszár divergences $\varphi(u,v) = vf\left(\frac{u}{v}\right)$
Bregman divergences $\varphi(u,v) = f(u) - f(v) - f'(v)(u-v)$
Burbea-Rao divergences $\varphi(u,v) = \frac{f(u)+f(v)}{2} - f\left(\frac{u+v}{2}\right)$

These three kinds are partially overlapping [PV97]. We focus our attention on Csiszár divergences [Csi63, Csi67], also called *f-divergence*s and denoted by $D_f(p,q)$.

3. *f*-Divergences

f-Divergences are based on convex functions and were independently introduced by Csiszár [Csi63], and Ali and Silvey [AS66]. These measures have been applied to different areas, such as medical image registration [Plu01] and classification and retrieval [HMMG01], among others. We introduce, in this section, the definition, properties, and particular instances which we will use in the following sections [LV87, Ö02a]. Accordingly §2.2, we take $\varphi(u,v) = vf\left(\frac{u}{v}\right)$ and

Definition 2 Let $(p,q) \in \mathcal{P}^2$ and $f \in \mathcal{F}$. The f-divergence of the probability distributions p and q is given by

$$D_f(p,q) = \sum_{x \in \mathcal{X}} q(x) f\left(\frac{p(x)}{q(x)}\right). \qquad (4)$$

By extension rules (2), for $p,q > 0$

$$0f\left(\frac{0}{0}\right) = 0 \quad qf\left(\frac{0}{q}\right) = qf(0) \quad 0f\left(\frac{p}{0}\right) = p \lim_{y \to \infty} \frac{f(y)}{y}. \qquad (5)$$

Then, $D_f(p,q)$ is well defined as a divergence measure [Vaj89]. Two important properties are:

- $D_f(p,q)$ is convex on (p,q).
  If (p,q) and (p',q') are two pairs of probability distributions, then

$$\lambda D_f(p,q) + (1-\lambda)D_f(p',q') \geq$$
$$D_f(\lambda p + (1-\lambda)p', \lambda q + (1-\lambda)q'). \qquad (6)$$

- Uniqueness.
  If $f,g \in \mathcal{F}$, then $D_f(p,q) = D_g(p,q) \Leftrightarrow \exists r \in \mathbb{R}. f(u) - g(u) = r(u-1)$.

f-Divergences have been studied in depth. The Research Group in Mathematical Inequalities and Applications (Victoria University, Melbourne, Australia) deserves a special mention since over recent years its members have made many contributions to this area [OV93, Dra00, BCDS02, Ö02a]. *f*-Divergences can be grouped together in terms of their convex functions. Considering the classification of Österreicher [Ö02a], we have the following types: χ^α-divergences, (symmetrised) dichotomy, Matusita's divergences, elementary divergences, Puri-Vincze Divergences, and Divergences of Arimoto-type. Within each type, other families of *f*-divergences can be created. We should mention particularly the subtype of f^α-divergences (dichotomy class) presented by Liese and Vajda [LV87].

divergence	field	definition
information	information theory	$I(p,q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}$
chi-squared	statistics	$\chi^2(p,q) = \sum_{x \in \mathcal{X}} \frac{(p(x)-q(x))^2}{q(x)}$
arithmetic-geometric	convex analysis	$AG(p,q) = \sum_{x \in \mathcal{X}} \ln \frac{(p(x)+q(x))/2}{\sqrt{p(x)q(x)}}$
Vasershtein-Ornstein	probability theory	$VO(p,q) = 1 - \sum_{x \in \mathcal{X}} \min\{p(x), q(x)\}$
L_α-norm	mathematics	$L\alpha(p,q) = \sum_{x \in \mathcal{X}} \mid p(x) - q(x) \mid^{\alpha} \quad \alpha \in \{1,2\}$

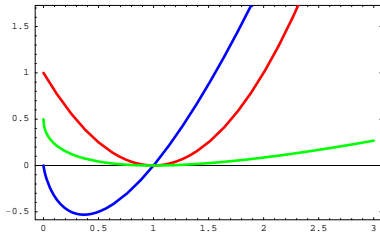**Table 1:** *A subset of useful divergences with its most common fields of application.*

Figure 1: *Plot for $x \in [0,3]$ of three strictly convex functions: $u \log u$ (blue), $(u-1)^2$ (red), and $\frac{1}{2}(\sqrt{u}-1)^2$ (green). The shared intersection is $f(1) = 0$. From these functions, the Kullback-Leibler, chi-square, and the Hellinger f-divergences are obtained, respectively.*

Next, we select three of the most important *f*-divergences [Dra00, GS01, Ö02a], called "distances" in the literature. They are built up from the convex functions in Fig. 1:

Kullback-Leibler $f(u) = u \log u$:

$$D_{\text{KL}}(p,q) = \sum_{x \in \mathcal{X}} p(x) \log \frac{p(x)}{q(x)}. \qquad (7)$$

Introduced by Kullback and Leibler [KL51], it corresponds to the relative entropy or Kullback-Leibler distance (see Part II). Based on continuity arguments, $0 \log \frac{0}{q(x)} = 0$ for all $q(x)$, and $p(x) \log \frac{p(x)}{0} = \infty$ for all $p(x) > 0$ (2). Hence, the measure takes values in $[0, \infty]$. It is not a metric, since it is not symmetric and does not satisfy triangle inequality but, despite of this, it has many useful properties [Rei89, CT91, GS01]. A square root version of Kullback-Leibler divergence has been used by Yang and Barron [YB99]. In Fig. 2.*a* we show the behaviour of this divergence by means of the contribution of a pair $(p(x), q(x))$. The maximum contribution is ∞ in $(p(x), 0)$, and the minimum is $-(e \ln 2)^{-1} \approx -0.531$ in $(\frac{1}{e}, 1)$. The contribution is null for any pair where $p(x) = 0$ or $p(x) = q(x)$. Note the relevance which the divergence takes with respect to p.

Chi-square $f(u) = (u-1)^2$:

$$D_{\chi^2}(p,q) = \sum_{x \in \mathcal{X}} \frac{(p(x)-q(x))^2}{q(x)}. \qquad (8)$$

Defined by Pearson [Pea00] (the history of this measure can be found in Liese and Vajda [LV87]), this measure takes values in $[0, \infty]$ due to the limit when $q(x) = 0$ (2). It is not symmetric. Reiss [Rei89] defined a divergence using the square root of D_{χ^2}. In Fig. 2.*b* we show the same representation as in the previous divergence. The maximum contribution is also ∞ in $(p(x), 0)$, but the minimum is 0 and it is attained in all the pairs where $p(x) = q(x)$. Observe that, even though a strong relevance with respect to p is maintained, the values of q take on more importance than in the case of D_{KL}.

Hellinger $f(u) = \frac{1}{2}(\sqrt{u}-1)^2$:

$$D_{h^2}(p,q) = \frac{1}{2} \sum_{x \in \mathcal{X}} \left(\sqrt{p(x)} - \sqrt{q(x)} \right)^2. \qquad (9)$$

The origins are in Hellinger [Hel09] (historical references can be found in Liese and Vajda [LV87] and Le Cam and Yang [LY90]). This symmetric measure takes values in $[0,1]$ due to the normalisation factor of $\frac{1}{2}$. If it is omitted in f, we obtain the general Hellinger form $2D_{h^2}$. The quantity $1 - D_{h^2}$ is called the Hellinger affinity, a measure popularised by Kakutani [Kak48] who also applied the square root to the general form of Hellinger obtaining a metric [Top00] (the normalisation factor is $\frac{1}{\sqrt{2}}$). In Fig. 2.*c*, the contribution of each pair is shown. The maximum contribution is $\frac{1}{2}$ in $(1,0)$ and $(0,1)$ and the minimum is 0 when $p(x) = q(x)$. Note how the relevance between p and q has balanced out due to the symmetry.

However, none of the above *f*-divergences are true distances. In Österreicher [Ö02b] there is a discussion about which *f*-divergences have a metric behaviour. Gibbs and Su [GS01] provide a summary of bounds between probability metrics and distances. Three relationships between the *f*-divergences presented are (Fig. 3):

- $D_{h^2}(p,q) \leq \frac{1}{2} D_{\text{KL}}(p,q)$ [Rei89]
- $D_{h^2}(p,q) \leq \sqrt{D_{\chi^2}(p,q)}$ [Rei89]
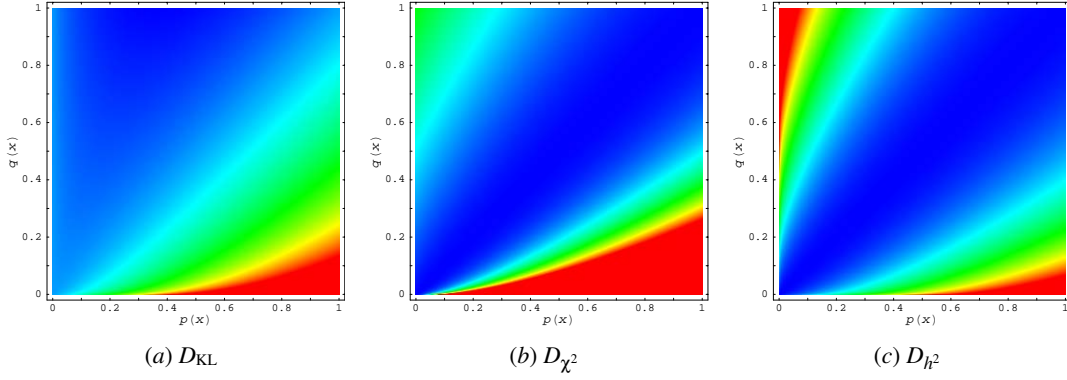- $D_{\text{KL}}(p,q) \leq \log(1 + D_{\chi^2}(p,q))$ [GS01]

(a) D_{KL} (b) D_{χ^2} (c) D_{h^2}

Figure 2: *Density maps of the contribution of a pair $(p(x), q(x))$ for all $x \in \mathcal{X}$: (a) D_{KL}, (b) D_{χ^2}, and (c) D_{h^2}.*
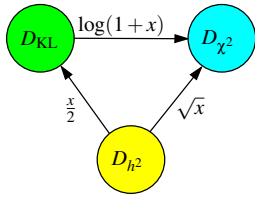

Figure 3: *Bounds between D_{KL}, D_{χ^2}, and D_{h^2}, where $D_f \xrightarrow{h} D_g$ means $D_f(p, q) \leq h(D_f(p, q))$.*
Credit: Adapted from Gibbs and Su [GS01].

4. *f*-Divergences in Radiosity

In this section, new refinement criteria based on *f*-divergences are introduced for hierarchical radiosity (object-space approach).

4.1. Method

Analysing the mutual information-based oracle $\rho_i \delta_{ij} B_j < \varepsilon$ (see Part III) [FRBS02] we observe that it can be rewritten from a Kullback-Leibler distance. In fact, the kernel of the oracle is based on the mutual information, which at the same time is defined as a the Kullback-Leibler distance and which, in accordance with (7), belongs to the *f*-divergences family. In order to obtain the new expression, we need to make the following considerations:

◦ Let $|\mathcal{S}_{i \times j}| = N_s$ be the number of samples of the area sampling form factor computation.
◦ Let $\widehat{F} = \sum_{k=1}^{N_s} F_{x_k \leftrightarrow y_k}$ be the form factor estimation where $(x_k, y_k) \in \mathcal{S}_{i \times j}$ ($\widehat{F} \approx \frac{F_{ji}}{A_i} = \frac{F_{ij}}{A_j}$).
◦ Let $p = \{p_k = \frac{F_{x_k \leftrightarrow y_k}}{\widehat{F}} \mid 1 \leq k \leq N_s\}$ be the probability distribution given for the contribution of every sample to $\widehat{F}$.

Note that $\mathrm{avg}_{1 \leq k \leq N_s} \{p_k\} = \frac{1}{N_s}$. Then, from the discretisation error (see Part III)

$$\delta_{ij} \approx \frac{A_i A_j}{A_T} \left(\frac{1}{|\mathcal{S}_{i \times j}|} \left(\sum_{(x,y) \in \mathcal{S}_{i \times j}} F_{x \leftrightarrow y} \log F_{x \leftrightarrow y} \right) \right.$$
$$\left. - \left(\frac{1}{|\mathcal{S}_{i \times j}|} \sum_{(x,y) \in \mathcal{S}_{i \times j}} F_{x \leftrightarrow y} \right) \log \left(\frac{1}{|\mathcal{S}_{i \times j}|} \sum_{(x,y) \in \mathcal{S}_{i \times j}} F_{x \leftrightarrow y} \right) \right),$$

(10)

we can rewrite

$$\delta_{ij} \approx \frac{A_i A_j}{A_T} (\mathrm{avg}_{1 \leq k \leq N_s} \{F_{x_k \leftrightarrow y_k} \log F_{x_k \leftrightarrow y_k}\}$$
$$- \mathrm{avg}_{1 \leq k \leq N_s} \{F_{x_k \leftrightarrow y_k}\} \log \mathrm{avg}_{1 \leq k \leq N_s} \{F_{x_k \leftrightarrow y_k}\}) \quad (11)$$
$$= \frac{A_i A_j}{A_T} \widehat{F} (\mathrm{avg}_{1 \leq k \leq N_s} \{p_k \log p_k\}$$
$$- \mathrm{avg}_{1 \leq k \leq N_s} \{p_k\} \log \mathrm{avg}_{1 \leq k \leq N_s} \{p_k\}) \quad (12)$$
$$= \frac{A_i A_j}{A_T} \widehat{F} \left(\mathrm{avg}_{1 \leq k \leq N_s} p_k \log p_k - \frac{1}{N_s} \log \frac{1}{N_s} \right)$$
$$= \frac{A_i A_j}{A_T N_s} \widehat{F} \left(\left(\sum_{1 \leq k \leq N_s} p_k \log p_k \right) - \log \frac{1}{N_s} \right)$$
$$= \frac{A_i A_j}{A_T N_s} \widehat{F} D_{\mathrm{KL}}(p, q), \quad (13)$$

where $q = \{q_k = \frac{1}{N_s} \mid 1 \leq k \leq N_s\}$ is the uniform distribution.

This fact suggests that we try other *f*-divergences in the kernel of the refinement oracle. These measures will give us the distance of the distribution of the point-to-point form factors, *p*, with respect to the uniform distribution, *q*. Thus, the Kullback-Leibler (7), chi-square (8), and Hellinger (9) distances have been tested. The Kullback-Leibler-based oracle has already been studied in [FRBS02, Fei02] from an information-theoretic perspective.

Definition 3   Three oracles for hierarchical radiosity, based on their respective f-divergences, are given by

- Kullback-Leibler divergence

$$\rho_i A_i A_j \widehat{F} D_{KL}(p,q) B_j < \varepsilon \qquad \text{(KL)}$$

- Chi-square divergence

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

- Hellinger divergence

$$\rho_i A_i A_j \widehat{F} D_{h^2}^2(p,q) B_j < \varepsilon \qquad \text{(HL)}$$

Observe that the constants $\frac{1}{A_T}$ and $\frac{1}{N_s}$ have been removed since they are specific constants for each scene and are implicit in the threshold.

It is important to note that the expression between parenthesis in (11) corresponds to Jensen's inequality with $f(x) = x \log x$ and $x \in p$. Moreover, we can also see that expression (12) is equal to the first term of the log-sum inequality (see Part II), taking $a_i = p_k$, $b_i = 1$, and $n = N_s$. Thus, $\delta_{ij} \geq 0$.

4.2. Results

For comparative effects, the kernel-smoothness-based oracle,

$$\rho_i \max\{F_{ij}^{max} - F_{ij}^{avg}, F_{ij}^{avg} - F_{ij}^{min}\} A_j B_j < \varepsilon, \qquad \text{(KS)}$$

is chosen as a representative of the oracles which work evaluating the variation of the radiosity kernel between a pair of elements (see Part III). This oracle and the f-divergence-based oracles have been implemented on top of the hierarchical Monte Carlo radiosity method. It should be noted that our oracles can be used with any hierarchical radiosity method.

In Fig. 4 we show a general view of the test scene obtained with the KL oracle. The left image (*i*) shows the Gouraud shaded solution, while the right one (*ii*) corresponds to the subdivision obtained. Each oracle has been evaluated with: $N_s = 10$ random lines between the corresponding pair of elements; an average of 2,684,000 rays to distribute the power have been cast for each solution; and the ε parameter has been tuned so that the meshes obtained have approximately 19,000 patches in all the methods.

For another view of the test scene, we present the results obtained with the f-divergence-based oracles KL, CS, and HL (Figs. 5.*a*–*c*, respectively) and the KS oracle (Fig. 6.*a*). We can see how the f-divergence-based oracles outperform the KS one, working the more complex light zones better and obtaining an improved sharpness in the objects. The meshes created are of higher quality and their precision in the corners and in the transitions of light show this feature. 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 the conditions of Jensen's inequality. This confers a distinct theoretical advantage on this 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 $F_{x_k \leftrightarrow y_k} \log F_{x_k \leftrightarrow y_k}$ by $F_{x_k \leftrightarrow y_k}^2$ in (11), the variance-based oracle is given by

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

where $V(p,q) = \text{avg}_{1 \leq k \leq N_s}\{p_k^2\} - \left(\frac{1}{N_s}\right)^2$. The results obtained are presented in Fig. 6.*b*, showing the inadequacy of this function.

In Table 2, the results of the previous methods are evaluated with the RMSE and PSNR measures. The improvement with regard to the KS and VT is reflected and the results of the KL oracle are noteworthy.

5. f-Divergences in Adaptive Sampling for Ray-Tracing

In this section, we apply the f-divergences to the refinement criteria based on ray-tracing (pixel-driven approach). To do this, we incorporate the divergences into the adaptive sampling scheme using the same basic idea as in hierarchical radiosity (§4) but considering the luminance information instead of the geometric information of the form factors. Therefore, we evaluate the homogeneity of a region of the image plane in accordance with the divergence between its luminance distribution and the uniform distribution. To make the comparison easier, we use the same framework as in the entropy-based refinement criteria for ray-tracing (see Part IV).

5.1. Method

The f-divergences defined in §3 will be used to evaluate the heterogeneity of a set of samples in a region. The scheme used is the following:

1. A first batch of N_s^p samples is cast through a pixel and the corresponding luminances $L_{i \in \{1,...,N_s^p\}}$ are obtained. For an sRGB colour system, the luminance corresponds to the value of Y in the corresponding XYZ system [Com98].
2. The f-divergences $D_f(p,q)$ are taken between the normalised distribution of the obtained luminances,

$$p = \{p_i = \frac{L_i}{\sum_{j=1}^{N_s^p} L_j} \mid 1 \leq i \leq N_s^p\}, \qquad (14)$$

and the uniform distribution $q = \{q_i = \frac{1}{N_s^p} \mid 1 \leq i \leq N_s^p\}$.
3. The refinement criterion, given by

$$\frac{1}{N_s^p} \overline{L} D_f(p,q) < \varepsilon \qquad (15)$$

is evaluated, where D_f represents the Kullback-Leibler,
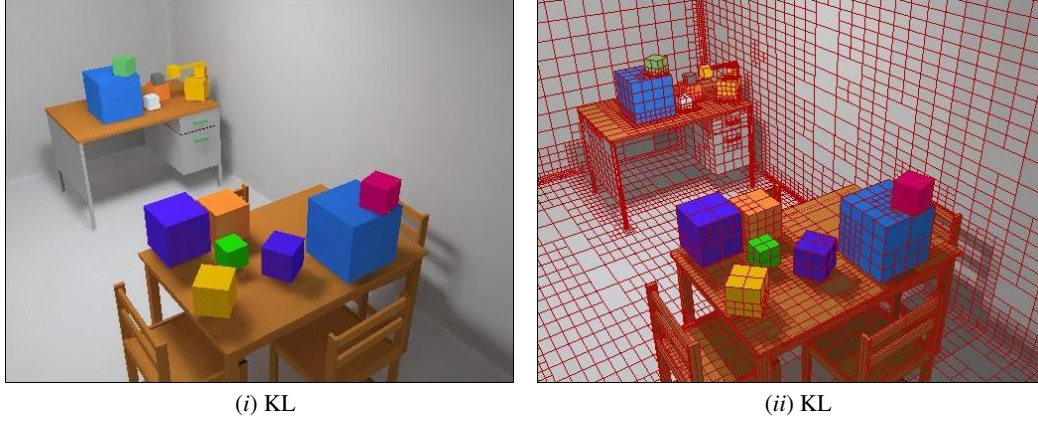
(*i*) KL   (*ii*) KL

Figure 4: *KL oracle for the test scene: (i) Gouraud shaded solution and (ii) its final adaptive mesh. The oracle has been evaluated with 10 random lines between elements.*

oracle	view$_1$				view$_2$			
	RMSE$_a$	RMSE$_p$	PSNR$_a$	PSNR$_p$	RMSE$_a$	RMSE$_p$	PSNR$_a$	PSNR$_p$
KL	9.475	8.698	28.599	29.342	9.712	8.956	28.385	29.088
CS	10.097	9.355	28.047	28.710	10.556	9.825	27.661	28.284
HL	9.990	9.217	28.139	28.839	10.404	9.687	27.787	28.407
KS	13.791	13.128	25.339	25.767	15.167	14.354	24.513	24.991
VT	16.414	15.898	23.826	24.104	17.829	17.378	23.108	23.331

Table 2: *The RMSE and PSNR measures of the f-divergence-based, KS, and VT oracles applied to the test scene. A set of images are shown in Fig. 4 (KL), Fig. 5 (KL, CS, and HL), and Fig. 6 (KS and VT). The oracles have been evaluated with 10 random lines between elements.*

chi-square, or Hellinger divergences, $\overline{L}$ is the average luminance

$$\overline{L} = \frac{1}{N_s^p} \sum_{i=1}^{N_s^p} L_i, \tag{16}$$

and ε is a predefined threshold for the refinement test. The divergence measure $D_f(p,q)$ in the kernel plays the role of a contrast. Note that to assign an importance to this value, we weight it with the average luminance (16), as in Glassner's version of classic contrast [Gla95], used also in the method CC (see Part IV). Division by the number of samples N_s^p in (15) ensures that the refinement process stops.

4. Successive batches of N_s^p rays are cast until the result of the test is true and no more refinement is necessary.

The new criteria give good visual results, but the RMSE obtained in our tests (see Table 3), although better than for the classic contrast, is higher than with the confidence test criterion (see Part IV). Our next logical step was to try the square root of Hellinger divergence, as it is a true metric. The results obtained were very encouraging and, by analogy, we extended the experimentation to the square root of the other divergences (also used in other fields [Rei89, YB99, Top00]). The results also improved the previous ones and

were also better than in the confidence test case. The square root versions of this set of *f*-divergences have already been used previously in statistics. Thus,

Definition 4 Three refinement criteria for adaptive raytracing, based on their respective f-divergences, are given by

- Square root of Kullback-Leibler divergence

$$\frac{1}{N_s^p} \overline{L} D_{KL}^{\frac{1}{2}}(p,q) < \varepsilon \tag{$KL^{\frac{1}{2}}$}$$

- Square root of chi-square divergence

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

- Square root of Hellinger divergence

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

5.2. Results

In Fig. 8 and Fig. 9 we present comparative results with different techniques for the test scene in Fig. 7. The following two methods are compared with the three *f*-divergence-based criteria ($KL^{\frac{1}{2}}$, $CS^{\frac{1}{2}}$, and $HL^{\frac{1}{2}}$ of Def. 4):
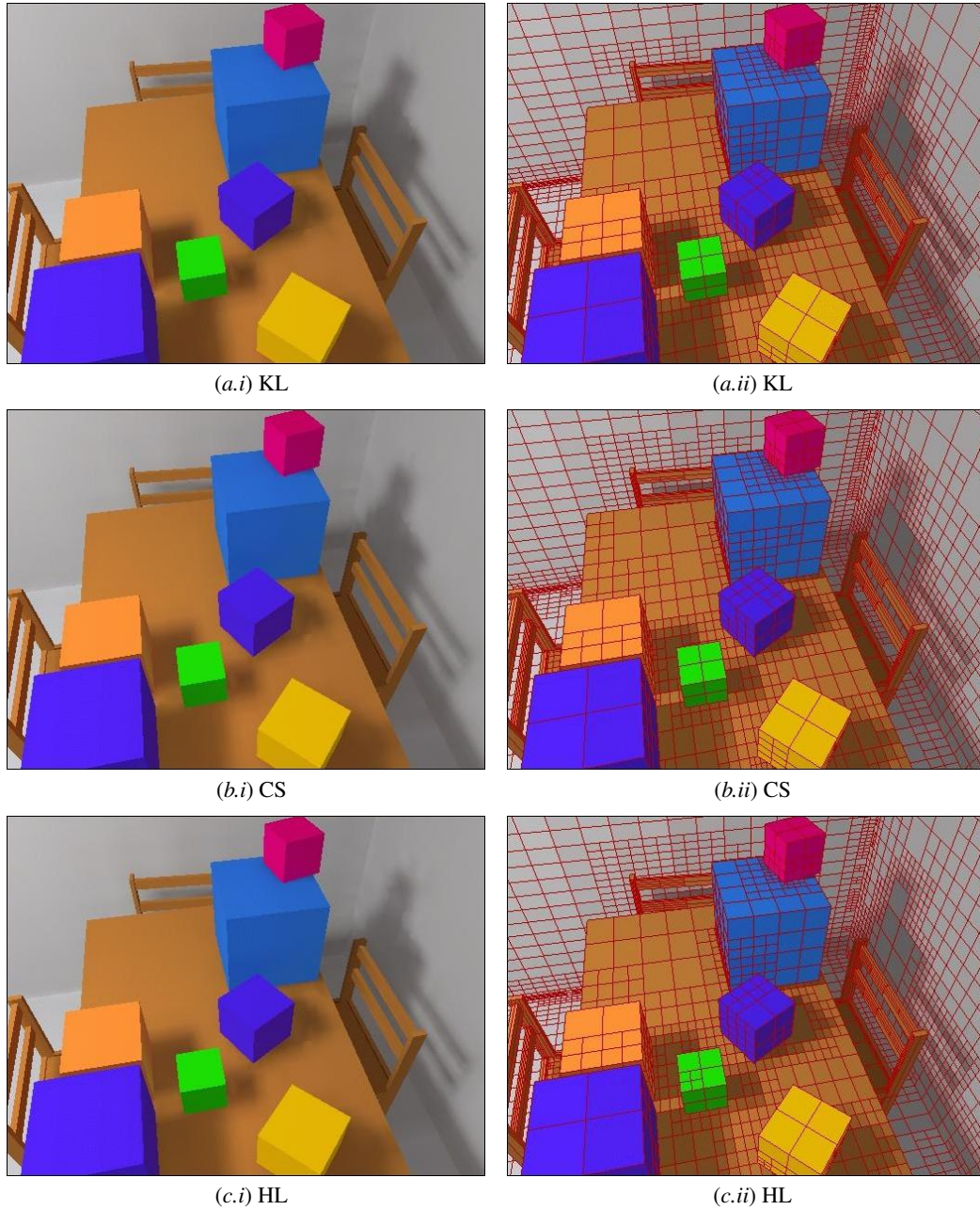
(*a.i*) KL

(*a.ii*) KL

(*b.i*) CS

(*b.ii*) CS

(*c.i*) HL

(*c.ii*) HL

Figure 5: *The view$_2$ of the scene for comparison of* f*-divergence-based oracles, (a) Kullback-Leibler (KL), (b) chi-square (CS), and (c) Hellinger (HL), versus KS and VT ones (Fig. 6). By columns, (i) Gouraud shaded solution and (ii) its final adaptive mesh. The oracles have been evaluated with 10 random lines between elements.*

Figure 6: *The view$_2$ of the scene for comparison of (a) kernel-smoothness-based (KS) and (b) variance-based (VT) oracles versus f-divergence-based ones (Fig. 5). By columns, (i) Gouraud shaded solution and (ii) its final adaptive mesh. The oracles have been evaluated with 10 random lines between elements.*
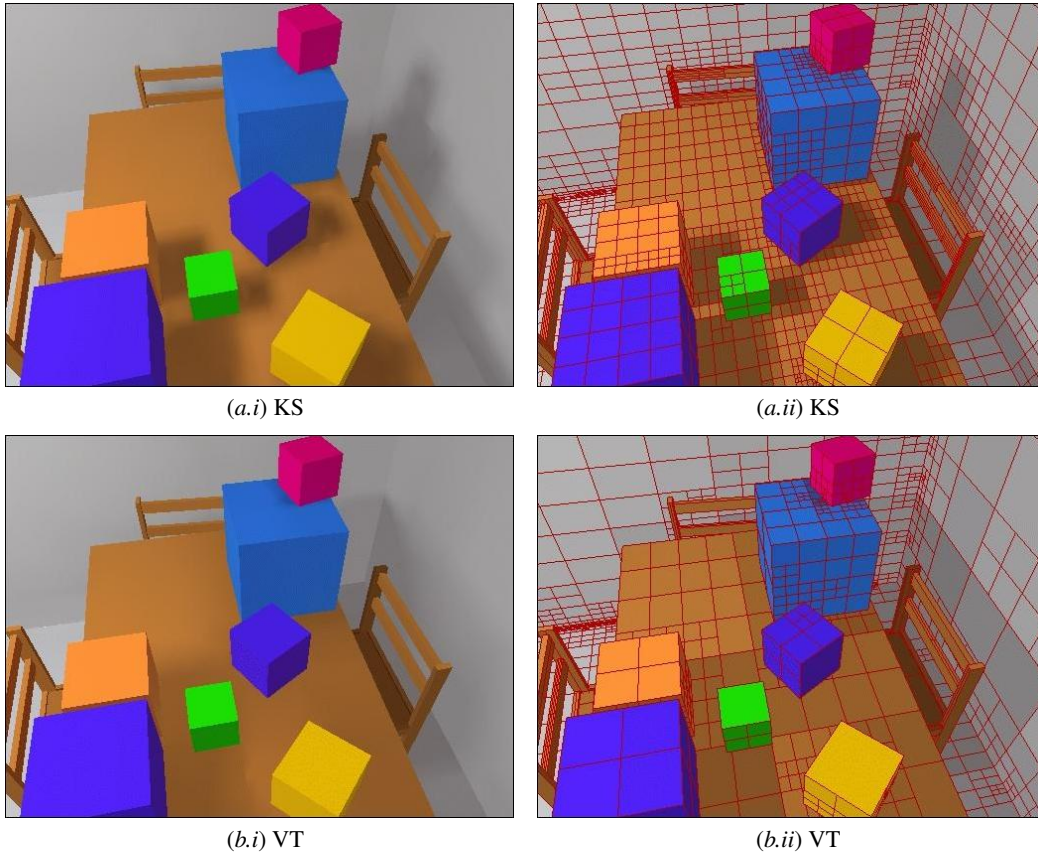
○ CC: Classic contrast (see Part IV) of the luminance weighted with the respective importance $\overline{L}$.

○ CT: Confidence test with a confidence level of $\alpha = 0.1$ and a tolerance $t = 0.025$ (see Part IV).

In order to evaluate their behaviour, the images are generated by a similar process to that of adaptive sampling ray-tracing in Part IV. Clearly, all the methods are directly applicable to adaptive sampling schemes such as that presented in Part IV. In all the methods, 8 initial rays are cast in a stratified way 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. An implementation of classic path-tracing with next event estimator was used to compute all images. The parameters were tuned so that all five test images were obtained with a similar average number of rays per pixel ($N_{\bar{s}}^{p} = 60$) and a similar computational cost. The reconstruction method applied is the piecewise-continuous image with box filter. Finally, the pixel value is the reconstructed signal average at pixel domain.

The resulting images are shown in Figs. 8.∗.*i* (CC and CT) and Figs. 9.∗.*i* (KL$^{\frac{1}{2}}$, CS$^{\frac{1}{2}}$, and HL$^{\frac{1}{2}}$), with the sampling density maps in Figs. 8.∗.*ii* and Figs. 9.∗.*ii*, respectively. The analysis of the critical points of the images shows how our sampling 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.

Comparison of the SDMs shows a better discrimination of complex regions of the scene in the three divergence cases against the classic contrast and confidence test cases. This explains the better results obtained by our approach. On the other hand, the confidence test approach also performs better than the classic contrast-based method. Its SDM also explains why it performs better than the contrast-based. However, it is unable to suitably render the reflected shadows under the mirrored pyramid and sphere with precision.

In Table 3, we show the RMSE and PSNR of the images obtained with classic (Figs. 8.∗.*i*), *f*-divergence, and square root of *f*-divergence (Figs. 9.∗.*i*) methods respective to the
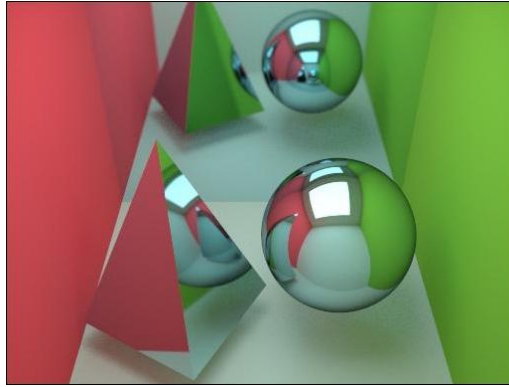
Figure 7: *Test scene for the ray-tracing comparison in Fig. 8 and Fig. 9, obtained with a path-tracing algorithm with 1,024 samples per pixel in a stratified way.*

(*a.i*) CC

(*a.ii*) SDM of (*a.i*)
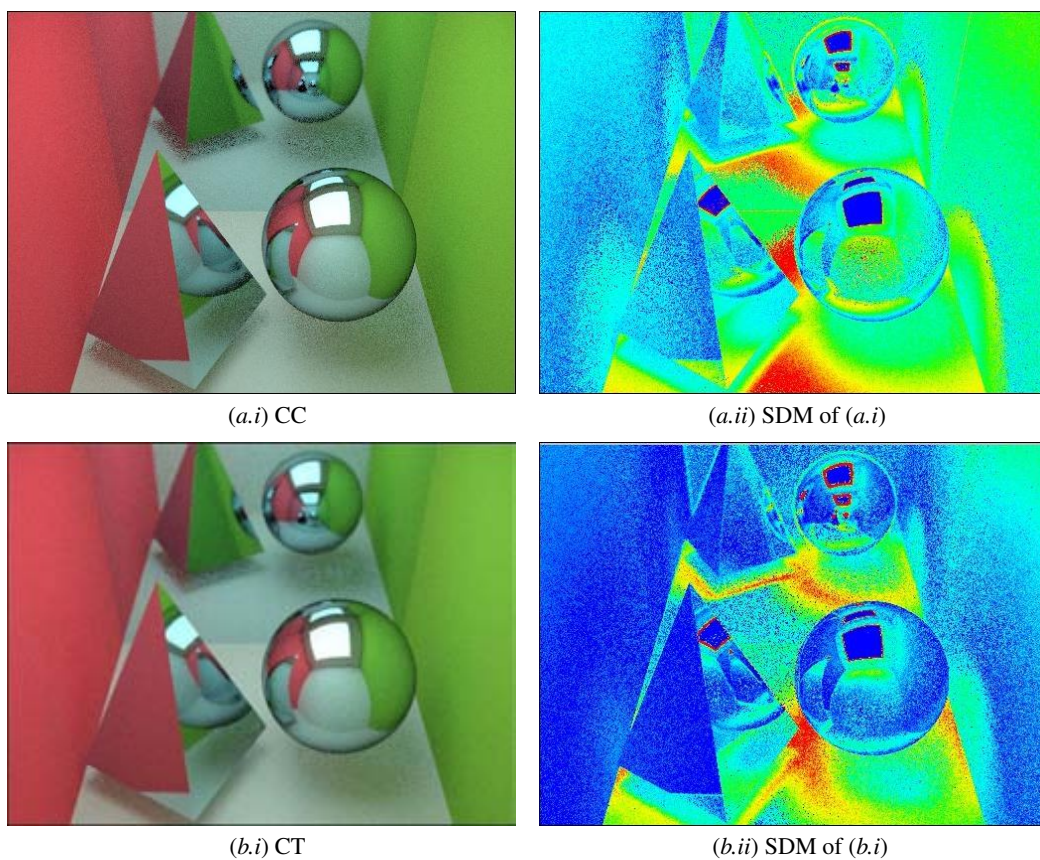

(*b.i*) CT

(*b.ii*) SDM of (*b.i*)

Figure 8: *Images of the test scene (Fig. 7) obtained with an adaptive sampling scheme based on (a) classic contrast (CC) and (b) confidence test (CT) methods. By columns, (i) shows the resulting images and (ii) the sampling density maps of (i). The average number of rays per pixel is 60 in all the methods. Compare with the images in Fig. 9.*

test image in Fig. 7. Visual comparison is in concordance with numerical data. The f-divergence-based criteria used in our experiments ($\mathrm{KL}^{\frac{1}{2}}$, $\mathrm{CS}^{\frac{1}{2}}$, and $\mathrm{HL}^{\frac{1}{2}}$) outperform both classic contrast and confidence test experiments. Finally, the better results of the $\mathrm{HL}^{\frac{1}{2}}$ criterion could be explained by the fact that it is a true distance.

References

[AS66] ALI M. S., SILVEY S. D.: A general class of co-efficient of divergence of one distribution from another. *Journal of Royal Statistical Society (Serie B) 28*, 1 (1966), 131–142.

[BB97] BREN M., BATAGELJ V.: *The Metric Index*. Tech. Rep. 35-561, Institute of mathematics, physics, and mechanics, Department of mathematics, University of Ljubljana, Ljubljana, Slovenia, 1997.

[BCDS02] BARNETT N. S., CERONE P., DRAGOMIR S. S., SOFO A.: Approximating Csiszár f-divergence by the use of Taylor's formula with integral remainder. *Mathematical Inequalities and Applications 5*, 3 (2002), 417–434.

[BOR99] BERNAOLA P., OLIVER J. L., ROMÁN R.: Decomposition of DNA sequence complexity. *Physical Review Letters 83*, 16 (October 1999), 3336–3339.

[Bur83] BURBEA J.: *J*-divergences and related concepts. In *Encyclopedia of Statistical Sciences*, Kotz S., Johnson N. L., Read C. B., (Eds.), vol. 4. Wiley Interscience, New York (NY), USA, 1983, pp. 290–296.

[Com98] COMMISSION I. E.: Default RGB colour space - sRGB. In *Colour Management in Multimedia Systems*. IEC, Geneva, Switzerland, 1998, ch. Part 2.1. IEC/PT61966(PL)34.

[Csi63] CSISZÁR I.: Eine Informationsheoretische Ungleichung und ihre Anwendungen auf den Beweis der Ergodizität von Markoffschen Ketten (An information-theoretic inequality and its application to the proof of the ergodicity of Markov chains). *Magyar Tudományos Akadémia Közleményei 8* (1963), 85–108.

[Csi67] CSISZÁR I.: Information-type measure of difference of probability distributions and indirect observations. *Studia Scientiarum Mathematicarum Hungarica 2* (1967), 299–318.

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[Dra00] DRAGOMIR S. S.: Some inequalities for the Csiszár f-divergence. In *Inequalities for Csiszár f-Divergence in Information Theory*. Research Group in Mathematical Inequalities and Applications, Melbourne, Australia, 2000. http://rgmia.vu.edu.au/monographs/csiszar_list.html.

[Fei02] FEIXAS M.: *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.

[FRBS02] FEIXAS M., RIGAU J., BEKAERT P., SBERT M.: Information-theoretic oracle based on kernel smoothness for hierarchical radiosity. In *Short Presentations (Eurographics '02)* (September 2002), pp. 325–333.

[Gla95] GLASSNER A. S.: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995.

[Gow85] GOWER J. C.: Measures of similarity, dissimilarity, and distance. In *Encyclopedia of Statistical Sciences*, Kotz S., Johnson N. L., Read C. B., (Eds.), vol. 5. Wiley Interscience, New York (NY), USA, 1985, pp. 397–405.

[GS01] GIBBS A. L., SU F. E.: On choosing and bounding probability metrics. *International Statistical Review 70* (2001), 419–435.

[Hel09] HELLINGER E. D.: Neue Begründung der Theorie der Quadratischen Formen von Unendlichen Vielen Veränderlichen (A new foundation of the theory of quadratic forms of infinite many variables). *Journal für Reine und Angewandte Mathematik 136* (1909), 210–271.

[HHK03] HE Y., HAMZA A., KRIM H.: A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing 51*, 5 (May 2003), 1211–1220.

[HMMG01] HERO A. O., MA B., MICHEL O., GORMAN J.: *Alpha-Divergence for Classification, Indexing and Retrieval*. Tech. Rep. CSPL-328, Communications and Signal Processing Laboratory, Ann Arbor (MI), USA, May 2001.

[Kak48] KAKUTANI S.: On equivalence of infinite product measures. *Annals of Mathematics 49* (1948), 214–224.

[KL51] KULLBACK S., LEIBLER R. A.: On information and sufficiency. *Annals of Mathematical Statistics 22* (1951), 76–86.

[LeG99] LEGLAND F.: Stability and approximation of nonlinear filters: an information theoretic approach. In *Proceedings of the 38th IEEE Conference on Decision and Control* (December 1999), pp. 1889–1894.

[LV87] LIESE F., VAJDA I.: *Convex Statistical Distances*. Teubner Verlagsgesellschaft, Leipzig, Germany, 1987.

[LY90] LE CAMP L. M., YANG G. L.: *Asymptotics in Statistics: Some Basic Concepts*. Springer-Verlag, New York (NY), USA, 1990.

[Mar01] MARTÍN J. A.: *Medidas de Diferencia y Clasificación no Paramétrica de Datos Composicionales (Measures of difference and non-parametric clasification of*
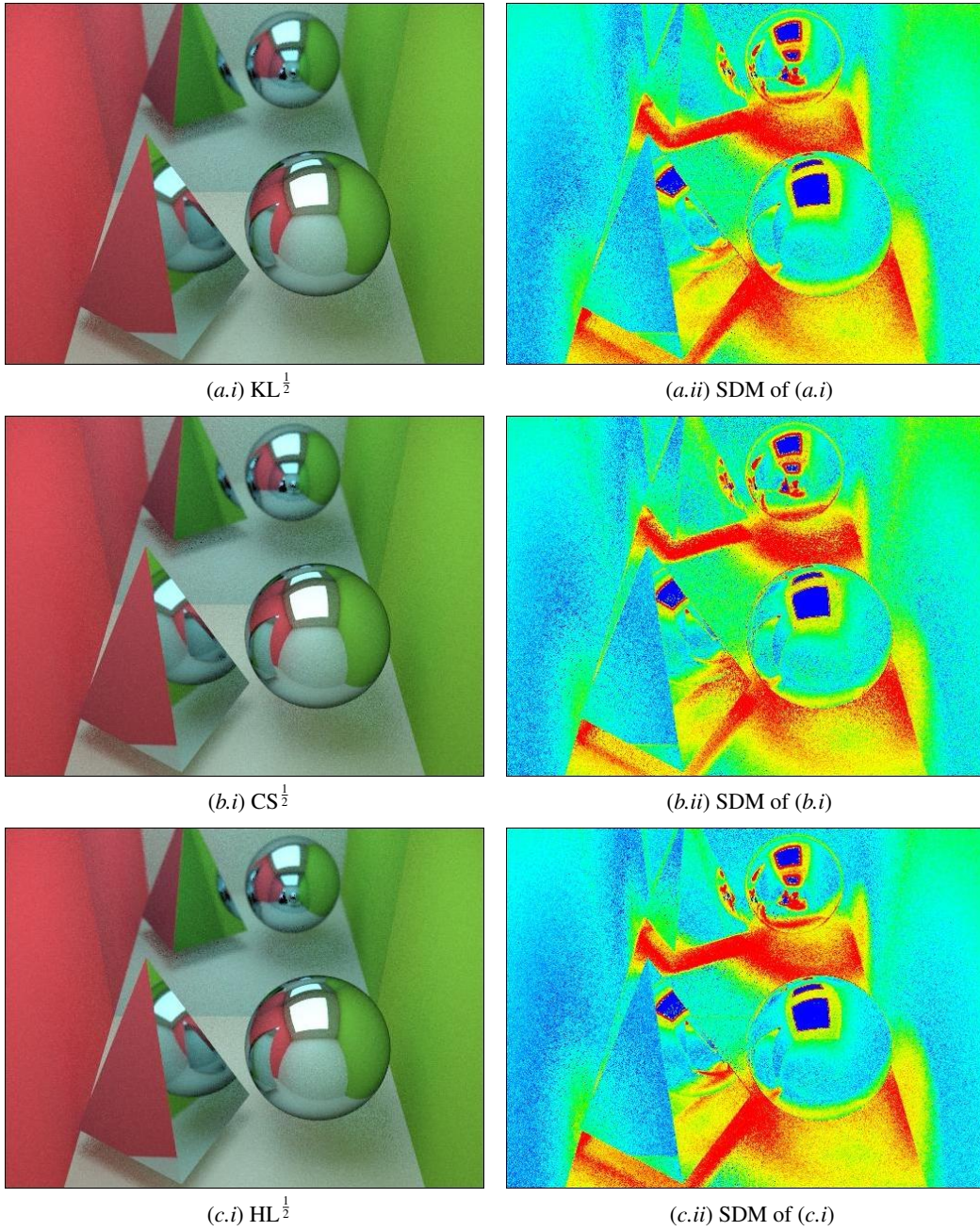
(*a.i*) KL$^{\frac{1}{2}}$

(*a.ii*) SDM of (*a.i*)

(*b.i*) CS$^{\frac{1}{2}}$

(*b.ii*) SDM of (*b.i*)

(*c.i*) HL$^{\frac{1}{2}}$

(*c.ii*) SDM of (*c.i*)

Figure 9: *Images of the test scene (Fig. 7) obtained with an adaptive sampling scheme based on square root of (a) Kullback-Leibler (KL$^{\frac{1}{2}}$), (b) chi-square (CS$^{\frac{1}{2}}$), and (c) Hellinger (HL$^{\frac{1}{2}}$) f-divergences. By columns, (i) shows the resulting images and (ii) the sampling density maps of (i). The average number of samples per pixel is 60 in all the methods. Compare with the images in Fig. 8.*

method	criterion	RMSE$_a$	RMSE$_p$	PSNR$_a$	PSNR$_p$
Classic	Contrast (CC)	6.157	6.126	32.344	32.387
	Confidence test (CT)	5.194	5.174	33.822	33.855
f-divergences	Kullback-Leibler (KL)	5.508	5.553	33.311	33.241
	chi-square (CS)	5.414	5.452	33.461	33.400
	Hellinger (HL)	5.807	5.862	32.852	32.770
Square root of *f*-divergences	Kullback-Leibler (KL$^{\frac{1}{2}}$)	4.824	4.793	34.463	34.519
	chi-square (CS$^{\frac{1}{2}}$)	4.772	4.736	34.557	34.623
	Hellinger (HL$^{\frac{1}{2}}$)	4.595	4.560	34.884	34.951

Table 3: *The RMSE and PSNR measures of the CC, CT, and* f*-divergence-based refinement criteria applied to Fig. 7. The images for the CC and CT methods are shown in Fig. 8, and for the* f*-divergence-based ones, in Fig. 9. The average number of samples per pixel is 60 in all the methods.*

compositional data). PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, 2001.

[Mer05] MERRIAM-WEBSTER, INC.: *Merriam-Webster Online Dictionary.* Springfield (MA), USA, 2005. http://www.m-w.com © 2005 by Merriam-Webster, Inc.

[Ö02a] ÖSTERREICHER F.: Csiszár's *f*-divergences - Basic properties. In *Workshop of the Research Group in Mathematical Inequalities and Applications* (Melbourne, Australia, November 2002).

[Ö02b] ÖSTERREICHER F.: *f*-Divergences - Representation theorem and metrizability. In *Workshop of the Research Group in Mathematical Inequalities and Applications* (Melbourne, Australia, November 2002).

[OBS98] O'SULLIVAN J. A., BLAHUT R. E., SNYDER D. L.: Information-theoretic image formation. *IEEE Transactions on Information Theory 44*, 6 (October 1998), 2094–2123.

[OV93] ÖSTERREICHER F., VAJDA I.: Statistical information and discrimination. *IEEE Transactions on Information Theory 39*, 3 (May 1993), 1036–1039.

[Pea00] PEARSON K.: 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. *Philosophical Magazine V*, 1 (1900), 157–175.

[Plu01] PLUIM J. P. W.: *Mutual Information Based Registration of Medical Images.* PhD thesis, Utrecht University, Utrecht, Holland, 2001.

[PV97] PARDO M., VAJDA I.: About distances of discrete distributions satisfying the data processing theorem of information theory. *IEEE Transactions on Information Theory 43*, 7 (1997), 1288–1293.

[PV03] PARDO M., VAJDA I.: On asymptotic properties of information-theoretic divergences. *IEEE Transactions on Information Theory 49*, 7 (2003), 1860–1868.

[Rei89] REISS R.-D.: *Approximate Distributions of Order Statistics: With Applications to Nonparametric Statistics.* Springer-Verlag, New York (NY), USA, 1989.

[Top00] TOPSØE F.: Some inequalities for information divergence and related measures of discrimination. *IEEE Transactions on Information Theory 46*, 4 (2000), 1602–1609.

[Vaj89] VAJDA I.: *Theory of Statistical Inference and Information.* Kluwer Academic Publishers, Boston (MA), USA, 1989.

[YB99] YANG Y., BARRON A.: Information theoretic determination of minimax rates of convergence. *Annals of Statistics 27* (1999), 1546–1599.

Applications of Information Theory to Computer Graphics Part III: Refinement Criteria for Radiosity

Miquel Feixas, Jaume Rigau, and Mateu Sbert

University of Girona, Spain

1. Radiosity Method

The radiosity method, first introduced in [GTGB84, NN85, CG85], solves the problem of illumination in an environment of diffuse surfaces. In this section, we look at the radiosity equation, the form factor computation and some refinement criteria for hierarchical radiosity.

1.1. Rendering Equation

The light transport in a virtual closed environment or scene (Figure 1) is described by the *rendering equation* [Kaj86], which is a second-order Fredholm integral equation. This equation, which describes all energy exchanges between surfaces, gives us the distribution of light at every point of a scene:

$$L(x, \omega_x) =$$

$$L_e(x, \omega_x) + \int_S \rho_{bd}(x, \omega_x, -\omega_{y \to x}) L(y, \omega_y) G(x, y) dA_y \quad (1)$$

where

- S is the set of surfaces that form the environment
- x and y are points on S
- dA_y is a differential area at point y
- ω_x is a given outgoing direction at point x and $\omega_{y \to x}$ is the outgoing direction at point y towards point x ($\omega_{y \to x}$ can also be seen as an incoming direction at point x coming from point y) (Figure 2)
- $L(x, \omega_x)$ is the *radiance* at point x in direction ω_x (radiance can be defined as the power arriving at or leaving from a surface per unit solid angle and per unit projected area, $\frac{W}{sr.m^2}$) and $L(y, \omega_{y \to x})$ is the radiance at point y in direction $\omega_{y \to x}$
- $L_e(x, \omega_x)$ is the *emitted radiance* at point x in direction ω_x
- $\rho_{bd}(x, \omega_x, -\omega_{y \to x})$, with units sr^{-1}, is the *bidirectional reflectance distribution function* (BRDF) at point x, which is the ratio between the outgoing radiance at x in direction ω_x and the incident radiant flux density (irradiance, $\frac{W}{m^2}$) at x from direction $\omega_{y \to x}$ (Figure 3a)
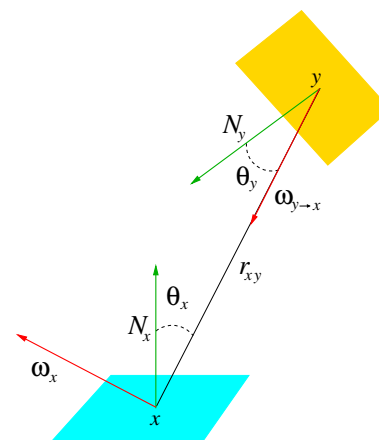
Figure 2: *Outgoing and incoming directions at point x.*

- $G(x, y)$ is the *geometric kernel*, equal to $\frac{\cos \theta_x \cos \theta_y}{r_{xy}^2} V(x, y)$, where θ_x and θ_y are the angles that the line joining x and y form with the normals at x and y respectively, r_{xy} is the distance between x and y, and $V(x, y)$ is a visibility function which is equal to 1 if x and y are mutually visible and 0 if not

This equation can be presented in slightly different forms (global illumination equation, radiance equation) [SP94, Gla95]. Observe that "the radiance distribution L is described implicitly, so we know what conditions it must satisfy, but we don't know what it actually is" [Gla95].

1.2. Continuous Radiosity Equation

For diffuse surfaces, the BRDF does not depend on the outgoing and incoming directions. Thus, the outgoing radiance $L(x, w_x)$ and the self-emitted radiance $L_e(x, w_x)$ are also independent of the outgoing direction (Figure 3b). From this simplification, the *rendering equation for diffuse surfaces*

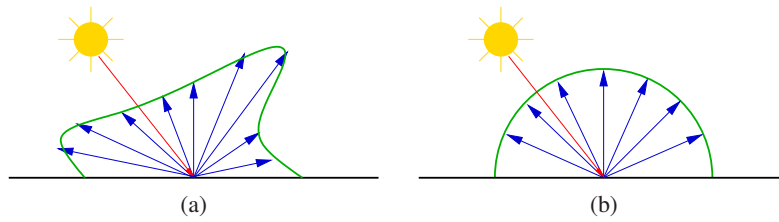**Figure 1:** *Two different illuminated scenes.*

Figure 3: *(a) Bidirectional reflectance distribution function. (b) Diffuse reflectance.*

can be expressed as

$$L(x) = L_e(x) + \int_S \rho_{bd}(x) L(y) G(x,y) dA_y \qquad (2)$$

If we integrate $L(x)$ on the whole hemisphere Ω_x of the outgoing directions w_x at point x, we obtain the total outgoing flux over the hemisphere per unit area, called the *radiosity* at point x (power per unit area) [SP94, Gla95]:

$$B(x) = \int_{\Omega_x} L(x) \cos\theta_x d\omega_x = \pi L(x) \qquad (3)$$

where $d\omega_x$ is the differential solid angle containing the direction ω_x and θ_x is the angle that the direction ω_x forms with the normal at x. In addition, the total self-emitted flux per unit area is expressed by $E(x) = \pi L_e(x)$, and is called the *emittance* at point x.

Note that $\rho_{bd}(x)$ is the ratio of outgoing radiance to incoming flux density. A more convenient quantity is the ratio of reflected to incoming total flux, which must be between 0 and 1 according to the energy conservation law (the energy reflected must be a fraction of the energy received, the other fraction is absorbed). This ratio is the diffuse reflectance, or simply *reflectance*, and is given by $\rho(x) = \pi\rho_{bd}(x)$ (Figure 3b).

The *radiosity equation* is then obtained by multiplying both sides of equation (2) by π:

$$B(x) = E(x) + \frac{\rho(x)}{\pi} \int_S B(y) G(x,y) dA_y \qquad (4)$$

where

- $B(x)$ and $B(y)$ are, respectively, the radiosities at points x and y ($\frac{W}{m^2}$)
- $E(x)$ is the emittance or emitted flux of energy per unit area at point x ($\frac{W}{m^2}$)
- $\rho(x)$ is the diffuse reflectance at point x (dimensionless)

The radiosity equation can also be written in a directional form [SP94]:

$$B(x) = E(x) + \frac{\rho(x)}{\pi} \int_{\Omega_x} B(y) \cos\theta_x d\omega_x \qquad (5)$$

In this conversion, $d\omega_x = \frac{\cos\theta_y}{r_{xy}^2} dA_y$ has been used.

1.3. Discrete Radiosity Equation and Form Factors

To solve the radiosity equation we can use a finite element approach, discretising the environment into n_p patches and considering the radiosities, emissivities and reflectances constant over the patches (Figure 4).

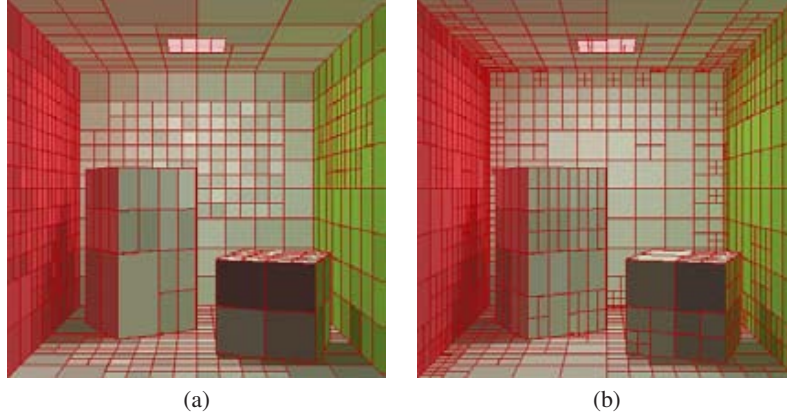With these assumptions, the integral equation (4) becomes

Figure 4: *A scene with two different discretisations.*

the system of radiosity equations [GTGB84]:

$$B_i = E_i + \rho_i \sum_{j=1}^{n_p} F_{ij} B_j \qquad (6)$$

where

- B_i, E_i, and ρ_i are respectively the radiosity, emittance (or emissivity), and reflectance of patch i, and B_j is the radiosity of patch j
- F_{ij} is the *patch-to-patch form factor*, only dependent on the geometry of the scene:

$$
\begin{aligned}
F_{ij} &= \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{G(x,y)}{\pi} dA_y dA_x \\
&= \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y) dA_y dA_x \\
&= \frac{1}{A_i} \int_{A_i} \int_{A_j} F(x,y) dA_y dA_x \\
&= \frac{1}{A_i} \int_{A_i} \int_{\Omega_{x \to j}} \frac{\cos\theta_x}{\pi} V(x,y) d\omega_x dA_x \qquad (7)
\end{aligned}
$$

where A_i and A_j represent, respectively, the surfaces and also the areas of patches i and j, x and y are, respectively, points on A_i and A_j, $F(x,y) = \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y)$ is the *point-to-point form factor*, and $\Omega_{x \to j}$ represents the set of directions going from x to patch j

Form factor properties

Form factors have the following properties:

- Reciprocity

$$A_i F_{ij} = A_j F_{ji} \quad \forall i, j \qquad (8)$$

- Energy conservation

$$\sum_{j=1}^{n_p} F_{ij} = 1 \quad \forall i \qquad (9)$$

- Additivity

$$F_{i(k \cup l)} = F_{ik} + F_{il} \qquad (10)$$

where i, k, and l are three disjoint patches. In general the reverse is not true

$$F_{(k \cup l)i} \neq F_{ki} + F_{li} \qquad (11)$$

In fact, if the patch i is divided into n_i subpatches, we obtain

$$\sum_{k=1}^{n_i} A_{i_k} F_{i_k j} = A_i F_{ij} \qquad (12)$$

As a direct consequence of this equation, if patch i is divided into n_i subpatches of equal area, we have

$$n_i F_{ij} = \sum_{k=1}^{n_i} F_{i_k j} \qquad (13)$$

or

$$F_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} F_{i_k j} \qquad (14)$$

In this case, F_{ij} is the average of the form factors between the subpatches of i and patch j.

Differential-area-to-area form factor

The form factor integral (7) can be considered as an average over the area of patch i of the inner integral. Thus, we have

$$F_{ij} = \frac{1}{A_i} \int_{A_i} F_{dA_x, A_j} dA_x \qquad (15)$$

where F_{dA_x, A_j} is the *differential-area-to-area form factor* and is equal to

$$
\begin{aligned}
F_{dA_x, A_j} &= \int_{A_j} \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y) dA_y \\
&= \int_{\Omega_{x \to j}} \frac{\cos\theta_x}{\pi} V(x,y) d\omega_x \qquad (16)
\end{aligned}
$$

If patches i and j are very distant from each other, then we can assume that F_{dA_x,A_j} is constant over patch i. So, we can evaluate this integral only at a point c (usually the center) of patch i, obtaining

$$F_{ij} = \frac{1}{A_i} \int_{x \in A_i} F_{dA_x,A_j} dA_x$$

$$\approx F_{dA_c,A_j} \frac{1}{A_i} \int_{x \in A_i} dA_x = F_{dA_c,A_j} \qquad (17)$$

The differential-area-to-area form factor is also called *point-to-patch form factor* and can be thought of as the limit of the patch-to-patch form factor when the area of one of the patches decreases to zero [SP94]. We define $F_j(x) \equiv F_{dA_x,A_j}$.

1.4. Form Factor Computation

The form factor computation is the most costly step of the radiosity method. More specifically, its cost is mainly due to the presence of the visibility term in the geometric kernel.

Analytical and deterministic numerical solutions

No analytical closed-form solution exists except for very simple shapes without occlusions. Schroeder and Hanrahan [SH93] solved the polygon-to-polygon case. In [SH92, Gla95] there is an extensive list of formulae for simple shapes. Here we only review one of them: the form factor between two patches of the interior of a sphere without occlusion, which is a paradigmatic case in our work.

From Figure 5, it can be easily obtained that the form factor F_{ij} between two spherical patches i and j is equal to $\frac{A_j}{A_S}$, where A_S is the area of the sphere. As θ_x and θ_y are equal and $\frac{\cos\theta_x}{r_{xy}} = \frac{\cos\theta_y}{r_{xy}} = \frac{1}{2R}$, where R is the radius of the sphere, the expression for the form factor becomes

$$F_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y) dA_x dA_y$$

$$= \frac{1}{\pi A_i} \int_{A_i} \int_{A_j} \frac{\cos\theta_x}{r_{xy}} \frac{\cos\theta_y}{r_{xy}} V(x,y) dA_x dA_y$$

$$= \frac{1}{4\pi R^2 A_i} \int_{A_i} \int_{A_j} dA_x dA_y = \frac{A_j}{A_S} \qquad (18)$$

Note also that $F(x,y) = \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} = \frac{1}{A_S}$ in a spherical scene.

When occlusions between patches exist, we can use deterministic numerical approximations to compute the form factors. Different methods can be found in the literature [SP94, Gla95]: Wallace's method, Nusselt's analogy, hemicube method.

Monte Carlo integration

In this section, the form factor integral (7) will be evaluated by the Monte Carlo method. So, we will give a brief overview of this technique. For a more detailed description,
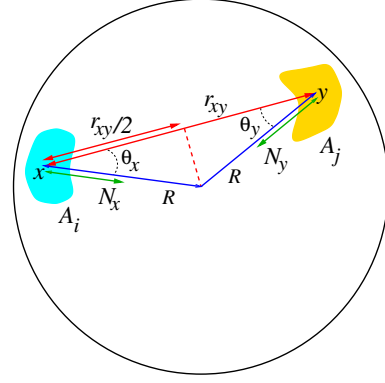

Figure 5: *Geometry for the form factor between two spherical patches.*

see [KW86]. Also [SP94, Gla95, Bek99] review it and give different techniques to sample a random variable.

Monte Carlo integration enables us to estimate integrals using random techniques. Let us suppose we want to solve the integral of a function $g(x)$. This can be written as

$$I = \int_D g(x) dx = \int_D \frac{g(x)}{f(x)} f(x) dx \qquad (19)$$

If $f(x) > 0$ ($\forall x \in D$) and $\int_D f(x) dx = 1$, then $f(x)$ can be considered as a *probability density function* (pdf) of a random variable X and the integral (19) can be read as the expected value of the random variable $\frac{g(X)}{f(X)}$ with respect to the pdf $f(x)$:

$$I = E_f\left[\frac{g(X)}{f(X)}\right] \qquad (20)$$

The term $\frac{g(x_1)}{f(x_1)}$, where x_1 is obtained by sampling from the pdf $f(x)$, is a *primary* estimator for the integral I:

$$I \approx \widehat{I} = \frac{g(x_1)}{f(x_1)} \qquad (21)$$

This estimator is unbiased, i.e., the expected value of this estimator is the value of the integral: $E[\widehat{I}] = I$. The variance of this estimator is given by

$$V[\widehat{I}] = E\left[\left(\frac{g(X)}{f(X)}\right)^2\right] - \left(E\left[\frac{g(X)}{f(X)}\right]\right)^2 = \int_D \frac{g(x)^2}{f(x)} dx - I^2 \qquad (22)$$

Averaging N independent primary estimators (obtained by sampling N independent values $x_1, x_2, \ldots, x_N$ from $f(x)$), we obtain the unbiased *secondary* estimator $\widehat{I}_N$

$$I \approx \widehat{I}_N = \frac{1}{N} \sum_{k=1}^{N} \frac{g(x_k)}{f(x_k)} \qquad (23)$$

with variance

$$V[\widehat{I_N}] = \frac{1}{N}V[\widehat{I}] = \frac{1}{N}\left(\int_D \frac{g(x)^2}{f(x)}dx - I^2\right) \quad (24)$$

So, we obtain better estimators as the number of samples increases. This result is according to the *weak law of large numbers*, which states that, for identically independent distributed (i.i.d.) random variables, $\frac{1}{N}\sum_{k=1}^N X_k$ is close to its expected value $E[X]$ for large numbers of N. Obviously the variance depends on the pdf chosen. When we use a pdf that resembles the integrand we are doing *importance sampling*, which can reduce dramatically the variance of our estimator [KW86].

With respect to the variance, let us remember that the standard deviation of X, which represents the error, is given by $\sigma = \sqrt{V[X]}$. It can be observed from (24) that σ decreases at a rate of $\frac{1}{\sqrt{N}}$ as the number of samples increases.

The mean square error *MSE* of an estimator $\widehat{\theta}$ of θ is given by

$$MSE(\widehat{\theta}) = E\left[\left(\widehat{\theta} - \theta\right)^2\right] \quad (25)$$

and it is equal to the variance when the estimator is unbiased.

Monte Carlo form factor computation

Three different ways of computing the patch-to-patch form factor F_{ij} are here reviewed (for a brief survey, see [Bek99]). Uniform area sampling and uniformly distributed local and global lines can be used to estimate the form factors:

● **Uniform area sampling**
  To calculate the patch-to-patch form factor (7)

  $$F_{ij} = \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2}V(x,y)dA_xdA_y$$

  we take random points x and y on patches i and j respectively (Figure 6). This means taking as pdf $f(x,y) = \frac{1}{A_iA_j}$, which is a uniform distribution.
  A primary estimator $\widehat{F_{ij}^1}$ is given by

  $$\widehat{F_{ij}^1} = \frac{1}{A_i}\frac{F(x,y)}{f(x,y)} = \frac{1}{A_i}\frac{F(x,y)}{\frac{1}{A_iA_j}} = A_jF(x,y) \quad (26)$$

  where (x,y) is a pair of random points. It is easy to see that this estimator is unbiased ($E[\widehat{F_{ij}^1}] = F_{ij}$) and its variance is given by

  $$V[\widehat{F_{ij}^1}] = \int_{A_i}\int_{A_j}(A_jF(x,y))^2\frac{1}{A_iA_j}dA_xdA_y - F_{ij}^2 \quad (27)$$

  We can see that a strong singularity for abutting patches is produced due to the term r_{xy}^4 in the denominator of the integrand. In this case, this technique is not satisfactory, as the variance can be very high [Bek99].

For N samples of pairs (x,y), the form factor integral is approximated by the secondary estimator:

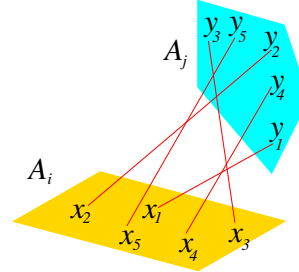$$\widehat{F_{ij}^1} = A_j\frac{1}{N}\sum_{k=1}^N F(x_k,y_k) \quad (28)$$

Figure 6: *Form factor F_{ij} can be computed by taking random points on patches i and j.*

● **Uniformly distributed lines**

  – **Local lines** A local line is a ray with its origin uniformly distributed on the surface of i and its direction distributed according to the cosine with respect to the normal at the origin. So, we estimate the integral (7)

    $$F_{ij} = \frac{1}{A_i}\int_{A_i}\int_{\Omega_{x\to j}}\frac{\cos\theta_x}{\pi}V(x,y)dA_xd\omega_x$$

    taking as pdf $f(x,w_x) = \frac{1}{A_i}\frac{\cos\theta_x}{\pi}$.
    An unbiased primary estimator $\widehat{F_{ij}^2}$ takes the value 1 if the local line hits the patch j directly and 0 if not. Let us recall that if a random variable X takes the values 1 and 0 with probabilities p and $1-p$, its variance is given by $V[X] = p(1-p)$ [Pap84]. Thus,

    $$V[\widehat{F_{ij}^2}] = F_{ij}(1-F_{ij}) \quad (29)$$

    A secondary estimator for F_{ij} is given by

    $$\widehat{F_{ij}^2} = \frac{N_{ij}}{N_i} \quad (30)$$

    where N_i is the number of local lines with origin on i and N_{ij} is the number of local lines with origin on i that hit j. It shows clearly that the form factor F_{ij} can be interpreted as the fraction of local lines with origin on i that have j as the nearest patch intersected (Figure 7a).

  – **Global lines**
    Global lines [Sbe93] can be generated by putting the scene within a sphere and selecting pairs of random points on the surface of this sphere. The lines connecting each pair of points are uniformly distributed throughout the scene. So, the form factor F_{ij} can also be considered as the probability of a global line that, crossing i, hits j (Figure 7b). It can be shown that each
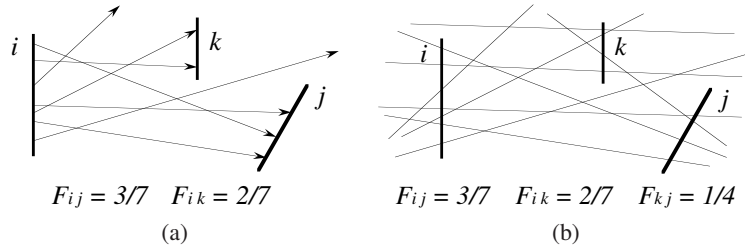
Figure 7: *Form factors can be computed with local (a) and global lines (b).*

line can contribute to the computation of several form factors (Figure 8).

Also, it is important to note that, from integral geometry [San76, Sbe96], the probability that, for a planar patch, a global line intersects patch i is proportional to A_i.

A secondary estimator for F_{ij} is given by

$$\widehat{F_{ij}^3} = \frac{N_{ij}}{N_i} \qquad (31)$$

where N_i is the number of global lines which cross i and N_{ij} is the number of global lines that, crossing i, hit j. Its variance is

$$V[\widehat{F_{ij}^3}] = \frac{1}{N_i} F_{ij}(1 - F_{ij}) \qquad (32)$$

To sample with global lines is equivalent to casting, for each patch, a number of local lines proportional to its area.

Observe that the variance will be higher for smaller patches as N_i is proportional to A_i. If we identify the lines connecting two patches with visibility, the form factor gives us the *visibility* between patches [Sbe96].

1.5. Solution to the Radiosity Equation

The classic radiosity method consists of the following steps [CW93, SP94, Gla95]:

- Discretise the environment into patches
- Compute the form factors F_{ij} for each pair of patches i and j (form factor matrix)
- Solve the system of linear equations
- Display the solution

In this method, the input data are the geometric information about the scene (for the form factors), the physical properties of the materials (for the emissivities and reflectances), and viewing conditions [SP94].

The radiosity equation (6), which refers to a single patch, can be expressed as a system of n_p linear equations. This linear system can be written in the form

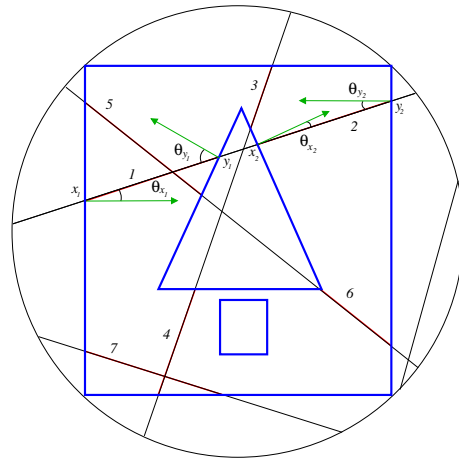$$\mathbf{B} = \mathbf{E} + \mathbf{R}\mathbf{B} \qquad (33)$$

Figure 8: *Each segment of a global line contributes to the computation of two form factors. Thus, the depicted line is used in four form factor computations.*

where $\mathbf{B}$ and $\mathbf{E}$ are, respectively, the vectors of radiosities and emittances, and $\mathbf{R}$ is the $n_p \times n_p$ matrix of the terms $\rho_i F_{ij}$. The solution $\mathbf{B}$ of such a system can be written as a Neumann series. As ρ_i is strictly less than 1, the matrix $\mathbf{R}$ has a norm strictly less than 1. In this case, the Neumann series converges and we can write the radiosity vector as a sum of an infinite series:

$$\mathbf{B} = \mathbf{E} + \mathbf{R}\mathbf{E} + \mathbf{R}^2\mathbf{E} + \cdots + \mathbf{R}^n\mathbf{E} + \cdots \qquad (34)$$

Since $\mathbf{R}$ represents the effect of one reflection on all the surfaces of the scene, $\mathbf{R}^n\mathbf{E}$ can be interpreted as the radiosity obtained after n rebounds of the emitted light through the scene.

In the literature, different iterative solution methods [SP94, Gla95] are available to solve the radiosity and power systems: Jacobi relaxation, Gauss Seidel relaxation, Southwell relaxation, and also their respective stochastic versions [Shi90, SP94, NFP95, Neu95, NNB97, Bek99].

1.6. Random Walks and Markov Chains

A discrete *random walk* [KW86, Rub81] is a Monte Carlo technique used to solve linear systems of equations. A *random walk* in a scene can be considered as a *Markov chain* [Col74, CT91, MR95]. This is a discrete stochastic process defined over a set of states $S = 1, 2, \ldots, n$ which is described by a *transition probability matrix* $\mathbf{P}$. In each step, the imaginary particle (or ray) makes a transition from its current state i to a new state j with *transition probability* P_{ij}. The transition probabilities only depend on the current state. A Markov chain can also be seen as a sequence of random variables $X_k, k = 0 \ldots \infty$, in which each $X_k, k \geq 1$, depends only on the previous X_{k-1} and not on the ones before. The random variables X_k indicate the probability of finding an imaginary particle in each state i after k steps from an initial distribution given by X_0.

Thus, for all $i, j \in S$, we have $\sum_{j=1}^{n} P_{ij} = 1$. Also, if we are in state i, the probability of being in state j after n steps is $(\mathbf{P}^n)_{ij}$ or P_{ij}^n. Under certain conditions, which are met in the context of this dissertation, the probabilities of finding the particle in each state i converge to a *stationary distribution* $\mathbf{w} = \{w_1, \ldots, w_n\}$ after a number of steps. The stationary or equilibrium probabilities w_i fulfil the relation $w_i = \sum_{j=1}^{n} w_j P_{ji}$ and also the reciprocity relation $w_i P_{ij} = w_j P_{ji}$.

A transition probability matrix is said to be aperiodic if it has no periodic state [Col74]. A periodic state is a state that can be visited back by a path starting from it only at multiples of a given period. A probability matrix is said to be irreducible if there is always a path between any two states.

The form factor matrix $\mathbf{F}$ meets all the required conditions to be a valid transition matrix of a random walk. The states of the random walk correspond to the patches of a scene and n_p denotes the number of patches. In order to determine the next state of a random walk, the form factors of the current patch need to be sampled. Such sampling can be carried out easily without having to compute the form factors explicitly [Shi90, PM92, FP93, Sbe97].

For the purpose of our work we are mainly interested in the following two properties [Sbe96]:

1. If the form factor matrix $\mathbf{F}$ is irreducible and aperiodic, then we have

$$\lim_{m \to \infty} (\mathbf{F}^m)_{ij} \to \frac{A_j}{A_T} = a_j \qquad (35)$$

for all the patches of a scene, where A_j is the area of patch j, $A_T = \sum_{i=1}^{n_p} A_i$, and a_j is the relative area of patch j. Thus, as the stationary or equilibrium distribution for a Markov chain is defined as the limit of the mth power of the transition matrix when m grows to infinity, if we know the areas of the patches, we also know the stationary distribution $\mathbf{a} = \{a_i\}$ of the random walk.
2. When the length of a random walk with transition matrix $\mathbf{F}$ grows to infinity, the number of hits on any patch i becomes proportional to a_i, independently of where the random walk started its trajectory.

A Markov chain with a stationary distribution is called ergodic. Thus, the form factors correspond to an ergodic Markov chain.

When the states form a countable set, as stated before, the Markov chain is called a *discrete* chain. When the states are not countable, the chain is called *continuous*. For instance, when taking infinitesimal areas dx at each point x on the surfaces S of the scene as the states and differential form factors $F(x, y)$, with $x, y \in S$ as transition probabilities, a continuous Markov chain with stationary distribution $w(x) = \frac{1}{A_T}$ results.

It can be shown that in flatland the stationary probabilities of the resulting discrete Markov chain are given by $w_i = \frac{L_i}{L_T} = l_i$, where L_T is the total length of all segments of the scene and L_i is the relative length of segment i. When taking infinitesimal lengths dx at each point x on the set of segments L of the scene as the states and differential form factors $F(x, y)$, with $x, y \in L$ as transition probabilities, a continuous Markov chain with stationary distribution $w(x) = \frac{1}{L_T}$ results.

1.7. Hierarchical Radiosity

The classic radiosity method means the entire matrix of form factors has to be computed before a solution can be obtained. This is the most costly step in all the process. To manage the complexity of this problem, different strategies can be used to reduce the number of form factors that need to be computed. Obviously, we also have to take into account the accuracy of the solution. A good strategy has to balance the reduction of the number of patches and the precision of the illumination.

Some techniques have been introduced in order to reduce the computational cost: progressive refinement, substructuring, adaptive refinement and hierarchical refinement. Other techniques try to reduce the number of form factors arriving at a solution within a given error bound [SP94].

The hierarchical refinement algorithm was first introduced in [HSA91] by Hanrahan and Salzman. Additional information can be found in [SP94, Gla95]. This algorithm is based on the objective of reducing the number of form factors needed to propagate the light through the environment: "At the first this may hardly seem possible; after all, the form factors describe the interaction of light between pairs of surfaces. How can we delete any of them and still hope to get an accurate solution?" [Gla95]. Hanrahan and Salzman observed that the N-body problem and the form factor problem were very similar. It is worth noting that both problems are based on the interaction between all pairs of objects and also that the gravitational force and the form factor have a similar mathematical expression. The idea behind both problems is that "small details don't matter when we are far away from

something" [Gla95]. Thus, the clustering algorithms of the *N*-body problem were applied to radiosity, resulting in the *hierarchical radiosity algorithm*.

If each of the *N* particles exerts interactions on the other $N - 1$ particles, thus there exist $\mathcal{O}(N^2)$ interactions to account for. But two distant groups of particles can be considered, in terms of interaction, as two single particles. In hierarchical radiosity the particles are substituted by patches and these are subdivided into smaller elements if necessary, in order to achieve an accurate light transport between them. The main objective is to obtain an accurate piecewise constant approximation of the radiosity on all the elements. To do this, the mesh is generated adaptively: when a constant radiosity assumption on patch *i* is not valid for the radiosity due to another patch, the refinement algorithm will refine *i* in a set of subpatches or elements. Finally, a multiresolution element mesh will enable us to accurately represent the energy transport between patches [Gla95].

An oracle or refinement criterion, based on an error estimation, informs us if a subdivision of the surfaces is needed. The oracle takes geometrical and visibility information about the patches and also the source radiosity and receiver reflectance, and returns whether or not the interaction is valid. Some of them will need further refinement, until a certain level where no further refinement is needed or a previously imposed bound on the area of the patches is reached. Its cost should not make the method prohibitive.

Bekaert et al. [BNN*98] have incorporated hierarchical refinement in Monte Carlo radiosity (more specifically in stochastic Jacobi radiosity) by means of *per-ray* refinement.

1.8. Refinement Criteria

In this section, we review some refinement criteria for hierarchical radiosity [†]. The cheapest and most widely-used oracle has been the power-based oracle [HSA91]. However, it leads to unnecessary subdivisions in smoothly illuminated unoccluded regions receiving a lot of power. As an alternative, oracles based on the smoothness of the geometrical kernel and the received radiosity have been proposed [SAS92, GSCH93, LTG93, PB95, LSG94, BW96, SSS97, HS98]. Nevertheless, oracles based on kernel smoothness also have the problem of unnecessary subdivisions where the kernel is unbounded, and the ones based on received radiosity rely on a costly accurate computation of form factors. All in all, the additional cost invested in both smoothness-based oracles, mainly through visibility computations, may outweigh the improvements obtained [Bek99].

The application of a good refinement criterion and strategy is fundamental for the efficiency of the hierarchical refinement algorithm. Next we review some oracles proposed in the past.

[†] This section follows closely the discussion in [Bek99]

Oracle based on transported power

Hierarchical refinement radiosity was initially presented for constant radiosity approximations by Hanrahan et al. [HSA91]. A cheap form factor estimate F_{ij} which ignores visibility was used to measure the accuracy of an interaction from an element *j* to an element *i*. If $\max(F_{ij}, F_{ji})$ exceeds a given threshold ε, the larger of the two elements *i* and *j* is subdivided using regular quadtree subdivision. Otherwise, the candidate link is considered admissible.

Hanrahan et al. [HSA91] also observed that the number of form factors can be reduced considerably without affecting image quality by weighting the link error estimates F_{ij} with the source element radiosity B_j and receiver element area A_i. Weighting with receiver reflectance ρ_i also further reduces the number of links without deteriorating image quality. Thus, the power-based criterion to stop refinement can be given by

$$\rho_i A_i F_{ij} B_j < \varepsilon \tag{36}$$

Other strategies [TH93, FH96] can also be used to reduce the number of form factors by taking visibility information about candidate interactions into account. We can see that power-based refinement criterion uses no information about the variation of the received radiosity across the receiver element. This results, for instance, in sub-optimal shadow boundaries and excessively fine refinement in smooth areas. The main advantage of criterion (36) is its very low computational cost while yielding a fair image quality.

Oracle based on kernel smoothness

In order to improve on power-based refinement, the variation of the radiosity kernel between a pair of elements is taken into account. In [SAS92], the refinement criterion is given by

$$\rho_i (F_{ij}^{max} - F_{ij}^{min}) A_j B_j < \varepsilon \tag{37}$$

where $F_{ij}^{max} = max_{x \in A_i, y \in A_j} F(x, y)$ and $F_{ij}^{min} = min_{x \in A_i, y \in A_j} F(x, y)$ are the maximum and minimum radiosity kernel values and are estimated by taking the maximum and minimum value computed between pairs of random points on both elements (A_i and A_j represent the surfaces of the elements), ε is a predefined threshold, B_j is the source radiosity and ρ_i the receiver reflectivity.

A similar approach was used in [GSCH93] in order to drive hierarchical refinement with higher-order approximations. When applied to constant approximations, the refinement criterion is given by

$$\rho_i max(F_{ij}^{max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{min}) A_j B_j < \varepsilon \tag{38}$$

where $F_{ij}^{av} = F_{ij}/A_j$ is the average radiosity kernel value. Kernel variation is a sufficient condition for received radiosity variation, but not a necessary condition [Bek99].

Oracle based on smoothness of received radiosity

Because bounding kernel variation is not a necessary condition for bounding received radiosity variation, we can expect that hierarchical refinement based on kernel smoothness will yield hierarchical meshes with more elements and links than required. Optimal refinement can be expected by directly estimating how well the radiosity $B_j(x)$, received at $x \in A_i$ from A_j, is approximated by a linear combination of the basis functions on A_i, i.e., by estimating the discretisation error directly.

This approach was first proposed by Lischinski et al. [LTG93] for constant approximations. Pattanaik and Bouatouch [PB95] proposed a similar strategy for linear basis functions. Other approaches are given in [LSG94, BW96, SSS97, HS98]. The computation cost of kernel and radiosity smoothness-based oracles has not yet been found to compensate for the gain in mesh quality [Bek99].

2. The Scene as a Discrete Channel

In order to apply information theory tools to a scene, we *model* the scene in two equivalent ways:

- *A random walk*
  A discrete random walk (section 1.6) in a discretised scene is a discrete Markov chain where the transition probabilities are the form factors and the stationary distribution is given by the relative area of patches (Figure 9).
- *An information channel*
  A scene can be interpreted as a discrete information channel where the input and output variables take values over the set of patches (the relative area of patches is the probability distribution of these variables) and the channel transition matrix is the form factor matrix.
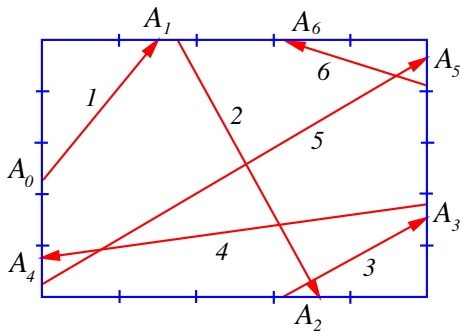

Figure 9: *Discrete random walk in a scene.*

In previous sections (section 1.6 and Part II), we reviewed the most basic concepts about a Markov chain and an information channel. Now, to work with a scene, the following mapping or transformation is done from those general definitions:

- For a discrete Markov chain:
  - number of states: $n \Longrightarrow$ number of patches: n_p
  - transition probability: $P_{ij} \Longrightarrow$ form factor: F_{ij}
  - stationary probability of state i: $w_i \Longrightarrow$ relative area of patch i: $a_i = \frac{A_i}{A_T}$
- For a discrete information channel with random variables X and Y:
  - probability distributions of X and Y: $\mathbf{p}$ and $\mathbf{q} \Longrightarrow$ relative area of patches: $\mathbf{a} = \{a_i\} = \{\frac{A_i}{A_T}\}$
  - conditional probability: $p_{j|i} \Longrightarrow$ form factor: F_{ij}

2.1. Discrete Scene Visibility Entropy

From the above assumptions, we define the *discrete scene visibility entropy rate*, or simply *scene visibility entropy*, as

$$H_S = H(Y|X) = -\sum_{i=1}^{n_p} a_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij} \qquad (39)$$

The scene entropy can be interpreted as the average uncertainty that remains about the destination patch of a random walk (or ray) when the source patch is known. It also expresses the average information content of a random walk in a scene. Moreover, H_S can also be seen as the average of the entropies of all patches, where the *entropy of a patch i* is defined by

$$H_i = H(Y|X=i) = -\sum_{j=1}^{n_p} F_{ij} \log F_{ij} \qquad (40)$$

and thus (39) can be written as

$$H_S = \sum_{i=1}^{n_p} a_i H_i \qquad (41)$$

The entropy of patch i expresses *the uncertainty (or ignorance) of a ray exiting from i about the destination patch*. In fact, it is the Shannon entropy of the form factors of patch *i*.

The Bayes theorem can be now expressed by the *reciprocity* property of the form factors (8)

$$p_{ij} = a_i F_{ij} = a_j F_{ji} \quad \forall i, j \qquad (42)$$

Also, we define the *scene visibility positional entropy* as

$$H_P = H(X) = H(Y) = -\sum_{i=1}^{n_p} a_i \log a_i \qquad (43)$$

which may be interpreted as *the uncertainty on the position (patch) of a ray traveling an infinite random walk*. It is the Shannon entropy of the stationary distribution.

The *joint entropy of a scene* is given by

$$H_J = H(X,Y) = -\sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log a_i F_{ij} \qquad (44)$$

This entropy can be interpreted as *the uncertainty about the pair position-target of a ray in an infinite random walk*. It is

the Shannon entropy of a random variable with probability distribution $\{a_i F_{ij}\}$.

2.2. Discrete Scene Visibility Mutual Information

The *discrete scene visibility mutual information* is defined as

$$
\begin{aligned}
I_S &= I(X,Y) = H(Y) - H(Y|X) \\
&= -\sum_{i=1}^{n_p} a_i \log a_i + \sum_{i=1}^{n_p} a_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij}
\end{aligned}
\tag{45}
$$

and can be interpreted as *the amount of information that the destination patch conveys about the source patch*, and vice versa. *I_S is a measure of the average information transfer in a scene*.

Let us remember that mutual information can be defined as a Kullback-Leibler distance (see Part II): $I(X,Y) = D_{KL}(\{p_{ij}\}\|\{p_i q_j\})$. Thus, scene mutual information is the *distance* or discrimination between the scene probability distribution $\{p_{ij}\} = \{\frac{A_i}{A_T} F_{ij}\}$ and the *independence* distribution of a scene $\{p_i q_j\} = \{a_i a_j\}$. It can also be expressed as

$$
\begin{aligned}
I_S &= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log \frac{a_i F_{ij}}{a_i a_j} \\
&= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log \frac{F_{ij}}{a_j}
\end{aligned}
\tag{46}
$$

2.3. Properties

In a discretised scene the following properties are met:

- From the reciprocity property of the form factors (8), the reversibility of the channel (and also of the Markov chain) can be obtained: $H_S = H(Y|X) = H(X|Y)$.
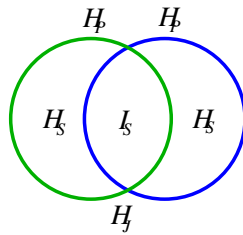- $H_J = H_P + H_S = 2H_P - I_S = 2H_S + I_S$ (Figure 10)

Figure 10: *Venn diagram of a scene.*

- If all the patches have the same area, then $a_i = \frac{1}{n_p}$ and $F_{ij} = F_{ji}$, for all i and j, and therefore

$$
H_S = -\frac{1}{n_p} \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} F_{ij} \log F_{ij} = \frac{1}{n_p} \sum_{i=1}^{n_p} H_i
\tag{47}
$$

and

$$
H_P = \log n_p
\tag{48}
$$

If all the F_{ij} are also equal, then H_S reaches its maximum value: $H_S = H_P = \log n_p$. The minimum value for H_S will be reached when all the form factors from any patch are near zero except one with value near 1: $H_S \approx 0$.

3. Randomness versus Correlation

As we have seen in the previous section, scene visibility entropy H_S is a general measure of the uncertainty or information content associated with a scene: "The entropy density provides an answer to the question: in a Markov chain, given the knowledge of the previous symbol, how uncertain are you, on average, about the next symbol?" [Fel97]. Thus, H_S can be seen as the intrinsic *unpredictability* or the irreducible *randomness* associated with the chain. H_S is also the expected minimum number of bits per symbol required to code a random walk in a scene.

On the other hand, scene mutual information I_S, which expresses the average information transfer, is a measure of the *dependence* or *correlation* between the different parts of a scene. According to W.Li [Li90], "it is now well understood that mutual information measures the general dependence, while the correlation function measures the linear dependence, and mutual information is a better quantity than the correlation function to measure dependence".

3.1. Maximum and Minimum Scene Entropy

It is especially interesting to ask about the extremal cases of maximum and minimum visibility entropy, which correspond to the maximum *disorder* (unpredictability or randomness in the ray path) and the maximum *order* (predictability), respectively. We must remark here that the concepts of order and disorder are not directly referred to the collocation of objects in space, but to visibility criteria. Maximum unpredictability can be obtained in scenes with no privileged visibility directions, and maximum predictability in the contrary case.

Both cases can be illustrated with the following two examples:

- The maximum entropy is exemplified by the interior of an empty sphere divided into equal area patches. Here all the form factors are equal and the uncertainty of the destination patch is maximum:

$$
H_S = H_P = \log n_p
\tag{49}
$$

This means that no visibility direction is privileged and the information transfer is zero: $I_S = 0$.

The sphere represents independence, equilibrium, homogeneity. It is modeled by a channel where the variables X and Y are independent, because in a sphere $F_{ij} = a_j$ and, thus, $p_{ij} = a_i F_{ij} = a_i a_j$. This is the expression of *independence* in a scene. Thus, if the independence is represented by a sphere, the discrete scene mutual information

expresses the *distance* between a given scene and a sphere discretised with the same number and area distribution of the patches.

Note that if the number of patches is doubled, the information content (entropy) of an empty sphere with equal area patches increases by one bit.

- The minimum entropy can be represented by a scene with almost touching objects, as for instance two near concentric spheres with a suitable discretisation. In this case there are strongly privileged visibility directions. This system is highly correlated and the information transfer is large.

3.2. Empirical Results

In this section we show the behaviour of the entropy and mutual information in simple scenes. In the following experiments, form factors have been computed using global lines (see section 1.4).

3.2.1. Different scenes but the same discretisation of the environment

In scenes with the same discretisation (as in Figure 11/Table 1, where we have a cubical enclosure with 512 interior cubes), and consequently with the same H_P, where the interior objects have simply been displaced, we can observe that the increase of entropy remains compensated by a mutual information decrease, and vice versa: more randomness means less correlation (Figure 11a), less randomness means more correlation (Figure 11c). The Venn diagrams in Figure 12 illustrate this behaviour.

Scene	H_S	I_S
Fig.11a	6.761	4.779
Fig.11b	5.271	6.270
Fig.11c	4.849	6.692

Table 1: *Entropy and mutual information for Figures 11a, 11b and 11c. For each scene, $H_P = 11.541$ and 10^7 global lines have been used to compute the form factors.*

3.2.2. The same scene but different discretisations of the environment

How does scene entropy behave with an increase in the number of patches? According to information theory, when the number of patches goes to infinity, the scene entropy also goes to infinity, but scene mutual information tends to a finite value (see the next chapter, section 7). So, in general, the increase in H_S has to be greater than the increase in I_S. This fact is partially illustrated in Table 2, corresponding to Figure 13, where we have a cubical enclosure with three different regular discretisations of its surfaces (600, 2400, and 5400 patches, respectively). We can see that, for each scene,

$H_P = \log n_p$, as all the patches have the same area. The Venn diagrams in Figure 14 illustrate the behaviour of the entropy and mutual information of these scenes.

Scene	H_S	I_S	H_P
Fig.13a	7.838	1.391	9.229
Fig.13b	9.731	1.498	11.229
Fig.13c	10.852	1.547	12.399

Table 2: *Results for a cubical enclosure with different discretisations of its surfaces (Figure 13). 10^9 global lines have been used to compute the form factors.*

3.2.3. Normalized measures

In order to account for changes in the proportion of randomness (H_S) and correlation (I_S) in a scene, these can be normalized by dividing them by the positional entropy H_P. So, they range from 0 to 1.

Normalized scene entropy can be defined as

$$\overline{H_S} = \frac{H_S}{H_P} \tag{50}$$

and *normalized scene mutual information* as

$$\overline{I_S} = \frac{I_S}{H_P} = 1 - \frac{H_S}{H_P} \tag{51}$$

In the literature, normalized mutual information is considered as a measure of *correlation* [CT91]. Also, it can be useful to normalize H_S and I_S with respect to the joint entropy H_J: $\widehat{H_S} = \frac{H_S}{H_J}$ and $\widehat{I_S} = \frac{I_S}{H_J}$. Obviously, they also range from 0 to 1. A similar approach has been used as a measure of 3D medical image alignment [Stu97].

In the sequence of results of Table 3, where we start with an empty cubical enclosure (Figure 13a) and then we add small interior cubes (Figures 15, 16, 11a, and 11c), we can observe how the normalized entropy decreases when we introduce progressively more cubes. This fact increases the correlation in the scene, to the detriment of its randomness, in spite of the fact that H_P also increases.
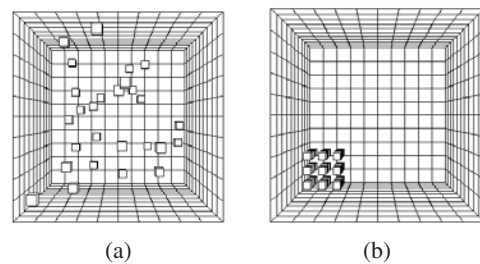

(a)   (b)

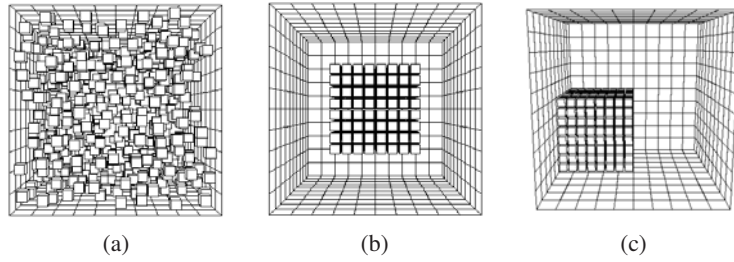**Figure 15:** *(a) Random and (b) clustered configurations with 27 cubes.*

| (a) | (b) | (c) |

Figure 11: *A random configuration and two clustered configurations with* 512 *cubes.*

| (a) | (b) | (c) |
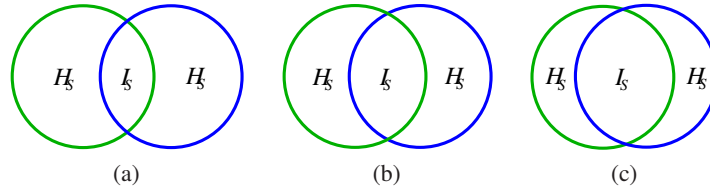
Figure 12: *Venn diagrams corresponding to different scenes with the same discretisation. The size of the circles (H_P) remains the same in all the diagrams.*

| (a) | (b) | (c) |
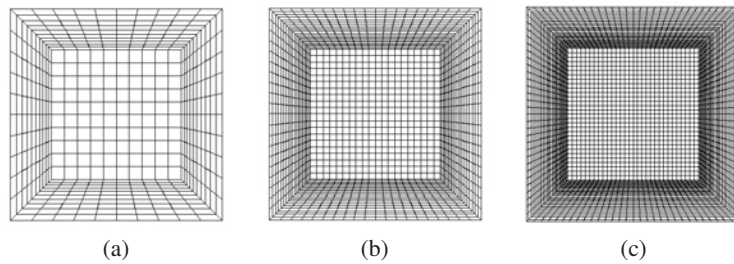
Figure 13: *Three empty cubical enclosures with their surfaces regularly discretised into (a)* 10×10*, (b)* 20×20*, and (c)* 30×30 *patches, respectively. The total number of patches is, respectively,* 600*,* 2400*, and* 5400*.*

| (a) | (b) | (c) |
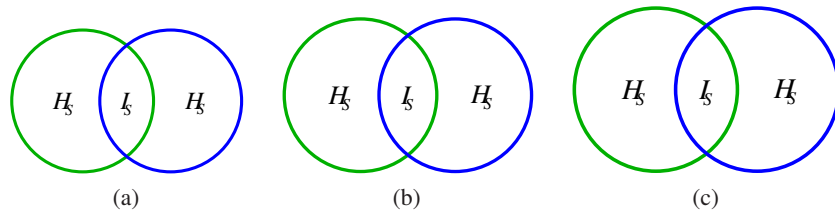
Figure 14: *Venn diagrams corresponding to the scenes of Figure 13 where we have a cubical enclosure with successive refinements. The size of the circles (H_P) increases with the number of patches.*

(a)   (b)

Figure 16: *(a) Random and (b) clustered configurations with 64 cubes.*

Scene	H_S	I_S	$\overline{H_S}$
Fig.13a	7.821	1.408	0.847
Fig.15a	7.780	1.669	0.823
Fig.15b	7.589	1.861	0.803
Fig.16a	7.709	2.009	0.793
Fig.16b	7.606	2.112	0.783
Fig.11a	6.761	4.779	0.586
Fig.11c	4.849	6.692	0.420

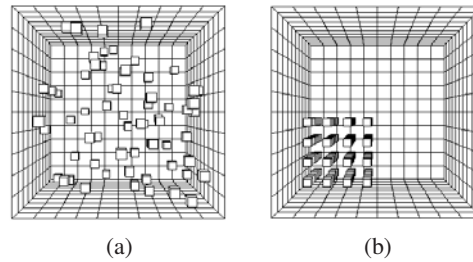**Table 3:** *Results for the empty scene of Figure 13a and the scenes with 27, 64 and 512 small cubes of Figures 13a, 15, 16, 11a, and 11c. For each scene, 10^7 global lines have been cast.*

From the previous results, we see that scene entropy (randomness) tends to increase with the *number of patches* and scene mutual information (correlation) tends to increase with the *number of objects* within a enclosure. So, the increase in the number of patches and the increase in the number of objects work in different (but complementary) directions.

4. Complexity of a Scene

Scene complexity has often been expressed as the number of patches into which a scene is subdivided. But, what do we really expect scene complexity to measure? In our context, scene complexity has to answer the question of how difficult it is to compute the visibility and radiosity of a scene with sufficient accuracy. Studying scene complexity will help to improve our knowledge about the behaviour of the visibility and radiosity of a scene.

To solve the illumination in a diffuse environment, we need to simulate the interreflection of light between all the surfaces. This simulation presents typical characteristics of complex behaviour. The *difficulty in obtaining a precise illumination solution* depends on

- the degree of dependence between all the surfaces

- how the interaction between these surfaces changes in dependence when the system is subdivided
- the degree of unpredictability

The two first considerations can be represented by a *statistical complexity measure*, which quantifies *correlation*, *structure*, or *interdependence* between the parts of a system, and the third one by the *entropy*, which measures *randomness* or *unpredictability*. In this work, the word *complexity* is reserved for a measure of *statistical complexity* and *entropy* is referred to as *randomness*.

The most representative measures of statistical complexity are excess entropy and mutual information. For scene visibility, the following proposition is fulfilled:

Proposition 1 From the point of view of scene visibility, the excess entropy (see Part II) becomes the mutual information.

Proof. From definitions in Part II and section 2, we have

$$H(X_1,\ldots,X_n) = H(X_1) + \ldots + H(X_n|X_1,\ldots,X_{n-1})$$
$$= H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1})$$
$$= H_P + (n-1)H_S$$

Thus, $E = \lim_{n\to\infty}(H_P + (n-1)H_S - nH_S) = H_P - H_S = I_S$. $\square$

So, *we propose taking the mutual information as a measure of scene complexity*.

According to Feldman, Rohilla, and Crutchfield:

- "It has become (in our sense) more broadly understood that a system's randomness and unpredictability fail to capture its patterns and correlational structure." [FC98]
- Entropy and mutual information are orthogonal or complementary: "Complexity and randomness each capture a useful property to describe how a process manipulates information" [SC99].

Entropy and mutual information express two basic aspects of scene complexity. But, we are sure that other measures could capture other perspectives of the scene complexity. Remember that one of the most basic formulae of information theory (see part II) relates entropy to mutual information:

$$H(X) = H(X|Y) + I(X,Y) \qquad (52)$$

Thus, complexity and randomness are combined in the same expression.

A very simple approach would be to consider that the complexity can be represented by the number of patches. It is true that scene entropy is strongly dependent on n_p and increases with it, with a maximum value of $\log n_p$. However, as we will see in this chapter, scene mutual information presents a very different behaviour with respect to n_p.

5. Continuous Scene Visibility Mutual Information

A scene is a continuous system. Thus, by discretising a scene into patches, a *distortion* or *error* is introduced. In a way, to discretise means to make it uniform, and consequently *some information is lost*. Obviously, the maximum accuracy of the discretisation is accomplished when the number of patches tends to infinity. Since the *continuous mutual information expresses with maximum precision the information transfer or correlation in a scene*, it will be considered as the main measure of the *scene complexity*. On the other hand, *discrete mutual information* will represent the *complexity of a discretised scene*.

Similarly to the previous chapter, now the scene is modeled by a continuous random walk (Figure 17) or by a continuous information channel. As we have seen in Part II, the mutual information between two continuous random variables X and Y is the limit of the mutual information between their discretised versions. On the contrary, the entropy of a continuous random variable does not equal the entropy of its discretised version in the limit of a finer discretisation.
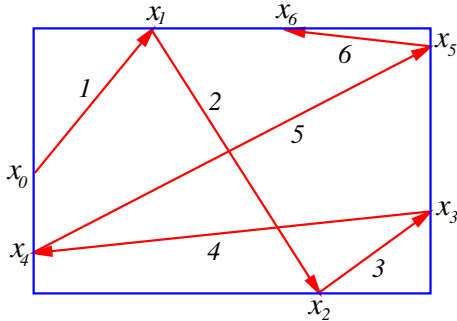

Figure 17: *Continuous random walk in a scene.*

Thus, in a scene, discrete mutual information I_S converges to continuous mutual information I_S^c when the number of patches tends to infinity:

$$I_S^c = \lim_{n_p \to \infty} I_S \qquad (53)$$

In this chapter we will see that this result is very important for this dissertation because it will enable us to calculate the distance to the ideal discretisation, represented by the continuous mutual information.

Scene visibility entropy however tends to infinity when the number of patches tends to infinity:

$$\lim_{n_p \to \infty} H_S = \infty \qquad (54)$$

As we have seen (section 1.6), when the states form an uncountable set, we deal with a continuous Markov chain. We can obtain the continuous formulae for the entropy and mutual information of a scene from the respective discrete definitions using the following substitutions:

- Each state by an infinitesimal area and each summatory by an integral.
- $w_i = \frac{A_i}{A_T} \Longrightarrow \frac{1}{A_T}$. This means substituting the discrete probability of taking patch i by the continuous probability of selecting any point.
- $F_{ij} \Longrightarrow F(x, y)$. This means substituting the patch-to-patch form factor by the point-to-point form factor. Remember that the value of $F(x, y)$ is $\frac{cos\theta_x cos\theta_y}{\pi r_{xy}^2}$ for mutually visible points, or zero otherwise, θ_x and θ_y being the angles which the normals at x and y form with the segment joining x and y, and r_{xy} the distance between x and y (see section 1.3).

In the same way, the continuous formulae for a scene can also be obtained from the continuous formulae of the entropy and mutual information (see Part II) by applying the following substitutions:

- $dx \Longrightarrow dA_x$, $dy \Longrightarrow dA_y$
- $p(x) \Longrightarrow \frac{1}{A_T}$
- $p(y|x) \Longrightarrow F(x, y)$
- $p(x, y) \Longrightarrow \frac{1}{A_T} F(x, y)$

Thus, we obtain

- *Continuous positional entropy*

$$H_P^c = - \int_S \frac{1}{A_T} \log \frac{1}{A_T} dA_x = \log A_T \qquad (55)$$

- *Continuous scene visibility entropy*

$$H_S^c = - \int_S \int_S \frac{1}{A_T} F(x, y) \log F(x, y) dA_x dA_y \qquad (56)$$

- *Continuous scene visibility mutual information*

$$\begin{aligned} I_S^c &= \log A_T + \int_S \int_S \frac{1}{A_T} F(x, y) \log F(x, y) dA_x dA_y \\ &= \int_S \int_S \frac{1}{A_T} F(x, y) \log (A_T F(x, y)) dA_x dA_y \end{aligned} \qquad (57)$$

For example, in the interior of an empty sphere, as any pair (x, y) fulfills $F(x, y) = \frac{1}{A_T}$, the result obtained is, as expected, $I_S^c = 0$. Remember that, in a sphere, $I_S = 0$ and thus $\lim_{n_p \to \infty} I_S = I_S^c = 0$.

Note that H_P^c and H_S^c are not invariant to a change in the scale of a scene. For our objectives, we are only interested in their discrete versions, which are always invariant to a change in the scale. I_S and I_S^c also have this desirable property.

6. Monte Carlo Computation of the Scene Visibility Complexity

Now we will show how the continuous mutual information can be computed using local or global lines.

6.1. Monte Carlo Integration

The continuous mutual information integral can be solved by Monte Carlo integration. Reparametrizing the integral, we have

$$
\begin{aligned}
I_S^c &= \int_S \int_S \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \\
&= \int_S \int_{\Omega_x} \frac{1}{A_T} \frac{\cos\theta_x}{\pi} \log(A_T F(x,y(x,\omega_x))) dA_x d\omega_x
\end{aligned}
$$
(58)

where $y(x,\omega_x)$ is the point visible from x in the ω_x direction. Now we can use $\frac{\cos\theta_x}{\pi A_T}$ as probability density function ($\int_S \int_{\Omega_x} \frac{\cos\theta_x}{\pi A_T} dA_x d\omega_x = 1$). Drawing samples according to this distribution means simply selecting first a random point in the scene upon the area and a direction upon the form factor distribution. This can be achieved with *local* lines or *global* lines. The result obtained is

$$
\begin{aligned}
I_S^c &\approx \frac{1}{N} \sum_{k=1}^N \log(A_T F(x_k, y_k(x_k, \omega_{x_k}))) \\
&= \frac{1}{N} \sum_{k=1}^N \log\left(\frac{A_T \cos\theta_{x_k} \cos\theta_{y_k}}{\pi r_{x_k y_k}^2} \right)
\end{aligned}
$$
(59)

In the global line case, N stands for the total number of segments of the global lines or the number of pairs of points considered, which is the total number of intersections divided by two (see Figure 8). In the local line case, N represents the total number of local lines used in a scene and the quantity of lines cast from each patch, N_i, proportional to its area ($N_i = \frac{A_i}{A_T} N$). In this chapter, the scene complexity has been calculated using both global and local lines.

6.2. Empirical Results

We begin by computing the complexity of platonic solids and the Cornell box (Figure 1a). In Table 4, we can observe that the minimum complexity corresponds to a sphere and the maximum complexity to a tetrahedron. As we expected, the polyhedra that are nearer to the sphere are less complex, i.e., they have less correlation. Thus, complexity appears to be inversely proportional to the number of faces. The complexity of the Cornell box is clearly greater than the one for the empty cube, as we have increased the cube complexity by introducing objects in its interior.

In addition, in Table 5, we show the complexity for the scenes of Figure 18. In Figure 18a, an object, made up of a table and four chairs, is situated in the middle of a room. In Figures 18b and 18c, arrays of 4 or 16 objects have been situated in the middle of the same room. In Figures 18d,

Scene	I_S^c
sphere	0
icosahedron	0.543
dodecahedron	0.825
octahedron	1.258
cube	1.609
tetrahedron	2.626
Cornell box	3.274

Table 4: *Complexity of platonic solids and the Cornell box. For each scene, 10^6 global lines have been cast.*

18e and 18f, the same 16 objects have been distributed in different ways. We can see that the introduction of objects increases the complexity and that the scenes with the same objects (18c, 18d, 18e and 18f) show similar complexities. In this case, the increase in complexity is produced when there are objects near the walls because this fact increases the correlation in the scene.

7. Complexity and Discretisation

Now, we will try to show that the scene complexity I_S^c is closely related to the difficulty in obtaining an accurate discretisation. In a way, to discretise a scene is to model it. "A system is not complex by some abstract criterion but because it is intrinsically hard to model" [BP97]. This point of view is compatible with W.Li's comment that: "An intuitively satisfactory definition of complexity should measure the amount of effort put in that generates *correlations* in a sequence. Of course, one cannot be sure that all the effort is spent on generating correlations. As a result, a measure of correlation typically provides a lower bound of a measure of complexity, and might be a reasonable estimate of the complexity" [Li91].

7.1. Continuous versus Discrete Mutual Information

From Part II, we know that

- The mutual information between two continuous random variables is the limit of the mutual information between their discretised versions.
- Refinement can never decrease the discrete mutual information.
- The continuous mutual information is the least upper bound for the discrete mutual information.

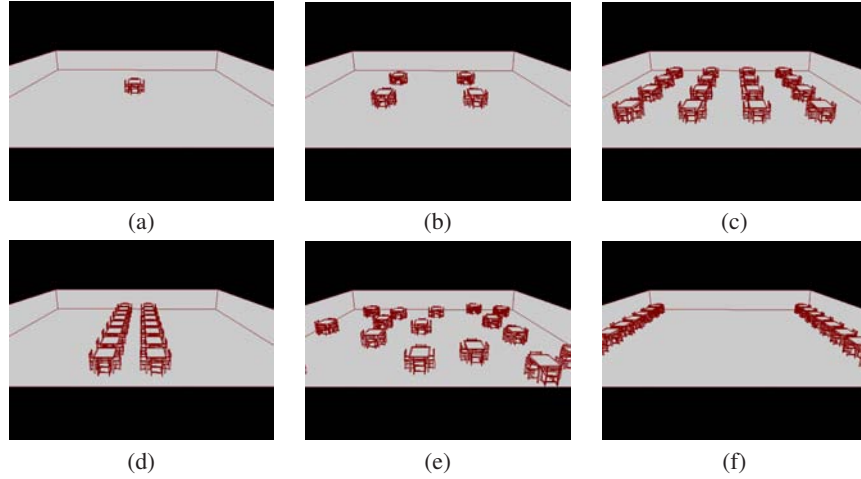Now, if we apply these results to scene visibility, we find that:

Figure 18: *(a) An object, composed of a table and four chairs, and (b) an array of 4 objects with the same composition, have been situated in the middle of a room. (c, d, e, f) The same 16 objects have been distributed in four different ways.*

Scenes	Fig.18a	Fig.18b	Fig.18c	Fig.18d	Fig.18e	Fig.18f
I_S^c	3.837	4.102	5.023	5.043	5.044	5.089

Table 5: *Complexity of the scenes of Figure 18. For each scene, 10^6 global lines have been cast.*

- If any patch is divided into two or more patches, the discrete mutual information I_S of the new scene increases or remains the same.
- The continuous scene visibility mutual information is the least upper bound to the discrete scene visibility mutual information.

Thus, a scene fulfils:

$$I_S^c - I_S \geq 0 \qquad (60)$$

Initially, these results were proved in [FS98, FdABS99]. In the next chapter, we give a general proposition proving these properties for visibility, radiosity and importance.

As we can see in Tables 6 and 7, corresponding to Figures 13 and 11(a), respectively, the computational cost of I_S^c is much lower than the cost of I_S: with few lines I_S^c can be computed with enough precision, unlike I_S which needs a lot of lines to get a precise measurement. Observe that I_S increases with the number of patches but is always less than I_S^c. We can also see that, due to the Monte Carlo error, the value of the discrete mutual information decreases (until convergence is achieved) with the increase in the number of lines cast. With few lines per patch, the values of the form factors give us an erroneous high correlation.

In Tables 8 and 9, corresponding to Figures 19 and 20, we also show how discrete mutual information I_S increases with the mesh refinement.

Scene	Lines (10^6)	H_S	I_S	I_S^c
Fig.13a	0.1	6.482	2.747	1.610
Fig.13a	10	7.821	1.408	1.612
Fig.13a	1000	7.838	1.391	1.610
Fig.13b	0.1	5.342	5.887	1.608
Fig.13b	10	9.420	1.809	1.611
Fig.13b	1000	9.731	1.498	1.610
Fig.13c	0.1	4.313	8.086	1.610
Fig.13c	10	9.684	2.715	1.611
Fig.13c	1000	10.852	1.547	1.610

Table 6: *Results for the cubical enclosure of Figure 13 with different discretisations of its surfaces. For each scene, 10^5, 10^7, and 10^9 global lines have been cast.*

Lines (10^6)	0.01	0.1	1	10
I_S	8.773	6.398	5.171	4.779
I_S^c	5.650	5.636	5.631	5.632

Table 7: *Discrete and continuous mutual information for a scene with 512 cubes (Figure 11a).*
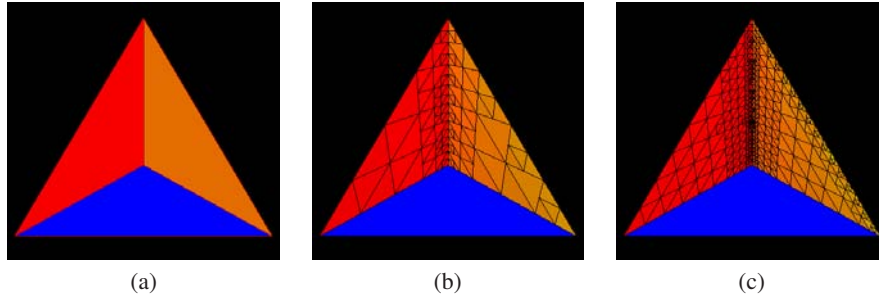
(a)　　　　　　　(b)　　　　　　　(c)

Figure 19: *Three different discretisations for a tetrahedron. The total number of patches is, respectively,* 4, 151, *and* 793.
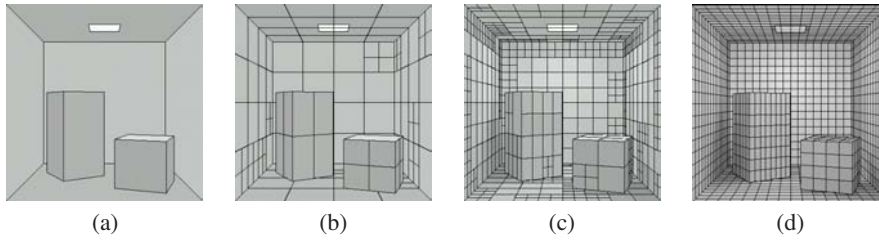

(a)　　　　　(b)　　　　　(c)　　　　　(d)

Figure 20: *Four different discretisations for the Cornell box. The total number of patches is, respectively,* 19, 196, 826, *and* 1924.

Scene	Patches	I_S	I_S^c
Fig.19a	4	0.415	2.626
Fig.19b	151	1.217	2.626
Fig.19c	793	1.445	2.626

Table 8: I_S^c *and* I_S *for the scenes in Figure 19. For each scene,* 10^7 *local lines have been cast.*

Scene	Patches	Lines (10^6)	I_S	I_S^c
Fig.20a	19	10	0.690	3.273
Fig.20b	196	10	2.199	3.273
Fig.20c	826	10	2.558	3.273
Fig.20d	1924	100	2.752	3.273

Table 9: I_S^c *and* I_S *for the scenes in Figure 20.*

7.2. Discretisation Accuracy

We know that the difference $I_S^c - I_S$ (60), always positive, expresses the *loss of information transfer* due to the discretisation. From this assumption, we can now take a leap forward with two fundamental proposals. The first appears naturally in an information-theory context and the second will be experimentally checked:

1. From an information-theory point of view, the ideal discretisation is the one that captures all the information transfer in a scene. Thus, between different discretisations of the same scene, the most precise will be the one that has a higher discrete mutual information I_S, i.e., the one that best captures the information transfer. With this in mind, we can express the *discretisation error* as the difference

$$\delta^v = I_S^c - I_S \qquad (61)$$

and the *relative discretisation error* as the quotient

$$\overline{\delta}^v = \frac{I_S^c - I_S}{I_S^c} \qquad (62)$$

The *relative discretisation accuracy* is given by $\frac{I_S}{I_S^c}$.

2. Continuous mutual information I_S^c expresses the difficulty in obtaining an accurate discretisation. The higher the I_S^c (i.e., when there is more information transfer in a scene), the more difficult it is to obtain an accurate discretisation, and probably more refinements will be necessary to achieve a given precision. From this point of view, the difficulty in discretising the interior of an empty sphere is null (the discretisation error is always equal to zero). The polyhedra that are "nearer" to the sphere are less complex than the others, and so easier to discretise.

8. Loss of Information Transfer due to the Discretisation

To obtain a refinement criterion based on the discretisation error between two patches, we need to consider both continuous and discrete patch-to-patch information transfers.

8.1. Mutual Information Matrix

As we have seen, discrete scene mutual information is given by

$$I_S = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} a_i F_{ij} \log\left(\frac{F_{ij}}{a_j}\right) \tag{63}$$

From this formula, the term

$$I_{ij} = a_i F_{ij} \log\left(\frac{F_{ij}}{a_j}\right) \tag{64}$$

can be considered as an element of a mutual information matrix, and it is easy to see that $I_{ij} = I_{ji}$. Each element represents the information transfer between patches i and j. Also, we can consider that

$$I_i = \sum_{j=1}^{n_p} a_i F_{ij} \log\left(\frac{F_{ij}}{a_j}\right) \tag{65}$$

expresses the information transfer from patch i. Thus, we can write

$$I_S = \sum_{i=1}^{n_p} I_i = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} I_{ij} \tag{66}$$

If we analyze the terms I_{ij}, we observe that negative values appear when $F_{ij} < a_j$. This situation reflects a very low interaction between the two patches involved. On the other hand, using the concavity property of the logarithm function (see Part II), it is easy to see that $I_i \geq 0$ (substituting a_k, b_k, and n by F_{ij}, a_j, and n_P, respectively).

The information transfer between two patches can be obtained more accurately if we consider the continuous mutual information between them. Thus, from the continuous mutual information, we obtain the following results.

Continuous information transfer:

$$\begin{aligned}
I_S^c &= \int_S \int_S \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \\
&= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y
\end{aligned} \tag{67}$$

Continuous information transfer due to patch i:

$$I_i^c = \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \tag{68}$$

Continuous information transfer between patches i and j:

$$I_{ij}^c = \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \tag{69}$$

This continuous measure expresses with maximum precision the visibility information transfer between two elements.

8.2. Discretisation Error Between Two Patches

As we have seen (section 7.2), a general discretisation error for a scene can be given by

$$\delta = I_S^c - I_S \geq 0 \tag{70}$$

In order to propose a refinement oracle for hierarchical radiosity, we are interested in the contribution to this discretisation error of the patch-to-patch interaction (i.e., discretisation error between two patches). So, we calculate, respectively, the difference between continuous and discrete patch-to-patch mutual information. For each one, three different Monte Carlo techniques can be used: patch-to-patch random lines, local lines and global lines.

1. **Patch-to-patch random lines:** The computation of (69)

$$I_{ij}^c = \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y$$

can be done with an area-to-area sampling (section 1.4), i.e., using random lines joining both elements i and j (the pdf is $\frac{1}{A_i A_j}$). For N_{ij} lines, we have

$$I_{ij}^c \approx \frac{A_i A_j}{A_T} \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \log\left(F(x_k, y_k) A_T\right) \tag{71}$$

where x_k and y_k are, respectively, the end-points on patches i and j of the k-th line.
From (64), I_{ij} can be expressed as

$$I_{ij} = \frac{A_i F_{ij}}{A_T} \log\left(\frac{F_{ij} A_T}{A_j}\right) = \frac{A_i A_j}{A_T} \frac{F_{ij}}{A_j} \log\left(\frac{F_{ij}}{A_j} A_T\right) \tag{72}$$

Now, taking $\frac{F_{ij}}{A_j} \approx \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k)$ (28), we obtain the

visibility discretisation error between patches i and j:

$$
\begin{aligned}
\delta_{ij}^v &= I_{ij}^c - I_{ij} \\
&\approx \frac{A_i A_j}{A_T} \left(\frac{1}{N_{ij}} \left(\sum_{k=1}^{N_{ij}} F(x_k, y_k) \log \left(F(x_k, y_k) A_T \right) \right) \right. \\
&\quad \left. - \frac{F_{ij}}{A_j} \log \left(\frac{F_{ij}}{A_j} A_T \right) \right) \\
&= \frac{A_i A_j}{A_T} \left(\frac{1}{N_{ij}} \left(\sum_{k=1}^{N_{ij}} F(x_k, y_k) \log (F(x_k, y_k)) \right) \right. \\
&\quad \left. - \frac{F_{ij}}{A_j} \log \left(\frac{F_{ij}}{A_j} \right) \right) \\
&= \frac{A_i A_j}{A_T} \left(\frac{1}{N_{ij}} \left(\sum_{k=1}^{N_{ij}} F(x_k, y_k) \log (F(x_k, y_k)) \right) \right. \\
&\quad \left. - \left(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \right) \log \left(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \right) \right) \geq 0
\end{aligned}
$$

(73)

where we have used the log-sum inequality (see Part II). This difference gives us the discretisation error between two elements and it is used as the basis for our mutual-information-based (MI-based) oracle. Observe also that δ_{ij}^v is symmetric: $\delta_{ij}^v = \delta_{ji}^v$.

2. **Local or global lines:**
   The computation of I_{ij}^c can also be done with uniformly distributed local or global lines (section 1.4). From (59), we obtain

$$
I_{ij}^c \approx \frac{A_i F_{ij}}{A_T} \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} \log (A_T F(x_k, y_k)) \tag{74}
$$

where N_{ij} is the number of local lines or segments of global lines which connect patches i and j. Hence, we find

$$
\begin{aligned}
\delta_{ij}^v &= I_{ij}^c - I_{ij} \\
&\approx \frac{A_i F_{ij}}{A_T} \left(\frac{1}{N_{ij}} \left(\sum_{k=1}^{N_{ij}} \log (F(x_k, y_k)) \right) - \log \frac{F_{ij}}{A_j} \right)
\end{aligned}
$$

(75)

As we expected, it is easy to see that the discretisation error between two spherical patches is equal to zero.

9. Mutual-Information-Based Oracle for Hierarchical Radiosity

We introduce in this section an information-theory oracle based on the radiosity kernel smoothness to be used in the hierarchical refinement algorithm. As the refinement strategy in hierarchical radiosity deals with one pair of elements at a time, we have to look for a similar interaction in our information theory framework.

The fundamental idea in our approach is the following: the difference between continuous and discrete patch-to-patch (or element-to-element) mutual information, i.e., discretisation error, gives us the *loss* of information transfer or the *maximum potential gain* of information transfer between two elements. Hence this difference can be interpreted as the *benefit to be gained by refining* and can be used as a decision criterion.

9.1. An Oracle Based on the Discretisation Error between Two Patches

To obtain a mutual-information-based oracle, we take a similar approach to the classic smoothness-based oracles, which multiplies $\rho_i B_j$ (from the radiosity equation (6)) by an expression of the visibility gradient between the two patches involved. In our case, the visibility gradient is given by the discretisation error $\delta_{ij}^v = I_{ij}^c - I_{ij}$, which also represents the *variation of the radiosity kernel*.

Our oracle will be based on the following considerations:

- In the radiosity equation (6)

$$
B_i = E_i + \rho_i \sum_{j=1}^{n_p} F_{ij} B_j
$$

the contribution of patch j to the radiosity of patch i is given by $\rho_i F_{ij} B_j$. Thus, the geometric factor, i.e., the radiosity kernel, is weighted by $\rho_i B_j$.

- The kernel-smoothness-based oracles reviewed in section 1.8, such as

$$
\rho_i (F_{ij}^{max} - F_{ij}^{min}) A_j B_j < \varepsilon
$$

and

$$
\rho_i max(F_{ij}^{max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{min}) A_j B_j < \varepsilon
$$

, try to capture the *variation of the radiosity kernel* using the maximum and minimum kernel values.

Our oracle proposal takes these two facts on board, weighting the variation of the radiosity kernel (expressed by the visibility discretisation error δ_{ij}^v between two patches) by $\rho_i B_j$. So, we find that the *mutual-information-based (MI-based) oracle* is given by

$$
\rho_i \delta_{ij}^v B_j < \varepsilon \tag{76}
$$

which can be computed with N_{ij} element-to-element random lines between elements i and j:

$$
\begin{aligned}
&\frac{\rho_i A_i A_j B_j}{A_T} \left(\frac{1}{N_{ij}} \left(\sum_{k=1}^{N_{ij}} F(x_k, y_k) \log (F(x_k, y_k)) \right) \right. \\
&\left. - \left(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \right) \log \left(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \right) \right) < \varepsilon
\end{aligned}
$$

(77)

Observe that in this expression the receiver area appears weighting the oracle and thus avoiding an excessively small receiver subdivision.

It is important to note that δ_{ij}^{v} can be used as an oracle for visibility. This only takes into account the variation of the radiosity kernel and the areas of the patches involved.

9.2. Empirical Results

To check the performance of the MI-based oracle, we have implemented a power-based oracle (36), a classic kernel-smoothness-based (KS-based) oracle (38) and our MI-based oracle (77) in the hierarchical Monte Carlo radiosity [BNN*98] method of the RenderPark [Com00] system (`www.renderpark.be`). It should be noted that our oracle can be used with any hierarchical radiosity method.

In our experiments, we use two scenes: the Cornell box (Figures 21 and 24) and the cube room (Figures 22, 23 and 25). Six different discretisations were generated for the Cornell box: three coarse (Figure 21I) and three finer ones (Figure 21II). These discretisations have been obtained from three meshing strategies based, respectively, on transported power (36) (Figures 21a.I and 21a.II), classic kernel smoothness (38) (Figures 21b.I and 21b.II), and mutual information (77) (Figures 21c.I and 21c.II). In a similar way, we compared our strategy with the KS-based strategy using two different views of the cube room scene (Figures 22 and 23). Both KS-based and MI-based oracles were evaluated for each discretisation decision with 10 element-to-element random lines (except in Figures 24 and 25, where only 4 rays were used). For the power-based oracle we used a cheap form factor estimate (see section 1.8).

In Figures 21I and 21II we see the behaviour of the three oracles for two different levels of discretisation. Using the power-based and KS-based oracles, the shadow of the small cube gets an accurate representation only at the finer level of discretisation, whereas the MI-based oracle already produces a good representation in the coarse mesh. The power-based oracle overdiscretises the rear wall and the top of the prism, as expected, while the smoothness-based oracles correct this effect. However, the MI-based oracle supports the change from a coarse to a finer mesh much better (see again the rear wall).

Figures 22 and 23 show the behaviour of the classic KS-based and MI-based oracle for the cube room scene. Observe the accurate representation of the shadow of the chair near the right wall (Figure 22b) and front wall (Figure 23b) obtained by the MI-based oracle. Observe also the much better discrimination in the mesh, seen for instance on the floor and walls, and how the shadows on the table are represented more accurately in Figure 23b.

In Figures 24 and 25, the robustness of the classic KS-based and MI-based oracle are tested against a decrease from 10 to 4 point-to-point form factor computations for each oracle evaluation. The performance of the classic KS-based oracle degenerates to a degree similar to the power-based oracle, see for instance the rear wall in Figure 24a (compare with

Figures 21b.I) and the same happens in Figure 25a (compare with Figures 22a). On the other hand, the MI-based oracle maintains most of its good performance (compare Figure 24b with Figures 21c.II). See also the shadow of the chair near the right wall in Figures 25b and 22b.

In Figure 26 we show a more accurate solution computed with the MI-based oracle, 10 element-to-element random lines for each oracle evaluation and 2684260 rays for radiosity computation.

References

[Bek99] BEKAERT P.: *Hierarchical and Stochastic Algorithms for Radiosity*. PhD thesis, Katholieke Universiteit Leuven, Leuven, Belgium, December 1999.

[BNN*98] BEKAERT P., NEUMANN L., NEUMANN A., SBERT M., WILLEMS Y. D.: Hierarchical Monte Carlo radiosity. In *Rendering Techniques'98 (Proceedings of the 9th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1998), Drettakis G., Max N., (Eds.), Springer-Verlag Vienna-New York, pp. 259–268. Held in Vienna, Austria.

[BP97] BADII R., POLITI A.: *Complexity. Hierarchical Structures and Scaling in Physics*. Cambridge University Press, 1997.

[BW96] BEKAERT P., WILLEMS Y. D.: Error control for radiosity. In *Rendering Techniques'96 (Proceedings of the 7th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1996), Pueyo X., Schröder P., (Eds.), Springer-Verlag Vienna-New York, pp. 153–164. Held in Vienna, Austria.

[CG85] COHEN M. F., GREENBERG D. P.: The hemicube: A radiosity solution for complex environments. *Computer Graphics (Proceedings of SIGGRAPH'85) 19*, 3 (July 1985), 31–40. Held in San Francisco (CA), USA.

[Col74] COLEMAN R.: *Stochastic Processes*. George Allen & Unwin Ltd., London, UK, 1974.

[Com00] COMPUTER GRAPHICS RESEARCH GROUP: *RenderPark: A Photorealistic Rendering Tool*. Katholieke Universiteit Leuven, Leuven, Belgium, November 2000.

[CT91] COVER T. M., THOMAS J. A.: *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[CW93] COHEN M. F., WALLACE J. R.: *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Boston (MA), USA, 1993.

[FC98] FELDMAN D. P., CRUTCHFIELD J. P.: *Discovering Noncritical Organization: Statistical Mechanical, Information Theoretic and Computational Views of Patterns in One-Dimensional Spin Systems*. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998.
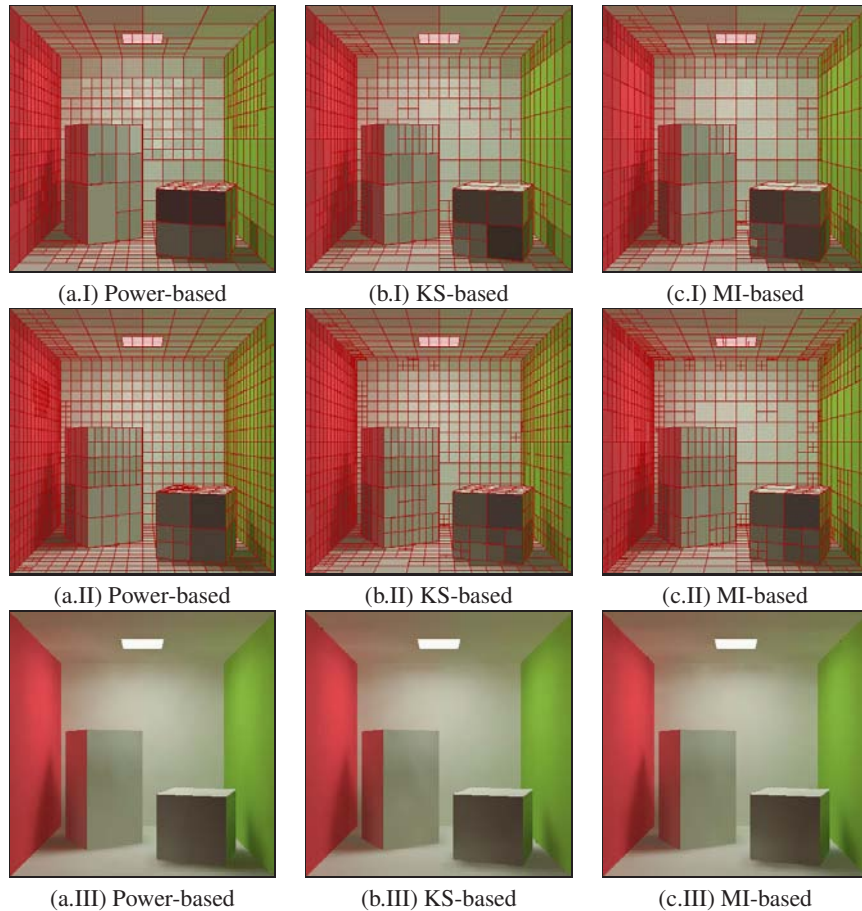
(a.I) Power-based   (b.I) KS-based   (c.I) MI-based

(a.II) Power-based   (b.II) KS-based   (c.II) MI-based

(a.III) Power-based   (b.III) KS-based   (c.III) MI-based

Figure 21: *Power-based (a), KS-based (b) and MI-based (c) methods with the Cornell box scene. A coarse mesh is shown in (I) with 1051 (a.I), 1039 (b.I), and 1047 (c.I) patches, with 19472 rays for the radiosity computation. A fine mesh is shown in (II) with 1979 (a.II), 1955 (b.II), and 1995 (c.II) patches, with 116780 rays for the radiosity computation. The Gouraud shaded solution for (II) is shown in (III). For images (b) and (c), 10 rays are cast for each oracle evaluation.*

[FdABS99] FEIXAS M., DEL ACEBO E., BEKAERT P., SBERT M.: An information theory framework for the analysis of scene complexity. *Computer Graphics Forum (Proceedings of Eurographics'99) 18*, 3 (September 1999), 95–106.

[Fel97] FELDMAN D. P.: A brief introduction to: Information theory, excess entropy and computational mechanics, 1997.

[FH96] FRISKEN S. F., HUBBOLD R. J.: Efficient hierarchical refinement and clustering for radiosity in complex environments. *Computer Graphics Forum 15*, 5 (1996), 297–310.

[FP93] FEDA M., PURGATHOFER W.: Progressive ray refinement for Monte Carlo radiosity. In *Eurographics Rendering Workshop 1993* (June 1993), Cohen M. F., Puech C., Sillion F. X., (Eds.), Eurographics Association, pp. 15–26. Held in Paris, France.

[FS98] FEIXAS M., SBERT M.: *Scene Continuous Mutual Information as Least Upper Bound of Discrete One*. Research Report IIiA–98–27–RR, Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 1998.

[Gla95] GLASSNER A. S.: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995.

[GSCH93] GORTLER S. J., SCHRÖDER P., COHEN M. F., HANRAHAN P.: Wavelet radiosity. *Computer Graphics (Proceedings of SIGGRAPH'93) 27* (August 1993), 221–230. Held in Anaheim (CA), USA.

[GTGB84] GORAL C. M., TORRANCE K. E., GREENBERG D. P., BATTAILE B.: Modelling the interaction of light between diffuse surfaces. *Computer Graphics (Proceedings of SIGGRAPH'84) 18*, 3 (July 1984), 213–222. Held in Minneapolis (MN), USA.

(a.I) KS-based   (b.I) MI-based

(a.II) KS-based   (b.II) MI-based

Figure 22: *KS-based (a) and MI-based (b) methods with the cube room scene showing the mesh (I) and Gouraud shaded solution (II). The number of patches is 13902 and 13878, respectively. For each scene, we cast 402650 rays for radiosity computation and 10 rays for each oracle evaluation.*

(a.I) KS-based   (b.I) MI-based

(a.II) KS-based   (b.II) MI-based

Figure 23: *A different view of the scene shown in Figure 22.*

| (a) KS-based | (b) MI-based | (c) MI-based |
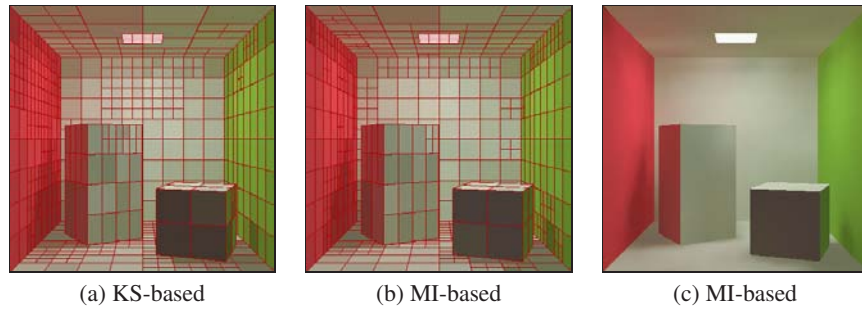
Figure 24: *KS-based (a) and MI-based (b) methods with the Cornell box scene showing the mesh. The number of patches is 875 and 891, respectively. For each scene, we cast 19458 rays for radiosity computation and 4 rays for each oracle evaluation. The Gouraud shaded solution for (b) is shown in (c).*

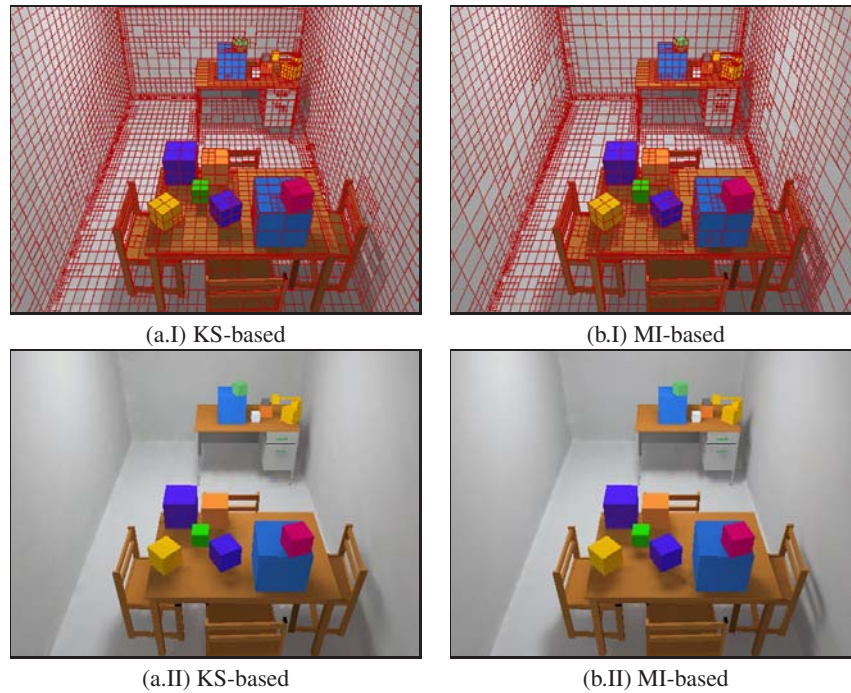| (a.I) KS-based | (b.I) MI-based |

| (a.II) KS-based | (b.II) MI-based |

Figure 25: *KS-based (a) and MI-based (b) methods with the cube room scene showing the mesh (I) and Gouraud shaded solution (II). The number of patches is 13690 and 13758, respectively. For each scene, we cast 402565 rays for radiosity computation and 4 rays for each oracle evaluation.*
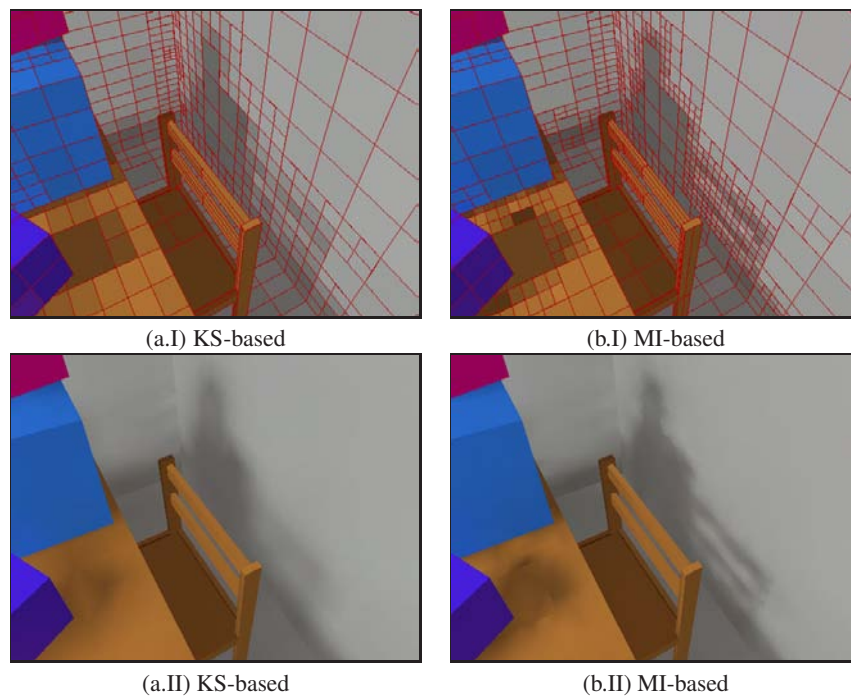
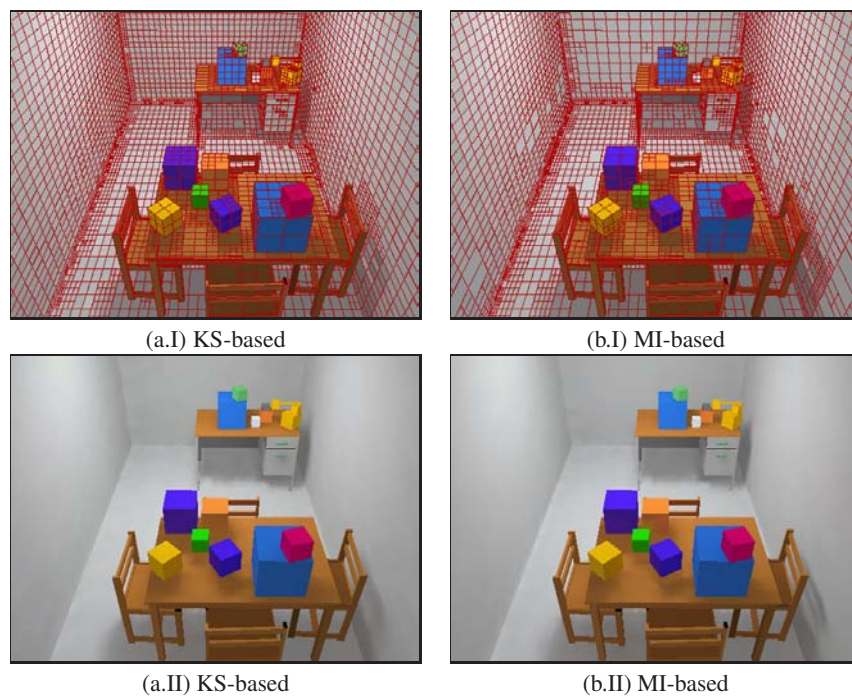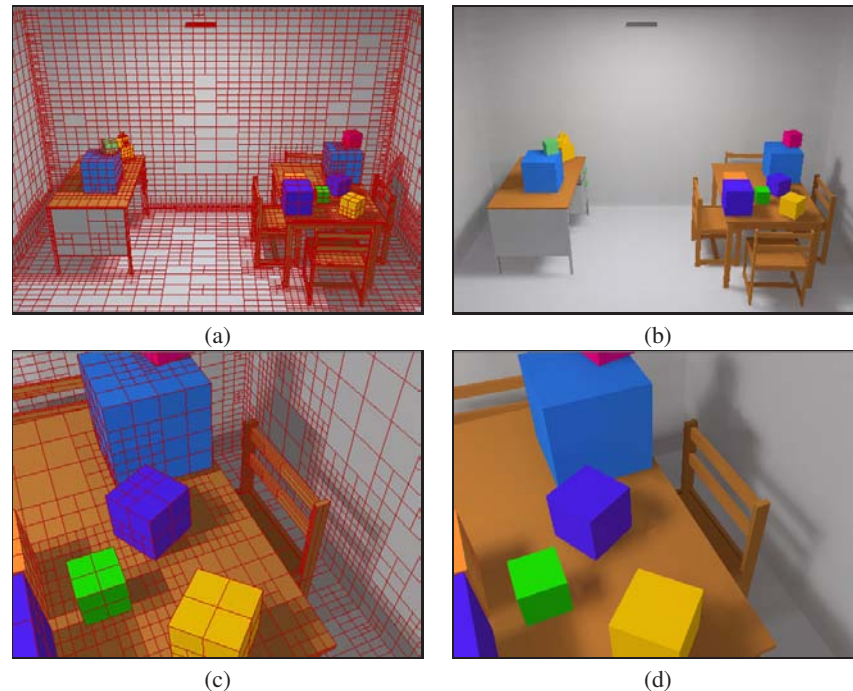**Figure 26:** *MI-based method with the scene shown in Figure* 22. *The number of patches is* 18338. *We cast* 2684260 *rays for radiosity computation and* 10 *rays for each oracle evaluation.*

[HS98] HOLZSCHUCH N., SILLION F.: An exhaustive error-bounding algorithm for hierarchical radiosity. *Computer Graphics Forum 17*, 4 (1998), 197–218.

[HSA91] HANRAHAN P., SALZMAN D., AUPPERLE L.: A rapid hierarchical radiosity algorithm. *Computer Graphics (Proceedings of SIGGRAPH'91) 25*, 4 (July 1991), 197–206. Held in Las Vegas (NV), USA.

[Kaj86] KAJIYA J. T.: The rendering equation. *Computer Graphics (Proceedings of SIGGRAPH'86) 20*, 4 (August 1986), 143–150. Held in Dallas (TX), USA.

[KW86] KALOS M. H., WHITLOCK P. A.: *The Monte Carlo Method.* John Wiley & Sons, 1986.

[Li90] LI W.: Mutual information functions versus correlation functions. *Journal of Statistical Physics 60*, 5/6 (1990), 823–837.

[Li91] LI W.: On the relationship between complexity and entropy for markov chains and regular languages. *Complex Systems 5*, 4 (1991), 381–399.

[LSG94] LISCHINSKI D., SMITS B., GREENBERG D. P.: Bounds and error estimates for radiosity. *Computer Graphics (Proceedings of SIGGRAPH'94) 28* (July 1994), 67–74. Held in Orlando (FL), USA.

[LTG93] LISCHINSKI D., TAMPIERI F., GREENBERG D. P.: Combining hierarchical radiosity and discontinuity meshing. *Computer Graphics (Proceedings of SIG-GRAPH'93) 27* (August 1993), 199–208. Held in Anaheim (CA), USA.

[MR95] MOTWANI R., RAGHAVAN P.: *Randomized Algorithms.* Cambridge University Press, New York (NY), USA, 1995.

[Neu95] NEUMANN L.: Monte Carlo radiosity. *Computing 55*, 1 (1995), 23–42.

[NFP95] NEUMANN L., FEDA M., PURGATHOFER W.: A new stochastic radiosity method for highly complex scenes. In *Photorealistic Rendering Techniques (Proceedings of the 5th Eurographics Workshop on Rendering)* (New York (NY), USA, 1995), Haas S., Mueller S., Sakas G., Shirley P., (Eds.), Springer-Verlag Vienna-New York, pp. 201–213. Held in Darmstad, Germany (June 1994).

[NN85] NISHITA T., NAKAME E.: Continuous tone representation of 3-d objects taking account of shadows and interreflection. *Computer Graphics (Proceedings of SIG-GRAPH'85) 19*, 3 (July 1985), 23–30. Held in San Francisco (CA), USA.

[NNB97] NEUMANN L., NEUMANN A., BEKAERT P.: Radiosity with well distributed ray sets. Fellner D., Szirmay-Kalos L., (Eds.), vol. 16, Blackwell Publishers, pp. 261–270. Held in Budapest, Hungary.

[Pap84] PAPOULIS A.: *Probability, Random Variables,*

and Stochastic Processes, 2 ed. McGraw-Hill, New York (NY), USA, 1984.

[PB95]　PATTANAIK S. N., BOUATOUCH K.: Linear radiosity with error estimation. In *Rendering Techniques'95 (Proceedings of the 6th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1995), Hanrahan P., Purgathofer W., (Eds.), Springer-Verlag Vienna-New York. Held in Dublin, Eire.

[PM92]　PATTANAIK S. N., MUDUR S. P.: Computation of global illumination by Monte Carlo simulation of the particle model of light. In *Proceedings of the 3th Eurographics Workshop on Rendering* (May 1992), pp. 71–83. Held in Bristol, UK.

[Rub81]　RUBINSTEIN R. Y.: *Simulation and the Monte Carlo Method*. John Wiley & Sons, New York (NY), USA, 1981.

[San76]　SANTALÓ L. A.: *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading (MA), USA, 1976.

[SAS92]　SMITS B. E., ARVO J., SALESIN D.: An importance-driven radiosity algorithm. *Computer Graphics (Proceedings of SIGGRAPH'92) 26*, 2 (July 1992), 273–282. Held in Chicago (IL), USA.

[Sbe93]　SBERT M.: An integral geometry based method for fast form-factor computation. *Computer Graphics Forum (Proceedings of Eurographics'93) 12*, 3 (1993), 409–420. Held in Barcelona, Spain.

[Sbe96]　SBERT M.: *The Use of Global Random Directions to Compute Radiosity. Global Monte Carlo Methods*. PhD thesis, Universitat Politècnica de Catalunya, Barcelona, Spain, November 1996.

[Sbe97]　SBERT M.: Error and complexity of random walk Monte Carlo radiosity. *IEEE Transactions on Visualization and Computer Graphics 3*, 1 (March 1997), 23–38.

[SC99]　SHALIZI C. R., CRUTCHFIELD J. P.: *Computational Mechanics: Pattern and Prediction, Structure and Simplicity*. Working Paper 99–07–044, Santa Fe Insitute, Santa Fe (NM), USA, July 1999.

[SH92]　SIEGEL R., HOWELL J. R.: *Thermal Radiation Heat Transfer*, 3 ed. Hemisphere Publishing Corporation, New York (NY), USA, 1992.

[SH93]　SCHRÖDER P., HANRAHAN P.: On the form factor between two polygons. *Computer Graphics Proceedings (Proceedings of SIGGRAPH'93) 27* (August 1993), 163–164. Held in Anaheim (CA), USA.

[Shi90]　SHIRLEY P.: A ray tracing method for illumination calculation in diffuse–specular scenes. In *Proceedings of Graphics Interface'90* (May 1990), pp. 205–212. Held in Halifax, Canada.

[SP94]　SILLION F. X., PUECH C.: *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1994.

[SSS97]　STAMMINGER M., SLUSALLEK P., SEIDEL H.-P.: Bounded radiosity-illumination on general surfaces and clusters. *Computer Graphics Forum (Proceedings of Eurographics'97) 16*, 3 (1997), 300–317. Held in Budapest, Hungary.

[Stu97]　STUDHOLME C.: *Measures of 3D Medical Image Alignment*. PhD thesis, University of London, London, UK, August 1997.

[TH93]　TELLER S. J., HANRAHAN P.: Global visibility algorithms for illumination computation. *Computer Graphics (Proceedings of SIGGRAPH'93) 27* (August 1993), 239–246. Held in Anaheim (CA), USA.

Applications of Information Theory to Computer Graphics
Part IV: Adaptive Refinement for Ray-tracing

Jaume Rigau, Miquel Feixas, and Mateu Sbert

University of Girona, Spain

1. Introduction

Although ray-tracing is a straightforward and powerful image synthesis technique, it usually requires many rays per pixel to eliminate the aliasing or noise in the final image. However, not all the pixels in the image require the same number of rays. The edge of an object, the contour of a shadow, and a high illumination gradient will require a much better treatment than a region with almost uniform illumination. To this effect, many pixel supersampling refinement criteria have been defined in the literature.

The measures used in these criteria are based on intensities (image space) and also on geometry (object space). They are also useful for an adaptive subdivision of image space for progressive refinement [PS89]. Some of them have been applied in the image based rendering field for weighting pixel colour for reconstruction [PCD*97] and adaptive sampling strategies [DC96, DCV97], and creating a priority scheme for sampling in interactive rendering [SS00]. The final objective is always to find the best final-image quality with a reasonable cost. In order to do this, we have to sample each pixel of the image plane carefully. It is essential to have a quantitative measure in order to evaluate when there is sufficient information about the pixel.

The data of a sample set through the pixel can be used to calculate a pixel homogeneity measure from two different points of view: *radiance* and *visibility*. The information which we will manipulate will be exclusively colour (radiance) and geometry (visibility), essential parameters for deciding on the "quality" of a pixel. In this context, the Shannon entropy will be interpreted as a measure of the degree of homogeneity of a pixel in the sense that the more heterogeneous the pixel, the more difficult it is to obtain its actual value. From it, we associate homogeneity with quality, so that the need for pixel refinement is proportional to the lack of quality (i.e., heterogeneity of the samples). The idea behind the new scheme is to obtain sufficient information in the refinement algorithm in order to find out the sampling needs.

Consequently, in this part we present a framework for entropy-based sampling applied to ray-tracing methods. First, definitions of new measures of *pixel quality* based on entropy are presented (§3). Next, we present the pixel quality as a measure of *pixel contrast* (§4). Then, this contrast is applied to classic *supersampling* ray-tracing (§5) and *adaptive sampling* (§6). This framework is easily adaptable to other stochastic processes which require measures of quality in order to reach decisions.

2. Previous Work

Three principal subproblems make up the process of obtaining a good quality image: efficient sample generation, adaptive control of the sampling rate, and filtering for image reconstruction [PS89]. Many approaches are to be found to deal with them:

1. Different pixel sampling methods have been introduced, among them: jittered sampling [CPC84, DW85], Poisson disk sampling [DW85, Mit87, MF92], hierarchical sampling [Kaj86], complete stratification at each refinement level [Sch91], importance sampling [Shi90], and quasi-Monte Carlo sampling [KH94, OA96].

2. 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 [DW85] 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 [Mit87] proposes a contrast [Cae81] based on the characteristics of the human eye; Lee et al. [LRU85], Purgathofer [Pur86], and Tamstorf and Jensen [TJ97] develop different methods based on the variance of the samples with their respective confidence intervals.

3. Samples are filtered to produce the final pixel values. Different filter shapes have been used in image reconstruc-

tion: box filter, triangular filter, Gaussian filter, multi-stage filter, etc. (see [Gla95]).

We review here three commonly used refinement criteria: contrast, depth difference, and variance of the samples.

Mitchell, in [Mit87], uses a contrast measure [Cae81] for each RGB channel defined by

$$C = \frac{I_{\max} - I_{\min}}{I_{\max} + I_{\min}}, \qquad (1)$$

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 [SS00], within an interactive rendering context, Simmons uses a *priority* value p_c based on the above concepts (contrast and perception) [Mit87, Gla95] defined by

$$p_c = .4\frac{r_{\max} - r_{\min}}{r_{\max} + r_{\min}}\overline{r} + .3\frac{g_{\max} - g_{\min}}{g_{\max} + g_{\min}}\overline{g} + .6\frac{b_{\max} - b_{\min}}{b_{\max} + b_{\min}}\overline{b} \qquad (2)$$

where max, min, and the overline represent, respectively, the maximum, minimum, and average values for r, g, and b colour channels.

On the other hand, a useful and simple geometric measure for refinement is *depth difference*, used recently in image based rendering [DC96, DCV97, PCD*97] and interactive rendering [SS00]. Depth difference is given by

$$p_d = 1 - \frac{d_{\min}}{d_{\max}} \qquad (3)$$

where $d_{\max}$ and $d_{\min}$ represent maximum and minimum distance. In [SS00], p_c and p_d measures are combined in

$$p_v = \delta p_c + (1 - \delta)p_d \qquad (4)$$

with $\delta = 0.9$.

The basic idea of variance-based methods [LRU85, Pur86, TJ97] is to continue sampling until the confidence level or probability that the true value L is within a given tolerance d of the estimate value $\widehat{L}$ is $1 - \alpha$:

$$\Pr[L \in (\widehat{L} - d, \widehat{L} + d)] = 1 - \alpha. \qquad (5)$$

Mitchell considers that variance is a poor measure of visual perception of local variation [Mit87]. Kirk and Arvo showed that these methods are biased and proposed a simple correction scheme [KA91].

Refinement criteria have also been applied in the image-based rendering field to weight pixel colour for reconstruction purposes [PCD*97] and adaptive sampling strategies [DC96, DCV97]. Also Bolin and Meyer [BM98] have developed a perceptually-based approach using statistical and vision models.

3. Pixel Quality

In this section we introduce a new pixel quality measure, the *pixel entropy*. This measure will be defined from the information provided by set of samples on the image plane. We use the following sets:

- Let P be the set of pixels of the image plane with $|\mathsf{P}| = N_{\mathsf{p}} > 0$.
- Let S_{p} be the set of samples of a pixel $\mathsf{p} \in \mathsf{P}$ with $|S_{\mathsf{p}}| = N_s^{\mathsf{p}} > 1$.
- Let S_{P} be the set of samples of the image plane where $S_{\mathsf{P}} = \cup_{\mathsf{p} \in \mathsf{P}} S_{\mathsf{p}}$ with $|S_{\mathsf{P}}| = N_s^{\mathsf{P}} = \sum_{\mathsf{p} \in \mathsf{P}} N_s^{\mathsf{p}}$.

The implementation of a sample consists in casting a ray r_v^{Θ} from a scene viewpoint v through the image plane and, in particular, through a pixel: $\Theta \in \Omega_{v \to \mathsf{P}}$. Let us consider that each sample $s \in S_{\mathsf{P}}$ that hits a scene surface gives us information about the colour, distance and orientation of the hit point with respect to the viewpoint.

The definition of entropy

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i, \qquad (6)$$

measures the expectation of the surprise of the distribution p and it can be considered also to be a measure of its homogeneity. From the sample set and from the entropy, two different quality measures are defined, *pixel colour entropy* and *pixel geometry entropy*, based on the colour and geometry respectively.

3.1. Pixel Colour Entropy

Our first objective is to define the pixel colour entropy. We start with a global definition of entropy concerning all the samples passing through the image plane. We consider that the colour belongs to a colour system **c** structured in components called *colour channels*. Without loss of generality, in the majority of cases our colour measures will refer to a single channel $c \in \mathbf{c}$, $c(s)$ being the colour channel data of a sample $s \in S$ (e.g., radiance, luminance, and RGB values).

Let us consider the probability of each image plane sample as its colour channel contribution relative to the whole of the image plane sample set. Thus,

Definition 1 The image plane channel entropy of a channel c is given by

$$H^c(\mathsf{P}) = -\sum_{i=1}^{N_s^{\mathsf{P}}} r_i \log r_i \qquad r_i = \frac{c(s_i)}{\sum_{j=1}^{N_s^{\mathsf{P}}} c(s_j)}, \qquad (7)$$

where r_i represents the channel colour fraction of sample $s_i \in S_{\mathsf{P}}$.

This measure can be interpreted as the colour channel homogeneity of the samples passing through the image plane. Analogously, at the pixel level, we consider the probability of each pixel sample as its colour channel contribution relative to the whole of the pixel sample set. Then,

Definition 2 The pixel channel entropy of a channel c is given by

$$H^c(\mathsf{p}) = -\sum_{i=1}^{N_s^\mathsf{P}} p_i \log p_i \qquad p_i = \frac{c(s_i)}{\sum_{j=1}^{N_s^\mathsf{P}} c(s_j)}, \qquad (8)$$

where p_i represents the channel colour fraction of sample $s_i \in S_\mathsf{p}$.

From the properties of the entropy, the image plane channel entropy ranges from 0 to $\log N_s^\mathsf{P}$ and the pixel channel from 0 to $\log N_s^\mathsf{p}$. The maximum values are obtained when the channel colour of all the samples is the same (i.e., we have an uniform probability distribution). Using the grouping property of entropy (see Part II) it is easy to see that image plane and pixel channel entropies can be related in the following way:

$$H^c(\mathsf{P}) = \sum_{i=1}^{N_\mathsf{p}} q_i H^c(\mathsf{p}_i) - \sum_{i=1}^{N_\mathsf{p}} q_i \log q_i = \sum_{i=1}^{N_\mathsf{p}} q_i H^c(\mathsf{p}_i) + H_1^c(\mathsf{P}),$$
$$(9)$$

where $q_i = \sum_{j=1}^{N_s^{\mathsf{p}_i}} r_j$ is the *importance* (sum of probabilities) of pixel p_i, $H^c(\mathsf{p}_i)$ is the channel entropy of pixel p_i, and $H_1^c(\mathsf{P}) = -\sum_{i=1}^{N_\mathsf{p}} q_i \log q_i$ is the *importance entropy* of the image plane calculated from the importance of each pixel. Thus, the global entropy of the image plane is the sum of all the pixel entropies, weighted by the importance of each pixel, and the importance entropy obtained from the importance of each pixel.

The image plane and pixel entropies can be interpreted as the colour *homogeneity* or *uniformity* measured by its sample set and thus can be considered measures of the *quality* of the colour channel (i.e., lack of heterogeneity and noise). We can also observe that the entropy increases with the number of samples. In order to give a pixel quality measure between 0 and 1, the pixel channel entropy can be normalised with $\log N_s^\mathsf{P}$. Thus,

Definition 3 The pixel channel quality of a channel c is given by

$$Q^c(\mathsf{p}) = \frac{H^c(\mathsf{p})}{\log N_s^\mathsf{P}}. \qquad (10)$$

If we want to consider the global quality of a pixel, we need only mix its set of channels. Then,

Definition 4 The pixel colour quality of a colour system $\mathbf{c}$ is given by the weighting of its pixel channel qualities:

$$\mathbf{Q^c}(\mathsf{p}) = \frac{\sum_{c \in \mathbf{c}} w^c Q^c(\mathsf{p})}{\sum_{c \in \mathbf{c}} w^c}, \qquad (11)$$

where w^c is the weight of channel c.

The weighted values depend on each colour system. Without a priori information, the same weight per channel can be considered, otherwise a weight based on human perception (for an sRGB system, $w^R = 0.2126$, $w^G = 0.7152$, and $w^B = 0.0722$). This measure will enable us to define a new colour contrast measure for pixel sampling (§4.1). Note that the larger the number of samples the more accurate the quality measure.

In Fig. 1.*b*, we present a *colour quality map* to show the colour quality of all the pixels from Fig. 1.*a* using an sRGB colour system with the same weight by channel. With respect to the colour scale used, the minimum quality corresponds to the blue and the maximum to the red (in order to observe more details in the colour quality maps, the *outliers* are reduced to the borders of the interval $[-k\sigma, k\sigma]$ where σ is the standard deviation of the results and k is a parameter that modulates the width of the interval). A low quality in shadow areas and edges can be observed.

3.2. Pixel Geometry Entropy

Similar concepts introduced in the above section can be defined in this one with respect to a geometric measure. If $x = \Lambda(v, \Theta)$ is the hitpoint of a sample ray $s = r_v^\Theta$, the geometric information of each sample is given by $\theta_{n_x}^{-\Theta}$ (i.e., the angle of the normal at the hit point) and by the distance r_{vx} between this point and the origin of the ray (i.e., ray length). We take

$$g(s) = \frac{\cos \theta_{n_x}^{-\Theta}}{r_{vx}^2} \qquad (12)$$

as a geometry factor of a sample. This value provides a quality measure of visibility of a scene point from the observer's point of view.

Let us define the probability of each image plane sample as its relative geometric contribution to the whole image plane sample set.

Definition 5 The image plane geometry entropy is given by

$$H^g(\mathsf{P}) = -\sum_{i=1}^{N_s^\mathsf{P}} r_i \log r_i \qquad r_i = \frac{g(s_i)}{\sum_{j=1}^{N_s^\mathsf{P}} g(s_j)}, \qquad (13)$$

where r_i represents the geometry fraction of sample $s_i \in S_\mathsf{P}$.

Considering the probability of each pixel sample as its relative geometrical contribution to the whole of the pixel sample set we have

Definition 6 The pixel geometry entropy is given by

$$H^g(\mathsf{p}) = -\sum_{i=1}^{N_s^\mathsf{P}} p_i \log p_i \qquad p_i = \frac{g(s_i)}{\sum_{j=1}^{N_s^\mathsf{P}} g(s_j)}, \qquad (14)$$

where p_i represents the geometry fraction of sample $s_i \in S_\mathsf{p}$.

Analogously to the pixel colour entropy (9), an identical relation can be established between the geometric entropies of the image plane and the pixel. We can also normalise the pixel geometry entropy and therefore,

Definition 7 The pixel geometry quality is given by

$$Q^g(\mathsf{p}) = \frac{H^g(\mathsf{p})}{\log N_s^\mathsf{p}}. \tag{15}$$

In Fig. 1.*c* we show the *geometry quality map* from Fig. 1.*a* based on a grey scale. The lowest entropy corresponds to the darkest part, the highest entropy to the lightest (the outliers have the same treatment as in the colour quality maps). Observe that the edges have a very low entropy and are very clearly emphasised.

4. Pixel Contrast

In this section we present new pixel contrast measures based on pixel entropy (§3). As the entropy represents the homogeneity of the information brought back by the samples (i.e., rays crossing a pixel), we can define a simple measure which expresses the diversity or contrast of a pixel.

4.1. Pixel Colour Contrast

In the colour theory, the *colour contrast* is the phenomenon that alters the observation of the colours depending on their surroundings (the origin of the colour contrast is the way in which the information is transmitted from the retinal photoreceptors to the brain and the name of this study is the *colour opponency* theory [Gla95]). We use the same words to express the degree of heterogeneity of the colour in the region defined by a pixel given that this value depends directly on the colours that are around it. As we have seen, $H^c(\mathsf{p})$ represents the entropy or the degree of colour homogeneity of pixel p. From this measure,

Definition 8 The pixel channel contrast of a channel c is given by

$$C^c(\mathsf{p}) = 1 - Q^c(\mathsf{p}) = 1 - \frac{H^c(\mathsf{p})}{\log N_s^\mathsf{p}}. \tag{16}$$

It represents the colour channel heterogeneity or contrast of a pixel with a range of $[0, 1]$. We can also introduce the pixel binary contrast from minimum and maximum colour channel probabilities captured by this pixel. This measure is obtained from the binary entropy of these values. Thus,

Definition 9 The pixel channel binary contrast of a channel c is given by

$$C_b^c(\mathsf{p}) = 1 - H_b^c(\mathsf{p}) H_b^c(\mathsf{p}) = H(\{\frac{p_{\min}}{p_{\min} + p_{\max}}, \frac{p_{\max}}{p_{\min} + p_{\max}}\}), \tag{17}$$

where $H_b^c(\mathsf{p})$ is the binary entropy of the minimum and maximum channel colour probabilities, $p_{\min}$ and $p_{\max}$, respectively.

Both measures, $H_b^c(\mathsf{p})$ and $C_b^c(\mathsf{p})$, range also between 0 and 1 due to the fact that, in this case, only two values are taken into account. As we will see in our experiments, this binary measure yields more radical contrast than $C^c(\mathsf{p})$.

Similarly to previous works [Mit87, Gla95, SS00], we can obtain the colour contrast of a pixel by averaging all the colour channel contrasts weighted by their respective importances (colour channel average). This avoids oversampling on the areas with small colour values. Then, considering all the colour channels,

Definition 10 The pixel colour contrast of a colour system **c** is given by the weighting of its pixel channel contrasts:

$$\mathbf{C^c}(\mathsf{p}) = \frac{\sum_{c \in \mathbf{c}} w^c \bar{c} C^c(\mathsf{p})}{\sum_{c \in \mathbf{c}} w^c \bar{c}} \qquad \bar{c} = \frac{1}{N_s^\mathsf{p}} \sum_{i=1}^{N_s^\mathsf{p}} c(s_i), \tag{18}$$

where the channel contrasts are weighted by perceptual coefficients w^c, and $\bar{c}$ is the colour average in channel c of all $s \in S_\mathsf{p}$.

Definition 11 The pixel colour binary contrast of a colour system **c** is given by

$$\mathbf{C}_b^\mathbf{c}(\mathsf{p}) = \frac{\sum_{c \in \mathbf{c}} w^c \bar{c} C_b^c(\mathsf{p})}{\sum_{c \in \mathbf{c}} w^c \bar{c}}. \tag{19}$$

In an sRGB system, the colour contrast measures ($\mathbf{C^{sRGB}}$ and $\mathbf{C}_b^{\mathbf{sRGB}}$) have three channels with coefficients w^R, w^G, and w^B. These values depend on the specific use of contrast, but in general they can take the values proposed in the pixel colour quality (11), or also, for a perceptual balance of the channels, they can take those of the thresholds proposed in [Mit87, SS00]: 0.4, 0.3, and 0.6, respectively.

This last option is chosen in the next examples where a *colour contrast map* is used as a visual representation of the contrast measures in the same way that the quality map is used for the quality measures. Thus, in Fig. 2 we show different colour contrast maps to compare the heuristic (2), p_c (Fig. 2.*b*), with measures $\mathbf{C^c}$ (Fig. 2.*c*), and $\mathbf{C}_b^\mathbf{c}$ (Fig. 2.*d*). Another comparison is shown in Fig. 3 with a more complex scene. We can observe how these measures present a very good behaviour in critical areas (represented by warm colours) such as object edges and shadow contours. With respect to Fig. 2.*b* and Fig. 3.*b*, our measures are more discriminating, especially the binary contrast.

4.2. Pixel Geometry Contrast

As we have seen in (14), H^g represents the entropy or the degree of geometric homogeneity of a pixel. From this measure,

Definition 12 The pixel geometry contrast is given by

$$C^g(\mathsf{p}) = 1 - Q^g(\mathsf{p}) = 1 - \frac{H^g(\mathsf{p})}{\log N_s^\mathsf{p}}. \tag{20}$$

Similarly to the above section, we introduce the pixel binary contrast from minimum and maximum geometry factor probabilities of this pixel. Thus,
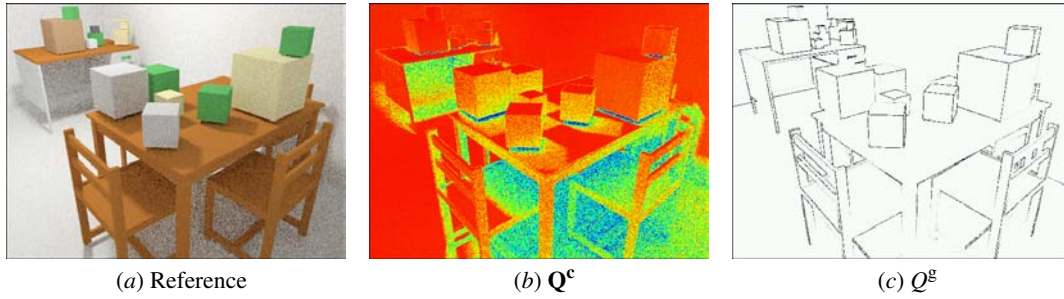
(*a*) Reference (*b*) $\mathbf{Q^c}$ (*c*) Q^g

Figure 1: *Colour and geometry quality maps. (a) Reference image obtained with $N_s^p = 8$. (b) Pixel colour quality $\mathbf{Q^{sRGB}}$ with the same weight per channel. (c) Pixel geometry quality Q^g.*

Definition 13 The pixel geometry binary contrast is given by

$$C_b^g(\mathsf{p}) = 1 - H_b^g(\mathsf{p}) H_b^g(\mathsf{p}) = H(\{\frac{p_{\min}}{p_{\min} + p_{\max}}, \frac{p_{\max}}{p_{\min} + p_{\max}}\}), \quad (21)$$

where $H_b^g(\mathsf{p})$ is the binary entropy of the minimum and maximum geometry factor probabilities, $p_{\min}$ and $p_{\max}$ respectively.

A third case can also be considered:

Definition 14 The pixel logarithmic-difference contrast is given by

$$C_{\log}^g(\mathsf{p}) = \log p_{\max} - \log p_{\min} = \log \frac{p_{\max}}{p_{\min}}. \quad (22)$$

This measure, introduced in Rigau et al. [RFBS01], is based on the gradient between the minimum and maximum complexity segments. As we will see, $C_{\log}^g(\mathsf{p})$ also shows a good behaviour.

For the previous reference scene in Fig. 2.*a*, we now show the geometry contrast measures using the corresponding maps in Figs. 4.*a–c*. These maps are compared with the map in Fig. 4.*d*, created using the depth difference heuristic p_d (3). The same comparison is carried out for the reference scene Fig. 3.*a* in Fig. 5. It can be seen that our measures capture the majority of edges because we take into account two components: distance and orientation. These geometry contrast maps have been generated by using the representation scale of the colour contrast maps in order to be able to compare, visually with each other, how the two types of contrast, colour (Fig. 3) and geometry (Fig. 5), work. The specialisation of each of the contrasts is evident: colour maps show the heterogeneity of regions while geometric maps identify edges.

4.3. Pixel Colour-Geometry Contrast

Finally, a combination of colour and geometry contrasts is considered. This combination enables us to graduate, with a coefficient δ between 0 and 1, the influence of both measures. Then,

Definition 15 The pixel contrast of colour system **c** is given by

$$C^{\mathbf{c}}(\mathsf{p}) = \delta \mathbf{C^c}(\mathsf{p}) + (1 - \delta) C^g(\mathsf{p}). \quad (23)$$

This combination can be made with any type of pixel colour contrast and geometry contrast. In general, good behaviour has been shown with binary contrasts (colour and geometry), and $\delta \in [0.8, 0.95]$.

We show for another scene, Fig. 6, two different lineal combinations. On the one hand, in Fig. 6.*a* we use the priority-value combination p_v (4), made up also of colour, p_c (2), and geometry, p_d (3). And, on the other hand, in Fig. 6.*b* we combine our measures $\mathbf{C_b^c}$ (19) and C^g (20). The same values $N_s^p = 4$ and $\delta = 0.9$ are used in both cases. A significant difference is observed: our combination tends to obtain more radical contrasts (highly or less complex cases) as opposed to the other option which takes values in a far more homogeneous interval. The explanation lies in the behaviour of the binary colour contrast which works exclusively with the extreme data.

5. Entropy-Based Supersampling

In this section, we apply the newly defined contrast measures to supersampling in a stochastic ray-tracing implementation.

5.1. Method

Ray-tracing is a point-sampling-based technique for image synthesis. Rays are traced from the camera through a pixel to sample radiance at the hitpoint in the scene, where radiance is usually computed by a random walk method. Since a finite set of samples is used, some of the information in the scene is lost. Thus, aliasing errors are unavoidable [DW85]. These errors can be reduced by using extra sampling, called *supersampling*, in regions where the sample values vary most.

In order to obtain reliable data to achieve photo-realistic effects (e.g., diffuse and specular interreflections, shadow
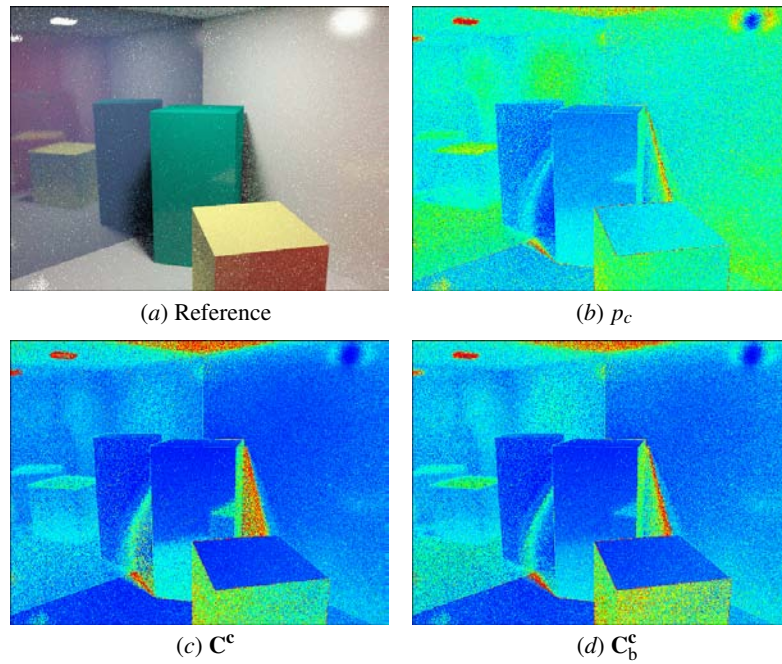
(a) Reference   (b) p_c

(c) $\mathbf{C^c}$ (d) $\mathbf{C_b^c}$

Figure 2: *Colour contrast maps. (a) Reference image obtained with $N_s^p = 8$. (b) Pixel colour contrast p_c (2). (c) Pixel colour contrast $\mathbf{C^c}$. (d) Pixel colour binary contrast $\mathbf{C_b^c}$.*

(a) Reference   (b) p_c
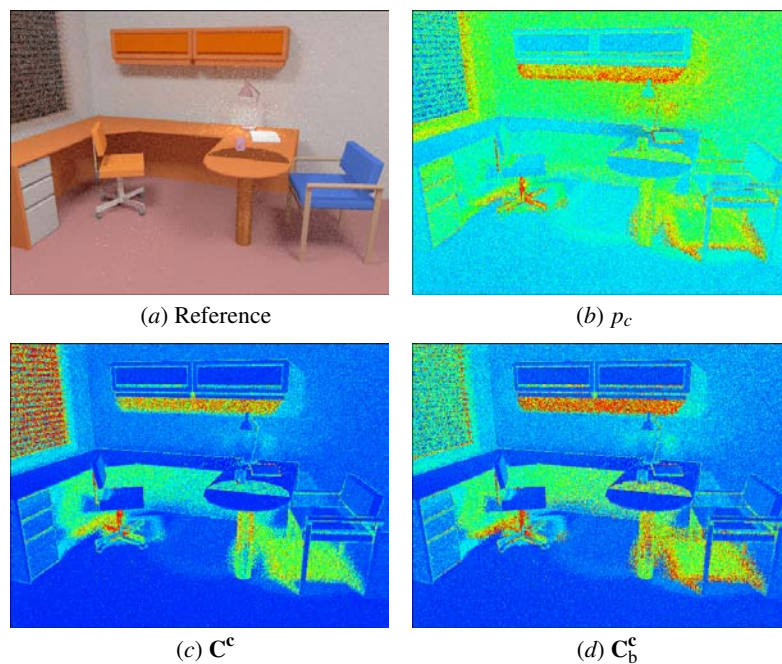
(c) $\mathbf{C^c}$ (d) $\mathbf{C_b^c}$

Figure 3: *Colour contrast maps. (a) Reference image obtained with $N_s^p = 8$. (b) Pixel colour contrast p_c (2). (c) Pixel colour contrast $\mathbf{C^c}$. (d) Pixel colour binary contrast $\mathbf{C_b^c}$.*
Credit: Modelled by Gregory J. Ward, Albany (CA), USA.

(a) C^g

(b) C^g_b
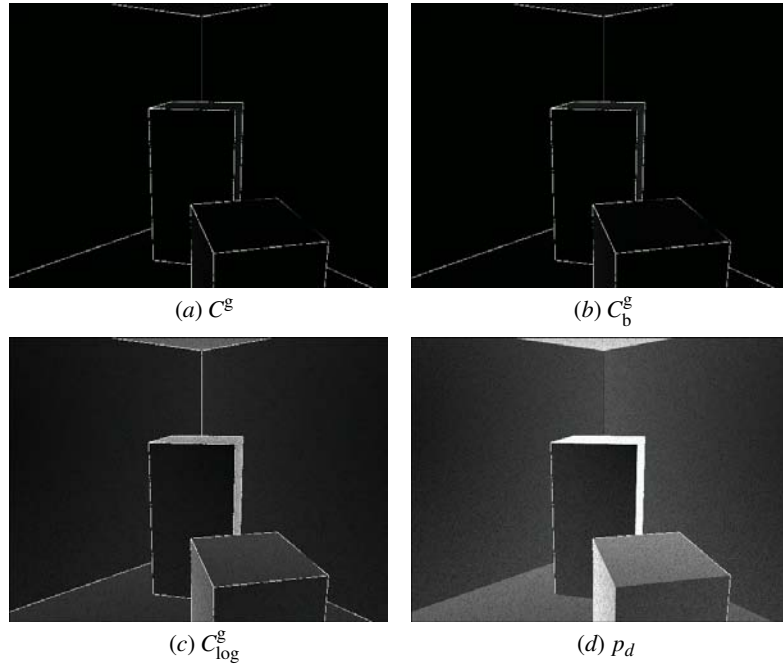
(c) $C^g_{\log}$

(d) p_d

Figure 4: *Geometry contrast maps from Fig. 2.a obtained with $N^p_s = 8$. (a) Pixel geometry contrast C^g. (b) Pixel geometry binary contrast C^g_b. (c) Pixel logarithmic-difference contrast $C^g_{\log}$. (d) Pixel depth difference p_d.*

(a) C^g

(b) C^g_b

(c) $C^g_{\log}$

(d) p_d

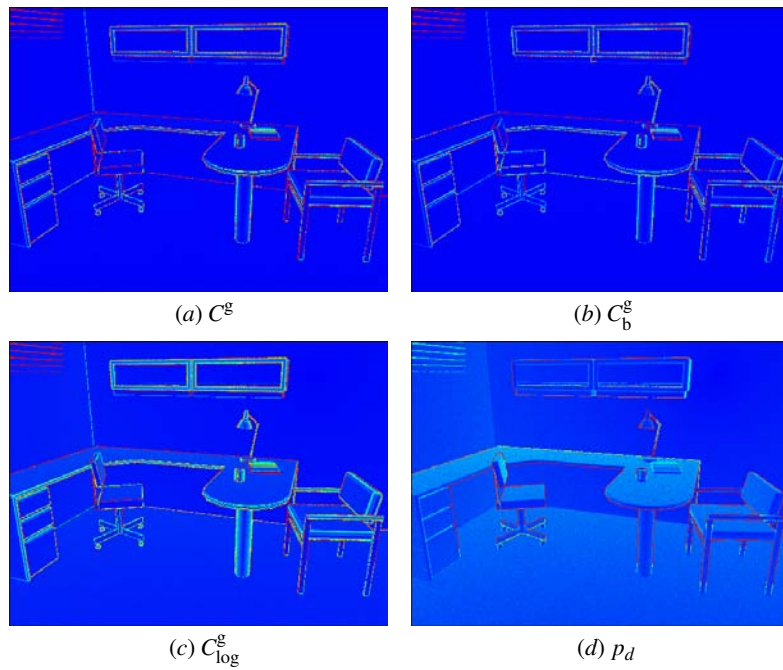**Figure 5:** *Geometry contrast maps from Fig. 3.a obtained with $N^p_s = 8$. For a visual comparison between the geometry and colour contrast in Fig. 3, the thermic scale is used. (a) Pixel geometry contrast C^g. (b) Pixel geometry binary contrast C^g_b. (c) Pixel logarithmic-difference contrast $C^g_{\log}$. (d) Pixel depth difference p_d.*
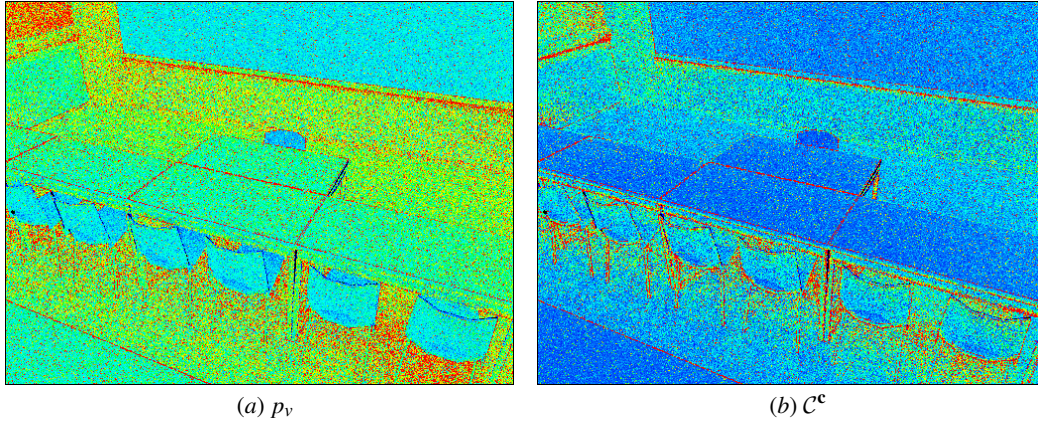*Credit: Modelled by Gregory J. Ward, Albany (CA), USA.*

(a) p_v (b) $\mathcal{C}^{\mathbf{c}}$

Figure 6: *Pixel contrast obtained with $N_s^{\mathsf{P}} = 4$ and a lineal combination with $\delta = 0.9$. (a) Pixel priority-value p_v (4). (b) Pixel contrast $\mathcal{C}^{\mathbf{c}}$ using $\mathbf{C}_b^{\mathbf{c}}$ and C^g.*
Credit: Model included in RenderPark [Com00], Computer Graphics Research Group, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium.

and penumbra, depth of field, motion blur, and translucency), the regions of the scene with the most complex illumination would need a more intensive treatment than a region with almost uniform illumination. This way of supersampling is called *adaptive sampling* [DW85, PS89]. A pixel is first sampled at a relatively low density. From this set of samples, 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. We can consider two kinds of adaptive sampling: first, when the refinement criterion plays the role of an oracle which decides the place and the quantity of supersampling necessary in one evaluation of the initial sampling only and, second, when the refinement criterion constantly evaluates the information received because of a supersampling and acts in consequence until it becomes satisfied. In this work we use the term *supersampling* exclusively for the first case and *adaptive sampling* for the second (§6).

We implement a simple supersampling technique: the sample set *S* will be proportionally distributed over the image plane with respect to the contrast $\mathcal{C}^{\mathbf{c}}$ estimated in each $\mathsf{p} \in \mathsf{P}$. This is equivalent to the use of the *pixel contrast* as an oracle. For definition of the measure itself (§4), the cost in samples is controlled by the diversity of colour and geometry in the pixel (i.e., low quality). Given that a high contrast is synonymous of low pixel quality and low contrast of high quality pixel, the measure adapts the densities of sampling to the necessity of improvement in pixel quality. Remembering the importance of each one of the samples and without more prior information, this system will improve, on average, the pixel quality in particular and the image in general. The generic procedure is made up of three sequentially quite different phases:

Oracle A pixel contrast is selected as oracle and a first estimate of actual contrast per pixel is obtained using an initial stratified sampling against the image plane. Usual values are 2, 4, and 8 (n_s^{P}). If we consider that the total number of samples N_s^{P} destined for the image is prefixed, the final value of the average of samples per pixel is $N_{\bar{s}}^{\mathsf{P}} = \frac{N_s^{\mathsf{P}}}{N_{\mathsf{p}}}$ and then, $1 < n_s^{\mathsf{P}} \ll \lfloor N_{\bar{s}}^{\mathsf{P}} \rfloor$ where the left side inequality is due to the definition of the measure of contrast and the right side inequality to being able to carry out supersampling. The result of this phase is the answer from the oracle: the contrast map.

Sampling The unused samples in the calculation of the contrast, $N_{\mathsf{p}}(N_{\bar{s}}^{\mathsf{P}} - n_s^{\mathsf{P}})$, are proportionally distributed to the values of contrast obtained per pixel. In each one of them, the sampling is also carried out with stratification. The distribution of the new set of samples gathers information from the scene in the regions of more diversity, with a proportional effort on this. The result of this phase is a supersampling directed exclusively by the contrast map generated in the previous phase.

Reconstruction The colour information gathered in the previous phase is put together with that obtained in the initial phase. Its evaluation allows us to achieve a more precise vision of the contents of the pixel and as a result an improvement in its quality. It only remains for the signal to be recuperated and to carry out the resampling process for each one of the pixels $\mathsf{p} \in \mathsf{P}$ with any of the reconstruction methods applicable to the sampling system used. The result of this phase is the solution for the image plane thanks to the assignment of the final colour to all of its pixels.

This proceeding is adaptable in any of its phases (e.g.,

stochastic ray-tracing method, pattern of initial sampling, supersampling method, and filters).

5.2. Results

Here, we show an example of our contrast measure $\mathcal{C}^{\mathbf{c}}$ (23) used as a supersampling oracle in path-tracing.

In Fig. 7.*a.i* we show a supersampling image obtained with $N_{\bar{s}}^{\mathsf{p}} = 32$ in the following way. First, a uniform stratified sampling with $n_s^{\mathsf{p}} = 8$ has been made in order to obtain the contrast map in Fig. 7.*a.ii*. Secondly, this map has been used in the supersampling process with an average of 24 rays per pixel. And thirdly, in order to analyse the behaviour of the contrast, the signal reconstruction in the last phase is carried out by a piece-wise continuous reconstruction using a box filter. The final pixel value corresponds to its signal average. The contrast measure used is a colour and geometry combination with $\delta = 0.5$ based on binary contrasts $\mathbf{C}_{\mathsf{b}}^{\mathbf{c}}$ (19) and $C_{\mathsf{b}}^{\mathsf{g}}$ (21). This means that the more critical the area, the more supersampled it is (warm colours), and the less critical, the more undersampled it is (cool colours, with a minimum of 8 rays per pixel). Two detailed regions are compared from the supersampling image (Figs. 7.*b–c.i*) and a similar image obtained by uniform stratified sampling with $N_s^{\mathsf{p}} = 32$ (Figs. 7.*b–c.ii*). We can observe a decrease in noise in the supersampled regions, and a better representation of shadow contour and edges.

6. Entropy-Based Adaptive Sampling

In (9) we have seen that image plane and pixel channel entropies are related thanks to the grouping property of entropy. It is important to note that this kind of decomposition can be applied recursively if the pixels are recursively subdivided. We will show in this section that this recursive decomposition provides us with a natural method of dealing with an adaptive sampling technique. Our scheme, valid for any pixel sampling and ray-tracing method, is applied to stochastic ray-tracing and compared with other options.

6.1. Adaptive Sampling

In order to obtain a realistic image, the aliasing has to be reduced by adapting the density sampling to the complexity of the region aiming at a good balance between cost and quality. We consider three phases in order to describe a generic process of adaptive sampling [Gla95] (Fig. 8), for which the scheme used in the supersampling procedure of §5.1 is a particular case:

Initial sampling An *initial sampling* pattern at a predetermined density is established. Normally, in order to choose its density we assume that the signal has a Nyquist rate similar to the frequency of the reconstruction samples (e.g., one sample per pixel (the minimum number of samples per pixel fulfils $N_s^{\mathsf{p}} > 1$ (§3)). It is also usual for this

density to be constant across the space, even though in the next phase, it should be increased locally if necessary.

Refinement tree The image space is divided up into regions (e.g., pixels). For each one of them, a *refinement test geometry* selects a subset of samples for evaluation. A *refinement test* is a criterion used for the evaluation of one or more characteristics which estimate the good quality of the current density. If the result is negative, a new set of samples are generated at the points indicated by the new *sampling geometry* and the process goes back to the refinement test geometry until the refinement criterion decides that the density of sampling in the region is accurate enough. The result of this process is a *refinement tree* of the image space where every node is a region with a density of sampling adapted to its own signal. In order to control extreme cases, it is usual to dispose of other criteria to finish the recursion (e.g., minimum area of the regions and/or maximum depth of tree).

Reconstruction The information of the signal collected at every region is unified by a *reconstruction* process and, if necessary, sent to a *filtering* process. Finally, a *resampling* process (e.g., centre of pixel) determines the final values for each of the pixels on the image plane.

Note that however much we increase the density of sampling locally, given that the signal is not usually band-limited, the sampling theory tell us that we can never capture it correctly. Thus, fine details of edges, shadings, textures, and others will hardly have enough quality in the final image. The approximation done by the method consists in looking for the minimum set of samples which estimates the signal locally accurately enough. A critical subproblem appears in each of its phases [PS89] and many approaches are found to deal with them (§2).

We focus our attention on obtaining an adaptive algorithm centred mainly on the refinement tree phase bringing a new perspective to the subproblem of controlling the sampling rate (new refinement criterion).

6.2. Recursive Entropy Tree

The natural way to represent information is by entropy, which in our context is interpreted as a measure of the degree of homogeneity of a region. Thus, using an entropy criterion means to evaluate the homogeneity (§3) or heterogeneity (§4) on a pixel. The fundamental idea behind our scheme is to capture the information in the refinement tree which results from the recursive decomposition of the entropy.

Generalising the grouping property of entropy, the entropy can be recursively decomposed in the following way: Let X be a discrete random variable over the set $\mathcal{X} = \{x_1, \dots, x_n\}$ with probability distribution $p = \{p_1, \dots, p_n\}$. Let us consider a partition of the set $\mathcal{X}$ in m-disjoint sets $\mathcal{Y} = \{\mathcal{Y}_1, \dots, \mathcal{Y}_m\}$ where $|\mathcal{Y}_j| = n_j$. Let us associate the discrete random variable Y to $\mathcal{Y}$ with probability distribution
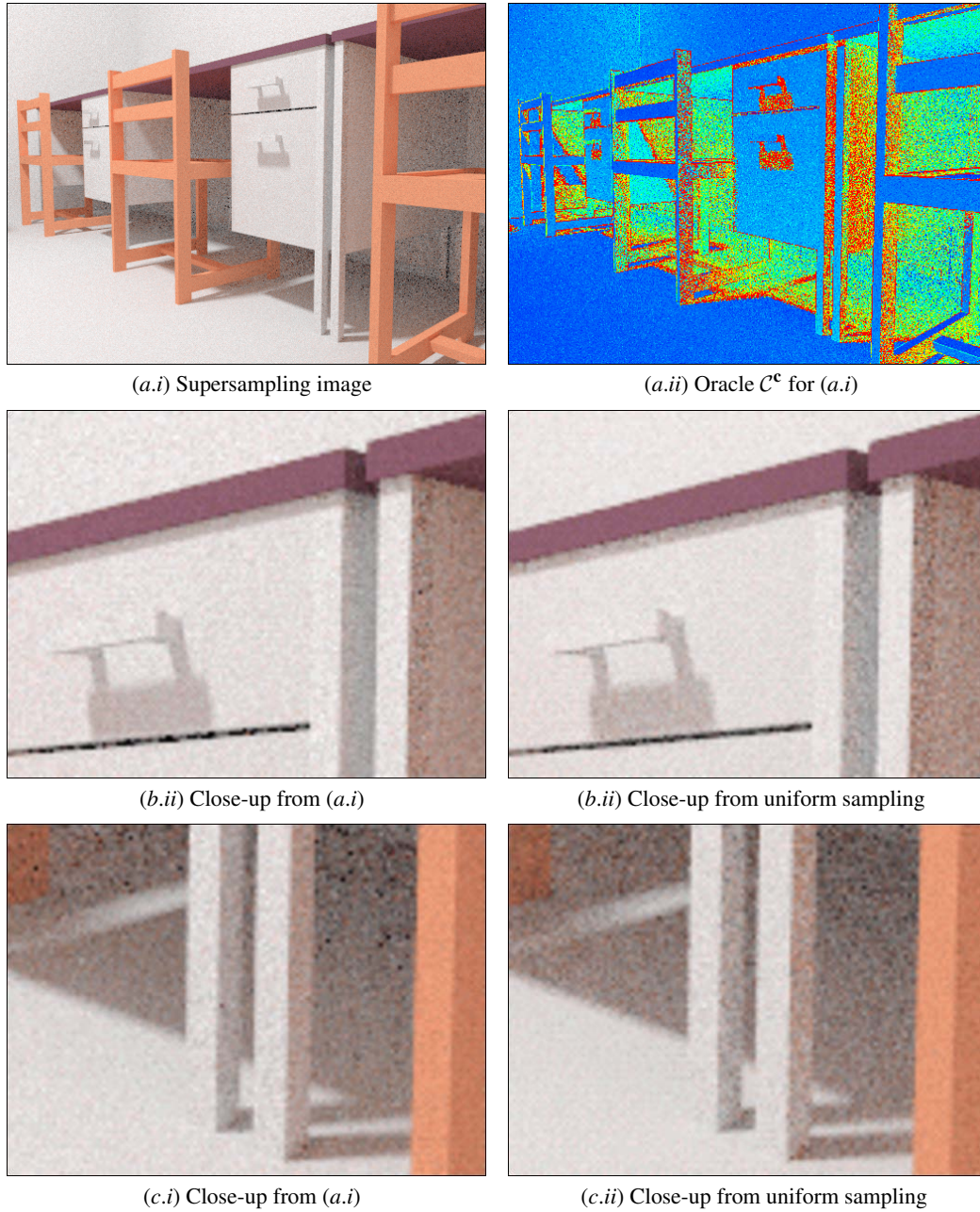
(*a.i*) Supersampling image

(*a.ii*) Oracle $\mathcal{C}^{\mathbf{c}}$ for (*a.i*)

(*b.ii*) Close-up from (*a.i*)

(*b.ii*) Close-up from uniform sampling

(*c.i*) Close-up from (*a.i*)

(*c.ii*) Close-up from uniform sampling

Figure 7: *Entropy-based supersampling versus uniform sampling. (a.i) Supersampling with an $N_{\bar{s}}^{\mathsf{p}} = 32$. (a.ii) Binary contrast map $\mathcal{C}^{\mathbf{c}}$ used as oracle to obtain (a.i). It has been calculated with $n_s^{\mathsf{p}} = 8$, $\mathbf{C}_b^{\mathbf{c}}$, C_b^g, and $\delta = 0.5$. Close-up details from (a.i) are shown in (b-c.i). They are compared with the same regions, (b-c.ii) respectively, taken from a uniform stratified sampling image with $N_s^{\mathsf{p}} = 32$.*
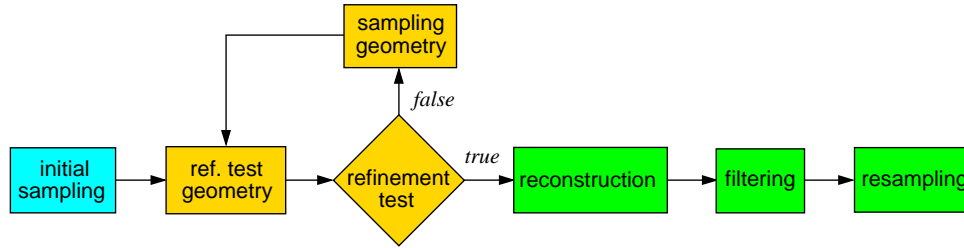
Figure 8: *Adaptive sampling process with three phases: initial sampling (blue), refinement tree (gold), and image reconstruction (green).*

$q = \{q_1, \ldots, q_m\}$ where $q_j = \sum_{k=1}^{n_j} p_{j_k}$ ($j_k \in \{1, \ldots, n\}$), and a new discrete random variable Y_j to each set $\mathcal{Y}_j$ with probability distribution $r_j = \{r_{j_1}, \ldots, r_{j_{n_j}}\}$ where $r_{j_k} = \frac{p_{j_k}}{q_j}$. Then

$$H(X) = \sum_{j=1}^{m} q_j H(Y_j) - \sum_{j=1}^{m} q_j \log q_j. \qquad (24)$$

This formula can be written as $H(X) = H_{\text{in}}(\mathcal{Y}) + H_{\text{out}}(\mathcal{Y})$ where $H_{\text{in}}(\mathcal{Y}) = \sum_{j=1}^{m} q_j H(Y_j)$ and $H_{\text{out}}(\mathcal{Y}) = H(Y) = -\sum_{j=1}^{m} q_j \log q_j$ represent, respectively, the hidden information (pending to be discovered) and the information already acquired in the descent of the tree created from an $\mathcal{Y}$ partition (Fig. 9).

In our case, (24) can also be interpreted taking into account only one colour channel (8) in the following way:

- $H(X)$ represents the entropy of the image plane.
- $H(Y_j)$ represents the entropy of each root pixel.
- Probability q_j is the proportion between the channel colour of pixel j and the sum of the channel colour of all pixels. It can be considered the "importance" of pixel j.

The decomposition of entropy can be recursively extended to the subpixels. This interpretation can also be applied to geometry entropy (14).

In our approach, probabilities are obtained by stochastic sampling. From the definition of entropy, we can see that when the number of samples tends to infinity, entropy also goes to infinity. In fact, we can consider that the original continuous scene contains infinite information. The following sampling algorithm will extract more information from the regions with more sample variation.

6.3. Algorithm

We present a new adaptive scheme for adaptive sampling, complementary to the entropy-based supersampling method (§5.1), with the important feature that it is based on the recursive expression of the Shannon entropy (i.e., the entropy tree). For the sake of simplicity, in the following analysis

we only consider the colour information of one channel, although in the final algorithm we will take the combination of colour and geometry contrasts into account (23).

A general description of our algorithm is as follows: On the image plane we sample each pixel to capture the colour of hitpoints and thus evaluate the information content (entropy) from the colour probability distribution. If the information of a pixel is high enough (i.e., the rays provide us with sufficient colour homogeneity on that pixel), refinement is not made, and the colour reconstruction of this pixel is done. When the information is not high enough, this pixel is subdivided into regions and we proceed in the same way for each region (subpixel). *"The approach will be to make sure that all the samples in a given region are similar in some specified way, so we can feel that we have captured what is happening in a region of the signal"* [Gla95].

This recursive process defines a tree with two well-separated phases for a pixel:

◦ *Pixel refinement.* Until enough information is extracted (tree descent).
◦ *Pixel colour.* Computation of the final colour (tree ascent).

The descent in the refinement tree can be interpreted as a progressive gain in information. The information acquired at each level is added together so that, at the end of the refinement process, the total information from the tree is the sum of the information obtained over all the branches (24). The measure used to capture the information will be the pixel contrast (§4).

Before introducing the algorithm we will give the definitions of the data used in it. Concerning the tree data structure, n represents the tree level where

◦ $n = 0$ is the image level (root).
◦ $n = 1$ is the pixel level (composed of N_{p} pixels of the image).
◦ $n > 1$ is the subpixel level.

We consider an n-node at any node of the tree with a level of $n > 0$ (i.e., no root). The set of data is described in Table 1. To compute the final colour of a pixel, we follow a path through the tree (Fig. 10). In the analysis below, we focus our
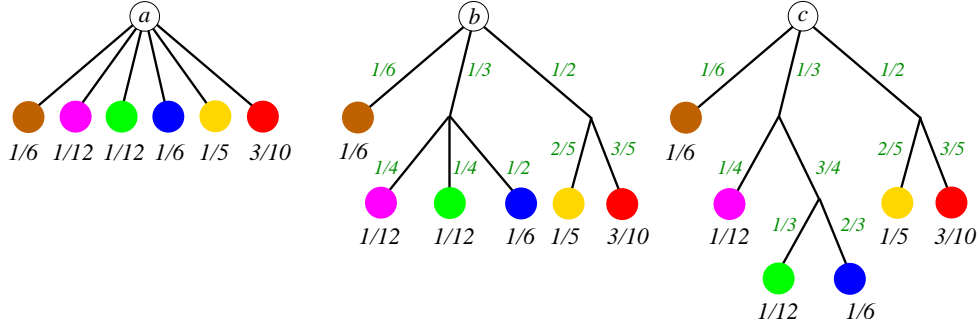
Figure 9: *Grouping property of entropy. The entropy of probability distributions of ⓐ is $H(\{\frac{1}{6}, \frac{1}{12}, \frac{1}{12}, \frac{1}{6}, \frac{1}{5}, \frac{3}{10}\})$, of ⓑ is $H(\{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\}) + \frac{1}{3}H(\{\frac{1}{4}, \frac{1}{4}, \frac{1}{2}\}) + \frac{1}{2}H(\{\frac{2}{5}, \frac{3}{5}\})$, and of ⓒ is $H(\{\frac{1}{6}, \frac{1}{3}, \frac{1}{2}\}) + \frac{1}{3}(H(\{\frac{1}{4}, \frac{3}{4}\}) + \frac{3}{4}H(\{\frac{1}{3}, \frac{2}{3}\})) + \frac{1}{2}H(\{\frac{2}{5}, \frac{3}{5}\})$. Accordingly to (24), they all have the same value: $H(a) = H(b) = H(c) = 2.445$.*

attention on the tree-path k of length m going from pixel k_0 to subpixel k_{m-1}. In this path, p_n represents the probability of the tree-branch at level n and q_n the *importance* of the n-node. In our algorithm, this quantity appears naturally due to recursive decomposition of the entropy (see (24), Fig. 9, and Fig. 10). The value of importance is given by the probability of the n-node:

$$q_n = \begin{cases} 1, & \text{if } n = 0, \\ p_0 \cdots p_{n-1} = \frac{\bar{c}_{0,k_0}}{\sum_{i \in R_0} \bar{c}_{0,i}} \prod_{\ell=1}^{n-1} p_\ell, & \text{if } n > 0. \end{cases} \quad (25)$$

For our purposes, q_n does not need to be normalised, thus we omit the normalisation constant $\sum_{i \in R_0} \bar{c}_{0,i}$ and we take $q_n = \bar{c}_{0,k_0} \prod_{\ell=1}^{n-1} p_\ell$.

Proposition 1 *The computation of q_n can be simplified to (in an abuse of notation, all the superindexes corresponding to arithmetic expressions must be interpreted as a power)*

$$q_n \approx \frac{\bar{c}_n}{N_r^{n-1}}. \quad (26)$$

Observe first that for a given path and $n > 0$, the colour $\bar{c}_n$ of an n-node is more accurate than the colour average of its respective region, k_{n-1}, in the preceding level. Thus, the accuracy of p_n, and at the same time of q_n, can be increased by substituting $\bar{c}_{n-1,k_{n-1}}$ for $\bar{c}_n$. Let us prove now (26) by induction.

Proof: For $n = 1$,

$$q_n = \bar{c}_{0,k_0} \approx \bar{c}_1 = \frac{\bar{c}_1}{N_r^0} = \frac{\bar{c}_n}{N_r^{n-1}}.$$

Hypothesis: $\forall_{0 < \ell < n}. q_\ell = \frac{\bar{c}_\ell}{N_r^{\ell-1}}$. Then, for $n > 1$

$$q_n = \bar{c}_{0,k_0} \prod_{\ell=1}^{n-1} p_\ell = q_{n-1} p_{n-1}$$

$$= \frac{\bar{c}_{n-1}}{N_r^{n-2}} \frac{\bar{c}_{n-1,k_{n-1}}}{\sum_{i \in R_{n-1}} \bar{c}_{n-1,i}}$$

$$\approx \frac{\bar{c}_{n-1}}{N_r^{n-2}} \frac{\bar{c}_n}{\bar{c}_{n-1} N_r} = \frac{\bar{c}_n}{N_r^{n-1}}. \quad \square$$

Now we can proceed to explain the algorithm. In the descent phase we sample an n-node and compute the contrast using expression $\mathcal{C}^{\mathbf{c}}$ (23). In (18) we must substitute the channel importance $\bar{c}$ by q_n and, according to §4, for a sRGB colour system we can take the perceptual coefficients $w^R = 0.2126$, $w^G = 0.7152$ and $w^B = 0.0722$ which capture the sensitivity of human colour perception [Com98].

Thus, for each n-node, the colour contrast (18) converts into

$$\mathbf{C}_n^{\mathbf{c}} = \sum_{c \in \mathbf{c}} w^c C_n^c q_n^c \quad (27)$$

and the colour and geometry combination (23) will be

$$\mathcal{C}_n^{\mathbf{c}} = \delta \mathbf{C}_n^{\mathbf{c}} + (1 - \delta) C_n^{\mathbf{g}}. \quad (28)$$

Note that this expression could also be calculated from the respective binary versions of colour and geometry contrasts (§4).

In the algorithm, we subdivide the pixel or subpixel when the contrast of an n-node is not less than a given threshold ($\mathcal{C}_n^{\mathbf{c}} \geq \varepsilon$). Thus, the phase of ascent begins when the test fails ($\mathcal{C}_n^{\mathbf{c}} < \varepsilon$). This happens because either the contrast (which represents the colour heterogeneity) or the importance ($q_n \rightarrow 0$ for growing n) are low. In this phase, each n-node in the path provides its colour estimation $\hat{c}_n$ from the signal reconstructed for each $c(s)$ where $s \in S_n$.

id	description	asserts
N_r	Number of regions in which an n-node can potentially be subdivided.	$N_r > 1 \wedge A_{1 \leq i \leq N_r} = \frac{A_{n-\text{node}}}{N_r}$
N_s^r	Number of samples of an n-node.	$N_s^r \geq N_r \wedge N_s^r \in N_r \mathbb{N}^+$
R_n	Set of regions of an n-node.	$|R_0| = N_p \wedge \forall_{n>0}.\, |R_n| = N_r$
S_n	Set of samples of an n-node.	$|S_0| = N_s^r N_p \wedge \forall_{n>0}.\, |S_n| = N_s^r$
$S_{n,i}$	Set of samples of an n-node region $i \in R_n$.	$S_n = \bigcup_{i \in R_n} S_{n,i}$ $|S_{n,i}| = \frac{|S_n|}{|R_n|} = \frac{N_s^r}{N_r}$
k	Path-tree $k = (k_0, \ldots, k_{m-1})$ where k_n is the region chosen at level n.	$m > 0 \wedge \forall_{n<m}.\, k_n \in R_n$
$\bar{c}_n$	Average colour channel data in an n-node.	$\bar{c}_n = \frac{1}{|S_n|} \sum_{s \in S_n} c(s)$
$\bar{c}_{n,i}$	Average colour in an n-node region $i \in R_n$.	$\bar{c}_{n,i} = \frac{1}{|S_{n,i}|} \sum_{s \in S_{n,i}} c(s)$ $\bar{c}_n = \frac{1}{|R_n|} \sum_{i \in R_n} \bar{c}_{n,i}$
p_n	Probability of region k_n of an n-node in a path k.	$p_n = \frac{\sum_{s \in S_{n,k_n}} c(s)}{\sum_{s \in S_n} c(s)} = \frac{\bar{c}_{n,k_n}}{\sum_{i \in R_n} \bar{c}_{n,i}}$
q_n	Probability of an n-node in a path k.	$q_n = \prod_{\ell=0}^{n-1} p_\ell$

Table 1: *Description of the data set of the refinement phase of entropy-based adaptive sampling in an image plane of N_p pixels. An n-node is a node of level $n > 0$ in the refinement-tree.*
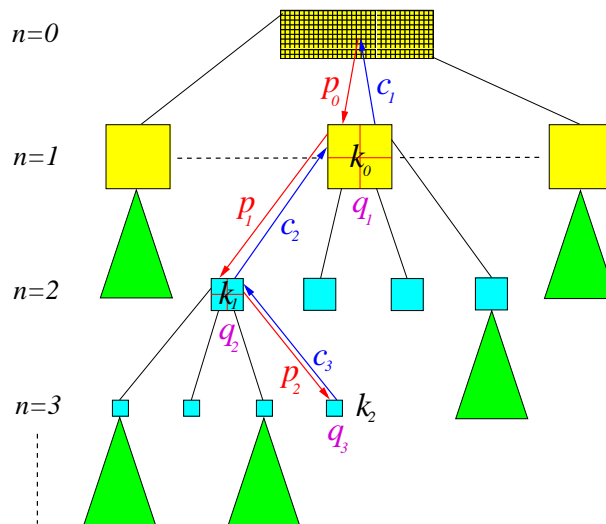

Figure 10: *A refinement-tree-path $k = (k_0, k_1, k_2)$ of length $m = 3$ of entropy-based sampling. The number of regions of an n-node is $N_r = 4$. We show the computation of the k_0-pixel colour: $c_{0,k_0} = c_1$ from the refinement (red) and reconstruction (blue) phases. The probabilities p_n and importances q_n are computed in the refinement phase to evaluate the entropy contrast (23).*

The final colour of an n-node is given by

$$c_n = \begin{cases} \widehat{c}_n, & \text{if } \mathcal{C}_n^{\mathbf{c}} < \varepsilon, \\ \sum_{i \in R_n} c_{n,i}, & \text{otherwise,} \end{cases} \quad (29)$$

where $c_{n,i}$ is the final colour of i-region of the n-node. Finally, we get c_1 for the colour of the pixels (or equivalently c_{0,k_0} in the path considered). An example of this process is shown in Fig. 10.

Observe that importance sampling is naturally integrated in the algorithm. Following importance sampling criteria a function should be sampled proportionally to its value which is what we obtain with our adaptive descent.

6.4. Results

For the purpose of comparison, in Fig. 12 we present the results for different techniques for the test scene in Fig. 11. We compare the following methods:

- *Classic contrast* (CC): A recursive adaptive sampling scheme based on contrast by channel (1) (with thresholds proportional to the visual system) weighted by its respective channel colour average [Gla95, SS00]. The maximum recursive level has been limited to 4 (Fig. 12.*a*).
- *Importance-weighted contrast* (IC): The same as in CC but each channel contrast is weighted with the respective importance q (26), as in our approach (Fig. 12.*b*).
- *Confidence test* (CT): Statistical approach based on a confidence interval (5) with a confidence level of $\alpha = 0.1$ and a tolerance $t = 0.025$ (Fig. 12.*c*).
- *Entropy-based contrast* (EC): Our approach (29) taking only colour contrast, $\delta = 1$ in (28) (Fig. 12.*d*).

Observe that the EC approach can be easily implemented on any standard hierarchical algorithm, using importance (26) and the new refinement criterion (28), with negligible additional cost.

In CC, IC, and EC, the number of subdivisions is $N_r = 4$ and the number of samples is $N_s^r = 8$. To compute the contrast measures for the refinement decision, the samples have been cast in a stratified way at each n-node (i.e., pixel or subpixel) and re-used at the next levels in the tree. In CT, groups of 8 samples were added in a stratified way until meeting the condition of the criterion. 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 ($N_s^p = 60$) and computational cost. The resulting images are shown in Figs. 12.*.*.i with close-ups in Figs. 12.*.*.ii. A sampling density map (SDM) for each one is given in Figs. 12.*.*.iii (generated under the same conditions as the quality (§3) and contrast (§4) maps; warm colours correspond to the highest sampling rate and cold colours to the lowest)

The overall aspect of the images in Figs. 12.*.*.i shows that our supersampling scheme performs best. Observe, for instance, the reduced noise in the shadows cast by the objects. This is further checked in the close-up images in Figs. 12.*.*.ii. Observe also the detail of the sphere shadow reflected on the pyramid. It is important to note that we managed to improve the classic contrast approach in CC greatly by including the importance used in our scheme (compare results in Fig. 12.*a* with Fig. 12.*b*). A comparison of the SDMs shows a better discrimination of complex regions of the scene in the entropy case against the classic contrast case. This explains the better results obtained by our approach. Moreover, the confidence test approach CT (Fig. 12.*c*) also performs better than the classic contrast-based methods CC (Fig. 12.*a*) and IC (Fig. 12.*b*). The SDM of CT also explains why it performs better. However, it is unable to render the reflected shadows under the mirrored pyramid and sphere with precision (see close-up in Fig. 12.*c*.ii).

In Table 2 we show two measures (error and quality) obtained in Figs. 12.*.*.i–ii with respect to the test scene in Figs. 11.*i–ii*, respectively. We select the root of the mean square error (it is calculated from the MSE of each colour channel) (RMSE) and the peak signal to noise ratio (PSNR) to evaluate the results (measure of the quality of a reconstructed image compared with an original image computing the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation). Because many signals have a very wide dynamic range, PSNR is usually expressed in terms of the logarithmic decibel scale: $10\log_{10}(I_{\max}^2/MSE)$ dB. For each one, we consider a weight balanced by every colour channel (RMSE$_a$ and PSNR$_a$) and a perceptual one (RMSE$_p$ and PSNR$_p$) in accordance with the sRGB system ($w^R = 0.2126$, $w^G = 0.7152$, and $w^B = 0.0722$. These measures reflect the good behaviour mentioned in CT and EC oracles (i.e., low RMSEs and high PSNRs). Although the error obtained using our approach is bigger than that with CT method, the visual results are better in the EC case (observe Figs. 12.*c–d*). This is due to the fact that the measures do not manage to reflect exactly the perceptual quality of the image. The EC images look better because the oracle distributes the samples in the perceptual critical regions more accurately (see SDMs).

Now, we present a test using the geometry component with $1 - \delta = 0.1$ in (28) and, at the same time, the binary contrast in colour and geometry. To do this, our approach is compared with the priority-value combination (4) made up of a colour contrast of the CC type and also by an usual geometry factor. Perceptual coefficients are taken equal as in our approach in both cases (§6.3). The tree depth level is set to 4 and the N_s^r is reduced by half (i.e., 4 samples) but maintaining the average per pixel (i.e., $N_s^p = 60$).

The images obtained are shown in Fig. 13. In Fig. 13.*a*, the entropy-contrast $\mathcal{C}_n^{\mathbf{c}}$ (28) with $\mathbf{C}^{\mathbf{c}}$ (18) and C^g (20). In Fig. 13.*b*, the binary-entropy contrast: $\mathcal{C}_n^{\mathbf{c}}$ using $\mathbf{C}_b^{\mathbf{c}}$ (19)

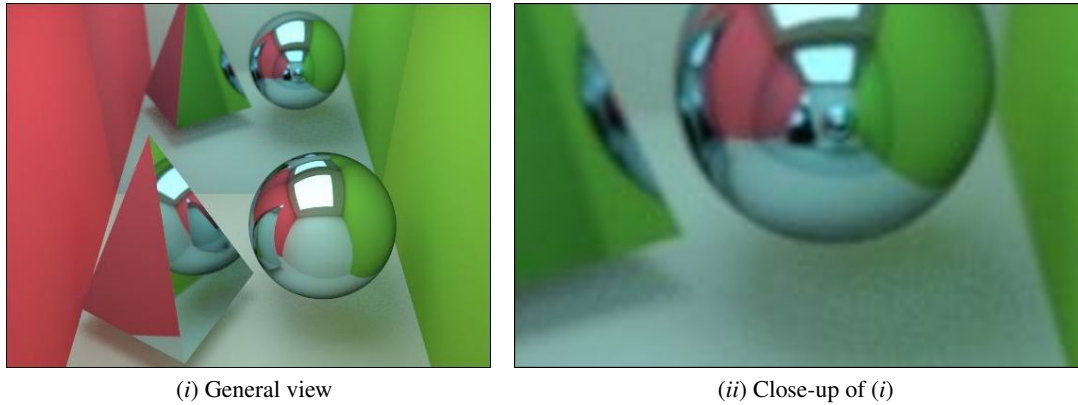| (*i*) General view | (*ii*) Close-up of (*i*) |

Figure 11: *Reference image used in the test in Fig. 12: (i) general view and (ii) close-up of (i). The image has been obtained with a path-tracing algorithm with 1,024 samples per pixel in a stratified way.*

oracle	general view				close-up			
	RMSE$_a$	RMSE$_p$	PSNR$_a$	PSNR$_p$	RMSE$_a$	RMSE$_p$	PSNR$_a$	PSNR$_p$
CC	13,727	13,599	25,379	25,461	20,276	20,024	21,991	22,100
IC	8,124	8,110	29,935	29,951	13,751	13,568	25,364	25,481
CT	5,194	5,174	33,822	33,855	8,407	8,338	29,638	29,710
EC	6,937	7,018	31,308	31,207	9,886	9,933	28,231	28,189

Table 2: *The RMSE and PSNR of CC, IC, CT, and EC oracles applied to the general view (Fig. 11.i) and close-up (Fig. 11.ii) of the test scene. The average number of rays per pixel is $N_{\bar{s}}^{p} = 60$ in all methods.*

and C_b^g (21). And, in Fig. 13.*c*, the priority-value approach p_v (4) with p_c (2) and p_d (3). The respective SDMs from Figs. 13.∗.*i* are shown in Figs. 13.∗.*ii*.

We see from comparing the images that the entropy contrast is much better than the classic contrast used here. Observe for instance the ceiling, the shadows and the mirroring wall. A drawback of our approach is the peaks of high radiance that we observe on the right wall because this region is undersampled in our method. However, this effect can be easily solved by a filtering technique. The comparison of the SDMs shows a better discrimination of complex regions of the scene in the entropy case (Figs. 13.*a–b.ii*) against the priority-value contrast case (Fig. 13.*c.ii*). This explains the better results obtained with our approach.

Finally, in Fig. 14.*a* we show another scene obtained with our approach using an average of $N_{\bar{s}}^{p} = 200$ and $\delta = 0.95$. Observe, in Fig. 14.*b*, how well the SDM works out both the geometry and colour details as in the shadow contours on the walls (the remaining spiked noise could easily be eliminated by filtering with an image smoothing method).

References

[BM98] BOLIN M. R., MEYER G. W.: A perceptually based adaptive sampling algorithm. In *SIGGRAPH '98 Conference Proceedings* (New York (NY), USA, July 1998), Cohen M., (Ed.), Annual Conference Series, ACM SIGGRAPH, ACM Press, pp. 299–309.

[Cae81] CAELLI T. M.: *Visual Perception: Theory and Practice*. Pergamon Press, Oxford, UK, 1981.

[Com98] COMMISSION I. E.: Default RGB colour space - sRGB. In *Colour Management in Multimedia Systems*. IEC, Geneva, Switzerland, 1998, ch. Part 2.1. IEC/PT61966(PL)34.

[Com00] COMPUTER GRAPHICS RESEARCH GROUP: *RenderPark: A Photorealistic Rendering Tool*. Software, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium, 2000. http://www.renderpark.be.

[CPC84] COOK R. L., PORTER T., CARPENTER L.: Distributed ray tracing. *Computer Graphics (Proceedings of SIGGRAPH '84) 18*, 3 (July 1984), 137–145.

[DC96] DARSA L., COSTA B.: Multi-resolution representation and reconstruction of adaptively sampled images. In *Proceedings of IX Brazilian Symposium on Computer Graphics and Image Processing (SIBGRAPI '96)* (October 1996), pp. 321–328.

[DCV97] DARSA L., COSTA B., VARSHNEY A.: Navigating static environments using image-space simplification and morphing. In *1997 Symposium on Interactive*
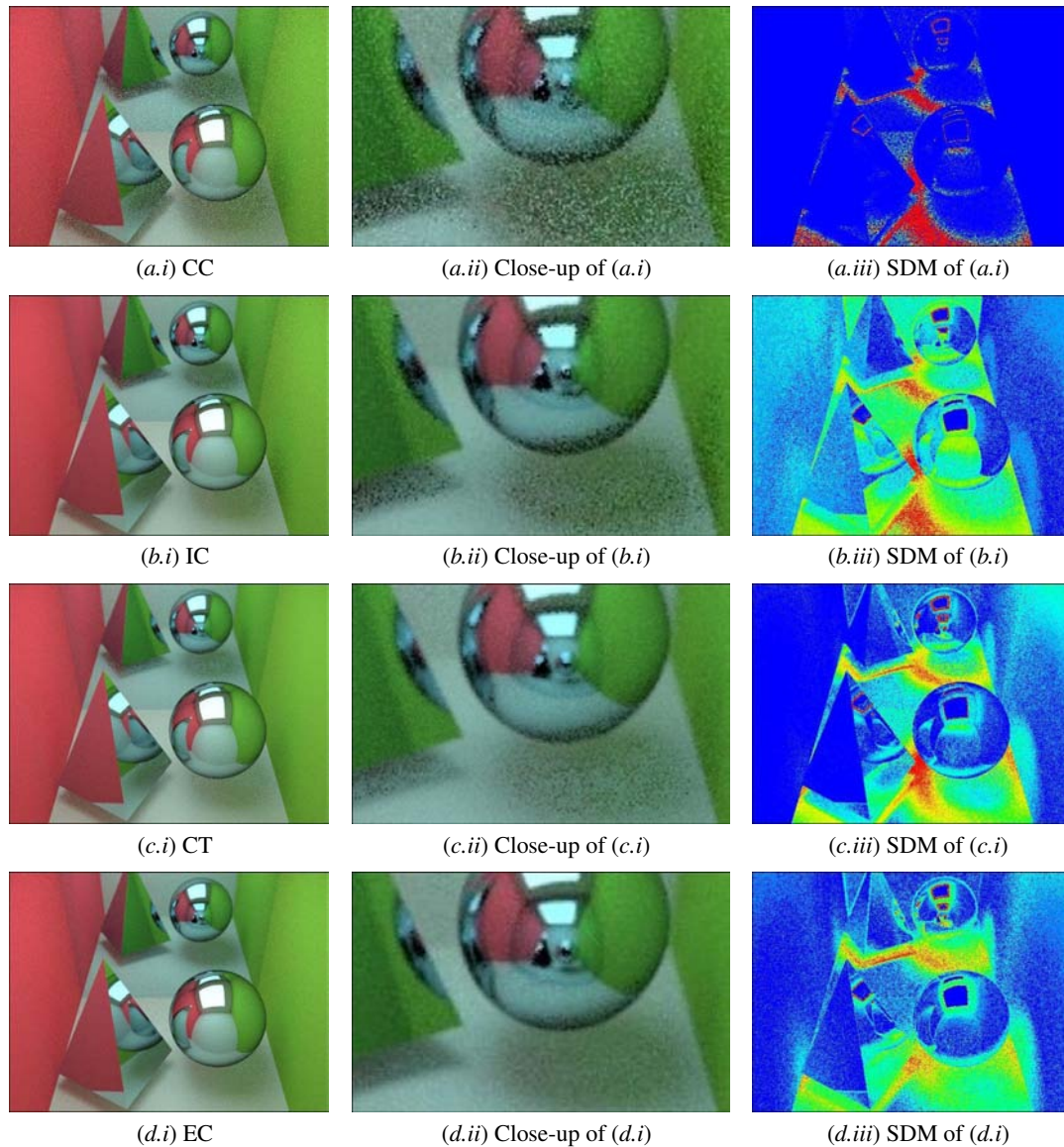
(*a.i*) CC   (*a.ii*) Close-up of (*a.i*)   (*a.iii*) SDM of (*a.i*)

(*b.i*) IC   (*b.ii*) Close-up of (*b.i*)   (*b.iii*) SDM of (*b.i*)

(*c.i*) CT   (*c.ii*) Close-up of (*c.i*)   (*c.iii*) SDM of (*c.i*)

(*d.i*) EC   (*d.ii*) Close-up of (*d.i*)   (*d.iii*) SDM of (*d.i*)

Figure 12: *Results of comparisons: (a) adaptive sampling scheme based on classic contrast (CC), (b) importance-weighted contrast (IC), same as in (a) but weighting with importance q (26), (c) confidence test method (CT), and (d) entropy-based method (EC) with colour contrast only (δ = 1). By columns: (i) shows the resulting images, (ii) close-up of regions of (i), and (iii) the sampling density maps of (i). The average number of rays per pixel is $N_{\bar{s}}^{\mathrm{p}} = 60$ in all methods, with a similar computational cost.*
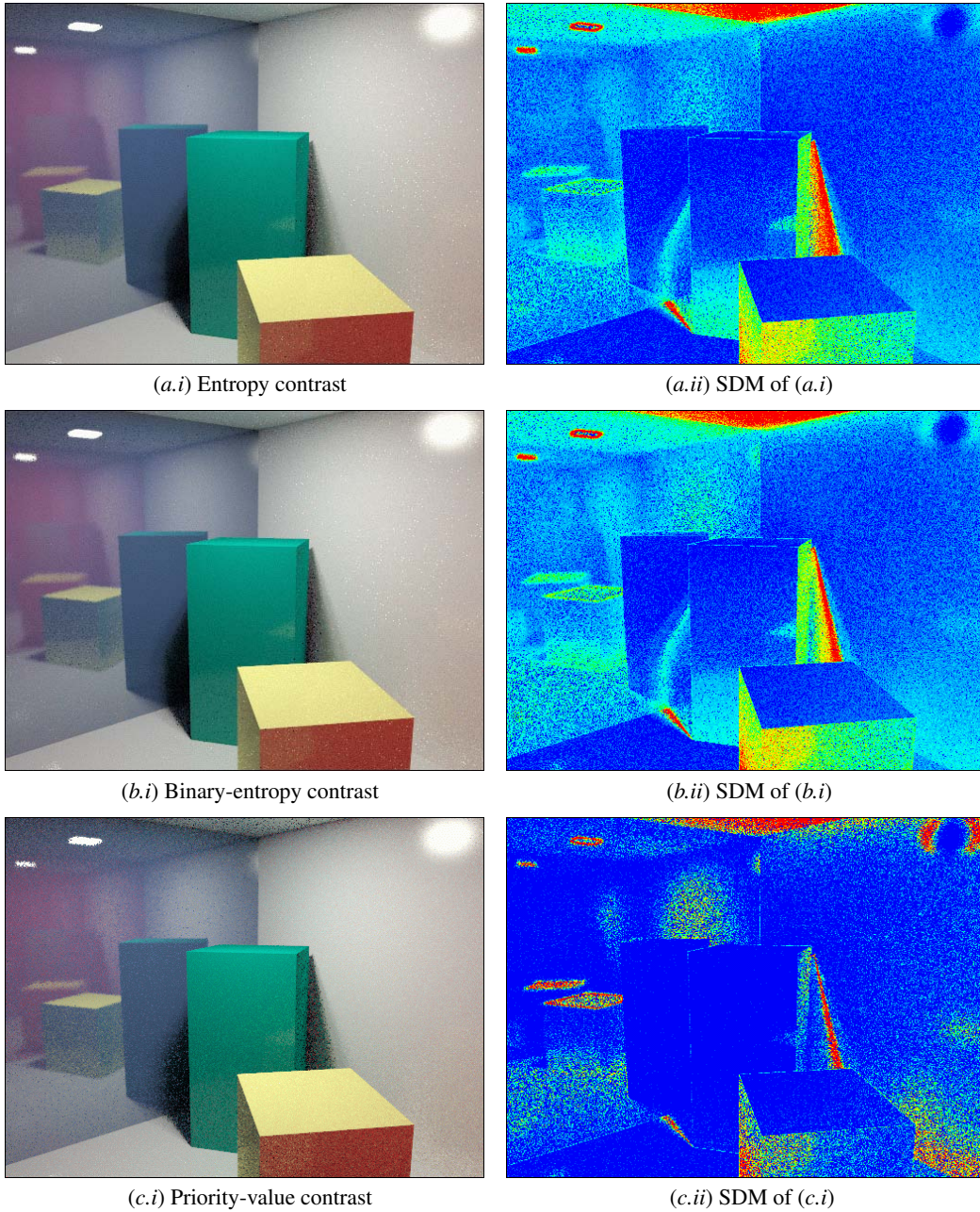
(*a.i*) Entropy contrast

(*a.ii*) SDM of (*a.i*)

(*b.i*) Binary-entropy contrast

(*b.ii*) SDM of (*b.i*)

(*c.i*) Priority-value contrast

(*c.ii*) SDM of (*c.i*)

Figure 13: *Images obtained with adaptive sampling where* $N_{\bar{s}}^{\mathrm{p}} = 60$ *and* $N_s^r = 4$*: (a) entropy contrast, (b) binary-entropy contrast, and (c) priority-value contrast. By columns: (i) Image sampled and (ii) sampling density map of (i).*
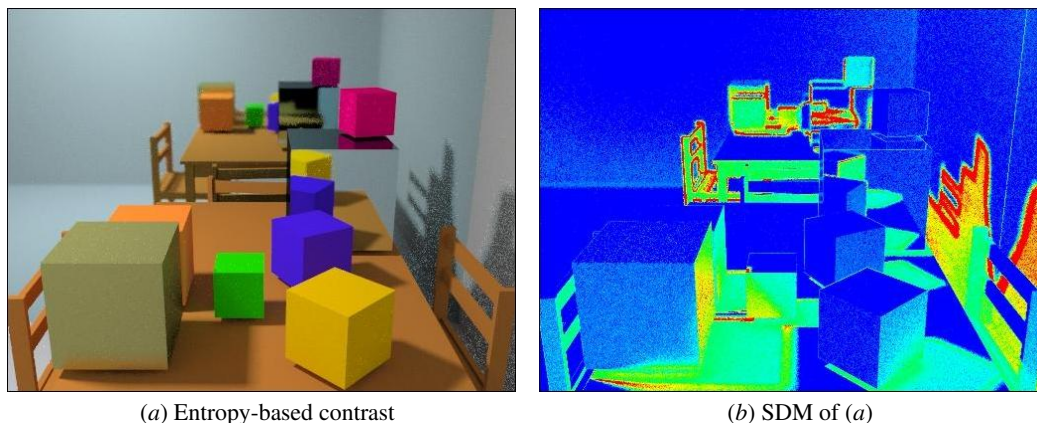
(*a*) Entropy-based contrast

(*b*) SDM of (*a*)

Figure 14: *Image obtained with entropy-based adaptive sampling where $N_{\bar{s}}^{\mathrm{p}} = 200$ and $\delta = 0.95$. (a) Sampled image. (b) Sampling density map of (a).*

3D Graphics (April 1997), Cohen M., Zeltzer D., (Eds.), ACM SIGGRAPH, pp. 25–34.

[DW85] DIPPÉ M. A. Z., WOLD E. H.: Antialiasing through stochastic sampling. *Computer Graphics (Proceedings of SIGGRAPH '85) 19*, 3 (July 1985), 69–78.

[Gla95] GLASSNER A. S.: *Principles of Digital Image Synthesis*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1995.

[KA91] KIRK D., ARVO J.: Unbiased variance reduction for global illumination. In *Proceedings of the 2nd Eurographics Workshop on Rendering* (May 1991), pp. 153–156.

[Kaj86] KAJIYA J. T.: The rendering equation. *Computer Graphics (Proceedings of SIGGRAPH '86) 20*, 4 (August 1986), 143–150.

[KH94] KELLER A., HEINRICH S.: *Quasi-Monte Carlo Methods in Computer Graphics, Part I: The QMC-Buffer*. Tech. Rep. 242/94, University of Kaiserslautern, Kaiserslautern, Germany, 1994.

[LRU85] LEE M. E., REDNER R. A., USELTON S. P.: Statiscally optimized sampling for distributed ray tracing. *Computer Graphics (Proceedings of SIGGRAPH '85) 19*, 3 (July 1985), 61–67.

[MF92] MCCOOL M., FIUME E.: Hierarchical Poisson disk sampling distributions. In *Proceedings of Graphics Interface '92* (May 1992), pp. 94–105.

[Mit87] MITCHELL D. P.: Generating antialiased images at low sampling densities. *Computer Graphics (Proceedings of SIGGRAPH '87) 21*, 4 (July 1987), 65–72.

[OA96] OHBUCHI R., AONO M.: *Quasi–Monte Carlo Rendering with Adaptive Sampling*. Tech. rep., Tokyo Research Laboratory, IBM Japan Ltd., Tokyo, Japan, 1996.

[PCD*97] PULLI K., COHEN M. F., DUCHAMP T.,

HOPPE H., SHAPIRO L., STUETZLE W.: View-based rendering: Visualizing real objects from scanned range and color data. In *Rendering Techniques '97 (Proceedings of the 8th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1997), Dorsey J., Slusallek P., (Eds.), Springer-Verlag, pp. 23–34.

[PS89] PAINTER J., SLOAN K.: Antialiased ray tracing by adaptive progressive refinement. *Computer Graphics (Proceedings of SIGGRAPH '89) 23*, 3 (July 1989), 281–288.

[Pur86] PURGATHOFER W.: A statistical method for adaptive stochastic sampling. *Eurographics '86: Proceedings of the European Conference and Exhibition 11*, 2 (August 1986), 157–162.

[RFBS01] RIGAU J., FEIXAS M., BEKAERT P., SBERT M.: View-dependent information theory measures for pixel sampling and scene discretization in flatland. In *Proceedings of Spring Conference on Computer Graphics '01* (Los Alamitos (CA), USA, April 2001), IEEE Computer Society, pp. 173–180.

[Sch91] SCHLICK C.: An adaptive sampling technique for multidimensional ray tracing. In *Proceedings of the 2nd Eurographics Workshop on Rendering* (May 1991), pp. 48–56.

[Shi90] SHIRLEY P.: A ray tracing method for illumination calculation in diffuse-specular scenes. In *Proceedings of Graphics Interface '90* (Toronto (Ontario), Canada, May 1990), Canadian Information Processing Society, pp. 205–212.

[SS00] SIMMONS M., SÉQUIN C. H.: Tapestry: A dynamic mesh-based display representation for interactive rendering. In *Rendering Techniques 2000 (Proceedings of the 11th Eurographics Workshop on Rendering)* (New

York (NY), USA, June 2000), Péroche B., Rushmeier H., (Eds.), Springer-Verlag, pp. 329–340.

[TJ97]   TAMSTORF R., JENSEN H. W.: Adaptive sampling and bias estimation in path tracing. In *Rendering Techniques '97 (Proceedings of the 8th Eurographics Workshop on Rendering)* (New York (NY), USA, June 1997), Dorsey J., Slusallek P., (Eds.), Springer-Verlag, pp. 285–295.

Applications of Information Theory to Computer Graphics
Part V: Viewpoint Selection and Mesh Saliency

Miquel Feixas, Mateu Sbert, and Francisco González García

University of Girona, Spain

1. Introduction

In computer graphics, several *viewpoint quality* measures have been applied in areas such as scene understanding [PB96, VFSH01, PPB*05], scene exploration [AVF04, SPT06], image-based modeling [VFSH03], and volume visualization [BS05, TFTN05, VFSG06, JS06]. In other areas, such as object recognition and mobile robotics, best view selection is also a fundamental task. Many works have demonstrated that the recognition process is view-dependent [PRC81, BET95, TBZB97, BTB99]. In [TBZB97], the authors found that "visual recognition may be explained by a view-based theory in which viewpoint-specific representations encode both quantitative and qualitative features". In robotics, the simultaneous localization and mapping problem (SLAM) requires that the robot decides on its own the necessary motions to construct the most accurate map possible. In [GBL02], an algorithm is proposed to guide the robot through a series of good positions, where 'good' refers to the expected amount and quality of the information that will be revealed at each new location.

The basic question underlying the viewpoint selection study and application is "what is a 'good' scene viewpoint?" Obviously, this question does not have a unique answer. Depending on our objective, the best viewpoint can be, for instance, the most *representative* one or the most *unstable* one, i.e., the one that maximally changes when it is moved within its close neighborhood [BS05]. [PRC81] and [BTB99] have presented different experiments demonstrating that observers prefer views (called *canonical views*) that avoid occlusions and that are off-axis (such as a three-quarter viewpoint), salient (the most significant characteristics of an object are visible), stable and with a large number of visible surfaces.

Extending the work initiated in [VFSH01, SPFG05], in this tutorial we will present a unified and robust framework to deal with viewpoint selection and mesh saliency. Given a set of viewpoints surrounding the object, we define an *in-*

formation channel between the viewpoints and the polygons of the object. From this channel, the *viewpoint mutual information* is used to obtain the best views of an object, to calculate the stability of a viewpoint, and to guide the object exploration. Then, we reverse the channel and we compute both the information and the saliency associated with each polygon. Finally, this polygonal saliency is used to calculate how salient is a viewpoint and is incorporated to viewpoint mutual information to drive the viewpoint selection. Our framework is also applicable to any set of viewpoints in a closed scene and, although only the geometric properties of an object have been considered, other aspects such as lighting could be incorporated.

In these notes, we only present some related work (Section 2) and the viewpoint channel (Section 3). Other aspects of the viewpoint framework (viewpoint similarity and unstability, object exploration, saliency, information-theoretic ambient occlusion, best view selection, etc.) can be found in the slides corresponding to this Part V.

2. Related Work

We review now some viewpoint quality measures for polygonal models. In [PB96], the quality of a viewpoint v of a scene is computed using the *heuristic measure* (HM) given by

$$C(v) = \frac{\sum_{i=1}^{n} \left\lceil \frac{P_i(v)}{P_i(v)+1} \right\rceil}{n} + \frac{\sum_{i=1}^{n} P_i(v)}{r} \quad , \qquad (1)$$

where $P_i(v)$ is the number of pixels corresponding to the polygon i in the image obtained from the viewpoint v, r is the total number of pixels of the image (resolution of the image), and n is the total number of polygons of the scene. In this formula, $\lceil x \rceil$ denotes the smallest integer, greater than or equal to x. The first term in (1) gives the fraction of visible surfaces with respect to the total number of surfaces, while the second term is the ratio between the projected area of the

scene (or object) and the screen area (thus, its value is 1 for a closed scene).

From the definition of entropy (see Part II), the *viewpoint entropy* (VE) [VFSH01] has been defined from the relative area of the projected polygons over the sphere of directions centered at viewpoint v. Thus, the viewpoint entropy was defined by

$$H_v = -\sum_{i=0}^{N_f} \frac{a_i}{a_t} \log \frac{a_i}{a_t}, \qquad (2)$$

where N_f is the number of polygons of the scene, a_i is the projected area of polygon i over the sphere, a_0 represents the projected area of background in open scenes, and $a_t = \sum_{i=0}^{N_f} a_i$ is the total area of the sphere. The maximum entropy is obtained when a certain viewpoint can see all the polygons with the same projected area. The best viewpoint is defined as the one that has maximum entropy. In molecular visualization, both maximum and minimum entropy views show relevant characteristics of a molecule [VFSL06].

A new viewpoint quality measure, called *viewpoint Kullback-Leibler distance* (VKL) [SPFG05], has been defined by

$$KL_v = \sum_{i=1}^{N_f} \frac{a_i}{a_t} \log \frac{\frac{a_i}{a_t}}{\frac{A_i}{A_T}}, \qquad (3)$$

where a_i is the projected area of polygon i, $a_t = \sum_{i=1}^{N_f} a_i$, A_i is the actual area of polygon i and $A_T = \sum_{i=1}^{N_f} A_i$ is the total area of the scene or object. The VKL measure is interpreted as the distance between the normalized distribution of projected areas and the 'ideal' projection, given by the normalized distribution of the actual areas. In this case, the background can not be taken into account. The minimum value 0 is obtained when the normalized distribution of projected areas is equal to the normalized distribution of actual areas. Thus, to select views of high quality means to minimize KL_v.

Apart from the previous references on viewpoint quality measures, [PPB*05] describe a number of different ways to measure the goodness of a view of an object. After analyzing different view descriptors, they conclude that no single descriptor does a perfect job and possibly a combination of them would amplify the advantage that each one has. Given a sphere of viewpoints, [YSY*06] compute the similarity between each two disjoint views using Zernike moments analysis and obtain a similarity weighted spherical graph. A view is considered to be stable if all edges incident on its viewpoint in the spherical graph have high similarity weights. [AVF04] and [SPT06] present two different exploration algorithms guided by viewpoint entropy and the total curvature of a visible surface, respectively. In the volume rendering field, [BS05], [TFTN05] and [JS06] use an extended version of viewpoint entropy and [VFSG06] introduce the viewpoint mutual information. [CSCF07] use view-

point entropy as a perceptual measure for mesh simplification.

Based on the investigation on canonical views, [GRMS01] present a new method for constructing images, where the viewpoint is chosen to be both off-axis and 'natural', and [LME06] obtain the viewing direction from the combination of factors such as saliency, occlusion, stability and familiarity. [LVJ05] have introduced the saliency as a measure for regional importance for graphics meshes and [KV06] presented a visual-saliency-based operator to enhance selected regions of a volume. [GCO06] introduced a method for partial matching of surfaces by using the abstraction of salient geometric features and a method to construct them.

3. Viewpoint Channel

In this section, we introduce an information channel between a set of viewpoints and the set of polygons of an object to deal with viewpoint selection. Then we define the viewpoint mutual information to select the most representative views of an object.

3.1. Viewpoint Mutual Information

Our viewpoint selection framework is constructed from an information channel $V \rightarrow O$ between the random variables V (input) and O (output), which represent, respectively, a set of viewpoints and the set of polygons of an object (see Figure 1(a)). This channel, which we call *viewpoint channel*, is defined by a conditional probability matrix obtained from the projected areas of polygons at each viewpoint. Viewpoints will be indexed by v and polygons by o. Throughout the development of this viewpoint framework, the capital letters V and O as arguments of $p()$ will be used to denote probability distributions. For instance, while $p(v)$ will denote the probability of a single viewpoint v, $p(V)$ will represent the input distribution of the set of viewpoints.

The viewpoint channel can be interpreted as an *observation channel* where the conditional probabilities represent the probability of *seeing* a determined polygon from a given viewpoint (see Figure 1(b)). The three basic elements of this channel are:

- Conditional probability matrix $p(O|V)$, where each element $p(o|v) = \frac{a_o}{a_t}$ is defined by the normalized projected area of polygon o over the sphere of directions centered at viewpoint v. Conditional probabilities fulfil $\sum_{o \in \mathcal{O}} p(o|v) = 1$. In our approach, background is not taken into account but it could be considered as another polygon.
- Input distribution $p(V)$, which represents the probability of selecting a viewpoint. In our experiments, $p(V)$ will be obtained from the normalization of the projected area of the object at each viewpoint. This can be interpreted as the

probability that a random ray originated at v hits (*sees*) the object. This assignment is consistent with the objective of selecting the viewpoints which *see* more projected area. Let us remember that this is a characteristic of a canonical view (see Section 1). The input distribution can also be interpreted as the *importance* assigned to each viewpoint v. For instance, the input distribution could also be defined by $p(v) = \frac{1}{N_v}$, where N_v is the number of viewpoints.

- Output distribution $p(O)$, defined by

$$p(o) = \sum_{v \in \mathcal{V}} p(v)p(o|v), \qquad (4)$$

which represents the average projected area of polygon o, i.e., the probability of polygon o to be hit (*seen*) by a random ray cast from the viewpoint sphere.

From the previous definitions, the *conditional entropy* is given by the average of all viewpoint entropies:

$$
\begin{aligned}
H(O|V) &= -\sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log p(o|v) \\
&= \sum_{v \in \mathcal{V}} p(v)H(O|v), \qquad (5)
\end{aligned}
$$

where $H(O|v) = -\sum_{o \in \mathcal{O}} p(o|v) \log p(o|v)$ is the *viewpoint entropy* H_v (2) and measures the degree of uniformity of the projected area distribution at viewpoint v. Let us observe that H_v has been now rewritten in a different form. Both entropies $H(O|v)$ and $H(O|V)$ tend to infinity when polygons are infinitely refined. This makes these measures very sensitive to the discretisation of the object and in general not appropriate to evaluate the quality of a viewpoint.

We now devote our attention to the *mutual information* between V and O, that expresses the degree of *dependence* or *correlation* between the set of viewpoints and the object. Mutual information is given by

$$
\begin{aligned}
I(V,O) &= \sum_{v \in \mathcal{V}} p(v) \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} \\
&= \sum_{v \in \mathcal{V}} p(v)I(v,O), \qquad (6)
\end{aligned}
$$

where we define

$$I(v,O) = \sum_{o \in \mathcal{O}} p(o|v) \log \frac{p(o|v)}{p(o)} \qquad (7)$$

as the *viewpoint mutual information* (VMI), which gives us the degree of dependence between the viewpoint v and the set of polygons, and it is a measure of the *quality* of viewpoint v. Consequently, mutual information $I(V,O)$ can be interpreted as the average viewpoint quality. *Quality* is considered here equivalent to *representativeness*. It is also important to indicate that the level of resolution of the viewpoint sphere will determine the accuracy of the measures.

In our framework, the best viewpoint is defined as the one that has *minimum* VMI. High values of the measure mean a high dependence between viewpoint v and the object, indicating a highly *coupled* view (for instance, between the

viewpoint and a small number of polygons with low average visibility). On the other hand, the lowest values correspond to the most *representative* or *relevant* views, showing the maximum possible number of polygons in a balanced way.

3.2. Discussion

Note that $I(v,O) = KL(p(O|v)|p(O))$, where $p(O|v)$ is the conditional probability distribution between v and the object and $p(O)$ is the marginal probability distribution of O, which in our case corresponds to the distribution of the average of projected areas. It is worth observing that $p(O)$ plays the role of the *target* distribution in the KL distance and also the role of the *optimal* distribution since our objective is that $p(O|v)$ becomes similar to $p(O)$ to obtain the best views. On the other hand, this role agrees with intuition since $p(O)$ is the average visibility of polygon o over all viewpoints, i.e., the *mixed distribution* of all views, and we can think of $p(O)$ as representing, with a single distribution, the knowledge about the scene. Note the difference between VMI (7) and VKL (3), due to the fact that in the last case the distance is taken with respect to the actual areas.

In [VFSG06], it has been shown that the main advantage of VMI over VE is its robustness to deal with any type of discretisation or resolution of the volumetric dataset. The same advantage can be observed for polygonal data. Thus, while a highly refined mesh will attract the attention of VE, VMI will be almost insensitive to changes in the mesh resolution. This behavior of both measures with respect to the discretization can be deduced from the mathematical analysis of VE and VMI. For instance, let us assume that a regular polygon o of the object is subdivided into two equal parts o_1 and o_2 such that $p(o_1|v) = p(o_2|v)$, $p(o_1) = p(o_2)$, $p(o|v) = p(o_1|v) + p(o_2|v)$ and $p(o) = p(o_1) + p(o_2)$. Assuming that only the term referred to polygon o can change in the formulas for VE (2) and VMI (7), we analyze their variation after the subdivision of o. The variation of VE is given by

$$
\begin{aligned}
\delta H(O|v) = &-p(o_1|v) \log p(o_1|v) - p(o_2|v) \log p(o_2|v) - \\
&(-p(o|v) \log p(o|v)) = p(o|v).
\end{aligned}
$$

Therefore, VE increases with a value $p(o|v)$ after the subdivision. On the other hand, the variation of VMI is given by

$$
\begin{aligned}
\delta I(v,O) = &\, p(o_1|v) \log \frac{p(o_1|v)}{p(o_1)} + p(o_2|v) \log \frac{p(o_2|v)}{p(o_2)} - \\
&p(o|v) \log \frac{p(o|v)}{p(o)} = 0.
\end{aligned}
$$

Thus, VMI remains invariant to the proposed subdivision. In general, if we compare both measures for different discretisations, mutual information will give similar results and VE will show an erratic behavior. Note that HM is also highly dependent on the discretisation, since the first term in (1) is
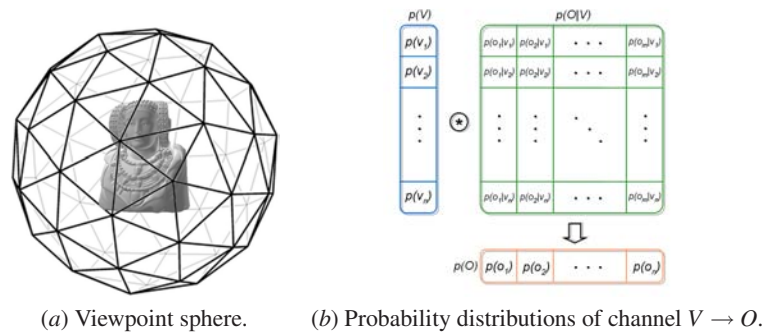
(*a*) Viewpoint sphere.   (*b*) Probability distributions of channel $V \rightarrow O$.

Figure 1: *Viewpoint information channel.*

given by the quotient between the number of visible polygons and the total number of polygons. The behavior of all these measures with respect to the discretisation will be experimentally shown in the next section.

References

[AVF04] ANDÚJAR C., VÁZQUEZ P. P., FAIRÉN M.: Way-finder: guided tours through complex walthrough models. *Computer Graphics Forum (Eurographics 2004)* (2004).

[BET95] BÜLTHOFF H., EDELMAN S., TARR M.: How are three-dimensional objects represented in the brain? *Cerebral Cortex 5* (1995), 247–260.

[BS05] BORDOLOI U. D., SHEN H.-W.: Viewpoint evaluation for volume rendering. In *Visualization, IEEE 2005* (May 2005), pp. 62–62.

[BTB99] BLANZ V., TARR M., BÜLTHOFF H.: What object attributes determine canonical views? *Perception 28* (1999), 575–599.

[CSCF07] CASTELLÓ P., SBERT M., CHOVER M., FEIXAS M.: Viewpoint entropy-driven simplification. In *Proceedings of WSCG 2007* (2007).

[GBL02] GONZÁLEZ-BAÑOS H. H., LATOMBE J.-C.: Navigation strategies for exploring indoor environments. *I. J. Robotic Res. 21*, 10-11 (2002), 829–848.

[GCO06] GAL R., COHEN-OR D.: Salient geometric features for partial shape matching and similarity. *ACM Trans. Graph. 25*, 1 (2006), 130–150.

[GRMS01] GOOCH B., REINHARD E., MOULDING C., SHIRLEY P.: Artistic composition for image creation. In *Rendering Techniques* (2001), pp. 83–88.

[JS06] JI G., SHEN H.-W.: Dynamic view selection for time-varying volumes. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 1109–1116.

[KV06] KIM Y., VARSHNEY A.: Saliency-guided enhancement for volume visualization. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 925–932.

[LME06] LU A., MACIEJEWSKI R., EBERT D. S.: Volume composition using eye tracking data. In *Eurographics/ IEEE-VGTC Symposium on Visualization* (2006), Ertl T., Joy K., Santos B., (Eds.).

[LVJ05] LEE C. H., VARSHNEY A., JACOBS D. W.: Mesh saliency. *Computer Graphics (Proceedings of SIGGRAPH'05)* (July - August 2005). Held in Los Angeles, USA.

[PB96] PLEMENOS D., BENEYADA M.: Intelligent display techniques in scene modelling. new techniques to automatically compute good views. In *International Conference GraphiCon'96* (July 1996). Held in St. Petersburg, Russia.

[PPB*05] POLONSKY O., PATANÈ G., BIASOTTI S., GOTSMAN C., SPAGNUOLO M.: What's in an image? *The Visual Computer 21*, 8-10 (2005), 840–847.

[PRC81] PALMER S., ROSCH E., CHASE P.: Canonical perspective and the perception of objects. *Attention and Performance IX* (1981), 135–151.

[SPFG05] SBERT M., PLEMENOS D., FEIXAS M., GONZÁLEZ F.: Viewpoint quality: Measures and applications. In *Proceedings of 1st Computational Aesthetics in Graphics, Visualization and Imaging* (May 2005). Held in Girona, Spain.

[SPT06] SOKOLOV D., PLEMENOS D., TAMINE K.: Methods and data structures for virtual world exploration. *The Visual Computer 22*, 7 (2006), 506–516.

[TBZB97] TARR M., BÜLTHOFF H., ZABINSKI M., BLANZ V.: To what extent do unique parts influence recognition across changes in viewpoint? *Psychological Science 8*, 4 (1997), 282–289.

[TFTN05] TAKAHASHI S., FUJISHIRO I., TAKESHIMA Y., NISHITA T.: Locating optimal viewpoints for volume visualization. In *Visualization, IEEE 2005* (May 2005).

[VFSG06] VIOLA I., FEIXAS M., SBERT M., GRÖLLER M. E.: Importance-driven focus of attention. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 933–940.

[VFSH01]   VÁZQUEZ P. P., FEIXAS M., SBERT M., HEI-
DRICH W.: Viewpoint selection using viewpoint entropy.
In *Proceedings of Vision, Modeling, and Visualization
2001* (Stuttgart, Germany, November 2001), Ertl T., Girod
B., Greiner G., Niemann H., Seidel H.-P., (Eds.), pp. 273–
280. Held in Stuttgart, Germany.

[VFSH03]   VÁZQUEZ P. P., FEIXAS M., SBERT M., HEI-
DRICH W.: Automatic view selection using viewpoint en-
tropy and its application to image-based modeling. *Com-
puter Graphics Forum* (Desember 2003).

[VFSL06]   VÁZQUEZ P. P., FEIXAS M., SBERT M., LLO-
BET A.: Realtime automatic selection of good molecular
views. *Computers & Graphics 30*, 1 (February 2006), 98–
110.

[YSY*06]   YAMAUCHI H., SALEEM W., YOSHIZAWA S.,
KARNI Z., BELYAEV A. G., SEIDEL H.-P.: Towards
stable and salient multi-view representation of 3d shapes.
In *SMI* (2006), p. 40.

Applications of Information Theory to Computer Graphics Part VI: View Selection in Scientific Visualization

Ivan Viola

University of Bergen, Norway

1. Introduction

Concepts for optimal viewpoint estimation in computer graphics have been intensively studied in the past decade. These concepts (reviewed in Part V) accommodate information-theoretic measures into viewpoint characteristics such as viewpoint entropy, which is based on projected area and number of visible polygons. These accommodation of information-theory measures have been used for estimating the most *informative* viewpoint or a minimal set of viewpoints that characterize given polygonal scene in a most expressive way.

Viewpoint selection is of increasing interest for the scientific visualization research. Approaches for estimation of most informative viewpoints for scientific data sets are somehow similar to those, developed for polygonal data, however there are differences that arise due to different focus of scientific visualization. One difference, as opposed to polygonal graphics, is that the underlying data is larger and more complex. Scientific data is originating from measurements and simulations that have very heterogeneous resulting output. Some data types can be aligned to a structure as a grid, other types are missing such a structure and are represented as sparse points without any connectivity information. Medical data sets are often represented as scalar values per sample point. In other areas, such as meteorological or flow simulations, several attributes per sample point are usual. Many natural phenomena are studied through their development in time. These studies are represented through time-varying data sets, of which visualization is often effective way of providing insights. Moreover scientific data often differ in the level of semantics. Some data contain solely measurements without any a priori knowledge about structures, other data sets contain information about most relevant structures (e.g., critical points in flow data). Summarizing, the heterogeneity among scientific data types, as opposed to *simple* polygonal scene, is aligned with heterogeneity in visualization approaches of viewpoint quality computation.

Second aspect of scientific visualization, as opposed to standard computer graphics, is that it serves gaining insights of complex underlying data. There is always a purpose that drives visualization, usually in form of needs abstracted from industry, sciences, or medicine. Here visualization can serve two purposes, exploration of new, partly unknown data by way of visual analysis by a group of domain users and dissemination of knowledge gained during the exploration process towards another spectrum of audience. Computer-guided viewpoint selection seems to be promising tool for addressing both scenarios, knowledge gain as well as knowledge dissemination.

In this part of the tutorial we are reviewing state-of-the-art view selection strategies for volume visualization. We start with approaches where viewpoint selection for polygonal data has been adapted for view selection on density volumes. Next we review approaches for viewpoint quality measures for time-varying volumes, in terms of estimation of static informative viewpoints for the entire time span or a dynamic viewpoint showing whole time series by continuously changing the viewpoint position to show most of the information.

After describing how information-theoretic measures can be applied for unclassified volumetric data, we give more space to approaches that incorporate higher level semantics into viewpoint selection. In particular we focus on guided navigation technique denoted as *importance-driven focus of attention* where information-theoretic measures work on volumetric data organized into *objects* of varying importance. We give an outlook towards semantics-driven viewpoint selection and review early work in this promising direction.

2. Optimal Views for Volume Data

Techniques to measure *informativeness* of a viewpoint reflect how much of the overall scene is *visible* through particular viewpoint. These metrics differ according to infor-

mation provided in the data as well as according to the aim of visualization. In the following we will review view selection techniques where knowledge about the data is limited to voxel values and visual mapping specification.

2.1. Viewpoint Estimation for Interval Volumes

One of the first approaches on view selection for volume data has been done on evaluation of viewpoint entropy on extracted iso-surfaces or interval volumes respectively [TFTN05]. This approach represents a transition between surface-based visibility estimation working on polygons and purely voxel-based visibility estimation. The viewpoint quality measure here is viewpoint entropy (see Part V), applied to faces of iso-surfaces or interval volumes. These interval volumes or iso-surfaces represent set of features directly derived from the original data values that can be optionally weighted by average opacity specified by transfer function as importance distribution among features. Viewpoint quality considers feature arrangement in order to avoid feature overlapping.

This technique performs quite well during the visibility estimation phase and the viewpoint selection calculations are nearly real-time. This might be of considerable advantage, however, this approach does not compute visibilities of the volume but of surfaces surrounding volumes instead. Therefore very little portion of information contained in the volume is processed. The only way how the volume information is integrated into the viewpoint estimation process is the importance computed by the average opacity of the interval volume. Figure 1 shows best and worst views at set of interval volumes extracted from the hydrogen data set including bounding sphere plot indicating viewpoints quality.

2.2. View Selection for Static and Temporal Data

In parallel to the interval volumes approach discussed above, the viewpoint selection algorithm [BS05] that evaluates visibility of the volume directly, has been proposed. Viewpoint selection for volumes is an adaptation of the viewpoint entropy approach for polygonal data to compute the visibility from volume, i.e., voxels, by substituting the area visibility distribution of faces by the distribution obtained from the quotient between the voxel visibility and the voxel importance (*noteworthiness* factor). Importance distribution is defined per-voxel and is equal to the opacity value specified in the transfer function. This means more opaque voxels will get more prominence than the low opacity ones. This work has additionally suggested information-theoretic measures for clustering views according to similarity using the Jensen-Shannon divergence (see parts II and V). In addition, this work proposes scheme for estimation of static viewpoint on time-varying data. This is realized through conditional entropy where the different random variable distributions are obtained from neighboring time-steps.

This work proposes to perform visibility calculations directly on the volume and thus conveying more information about the volume as compared to the interval volume approach. In addition, the partitioning approach enables to capture entire scene by small set of snapshots. The viewpoint estimation is considering the temporal aspect in the evaluation instead of evaluating viewpoint for every time step separately. The importance distribution, however, is on the per-voxel level meaning that the data is not organized into higher semantics other than original data elements. Figure 2 shows tooth data set from several viewpoints that together capture the most of information about the scene and a static viewpoint on time-varying shockwave data set.

Previously mentioned approach on view selection for volumes has been recently extended to support dynamic viewpoint change for static and time-varying data [JS06] enabling guided flythrough over most interesting viewpoints. The new approach in addition modifies the voxel relevance function by incorporating shape characteristics and color in addition to opacity value.

3. View Selection Using Higher-Level Semantics

Above mentioned approaches are focusing on how to estimate good viewpoints driven by values in the data or assigned visual representation. They do not incorporate additional gained knowledge about the data and do not specifically address particular application scenario. In this section we will focus on view selection techniques that incorporate higher-level knowledge. Voxels or polygons are organized into higher-level semantics, i.e., *objects*. Sufficient knowledge about the data is gained beforehand in a user-steered classification process. View selection explicitly targets gained knowledge dissemination in form of guided navigation with enhancement of features in focus.

First we will discuss focus of attention approach based on varying importance distribution among objects. The second discussed approach targets guided navigation for knowledge dissemination for medical intervention planning.

3.1. Importance-Driven Focus of attention

Importance-driven focus of attention [VFSG06] provides a guided navigation through pre-classified features in the volumetric data set. Object of interest (focus) is directly specified by user. A characteristic viewpoint for this object is selected in combination with a visually pleasing discrimination of the focus from the remaining context information. By changing the object of interest, both viewpoint settings and visual parameters are smoothly modified to put emphasis on the newly selected object of interest. Characteristic viewpoints are estimated in an information-theoretic framework using viewpoint mutual information, a measure adapted from information theory (see part V). Both stages, i.e., the interactive focusing approach and the identification of character-

istic viewpoints, are controlled by an intuitive importance distribution among structures within the volumetric data.

Focus of attention needs visual discrimination of interesting objects from other elements in the scene. It is realized through a visual emphasis of the object of interest, while other objects presented as context are visually suppressed. Changes in opacity and color saturation are used to control the visual emphasis.

In addition to visual discrimination, objects in focus have to be shown from a good and characteristic view where most of the focus structures are perceivable. The most interesting object must not be occluded by less relevant parts. If possible the focus should be in front of other features. In case where the feature of interest is always occluded by other features, cut-away views or other concepts from illustration are be included into the visualization. Furthermore a proper orientation of the up-vector of the viewpoint and a proper positioning of the focus are important to consider in the viewpoint specification.

Finding a viewpoint where the characteristics of a specific feature are clearly visible naturally requires the visibility estimation of the feature under specific viewing settings. It requires ray casting of the whole data set from various viewpoints similarly to view selection for volumes discussed above (see Section 2.2). Visibility computation is based on the opacity contribution of each voxel and object visibility is computed as the sum of voxel visibilities belonging to the object. Voxel membership to a particular object is a fuzzy value, i.e., voxel may belong to several objects at the same time. Additionally two weights influence the visibility of an object, i.e., image-space weight and object-space weight. Image-space weight penalizes the visibility of objects when they are located outside the image center. Object-space weight assigns higher visibility to objects which are closer to the viewing plane and penalizes those that are more far away.

Object visibility is then mapped to a conditional probability of the object for a given viewpoint. These values are used for computation of good viewpoints for a given object by using information-theoretic framework based on viewpoint mutual information and combined with object importance information.

After selecting visual representations of objects and by identifying representative viewpoints, the crucial information to perform interactive focus of attention is available. The importance distribution is in the interactive part a direct mapping of user's interest. Importance is directly mapped to focus discrimination and level of ghosting in cut-away views. The viewpoint transformation is also controlled by importance distribution, smoothly changing to characteristic viewpoint of selected object.

Guided navigation drives user's focus to the object of interest, although still allowing to manipulate the viewpoint

manually for custom inspection. In this scenario the object of interest is represented densely and clearly visible by using interactive cut-away views. After the manual viewpoint manipulation is finished, the viewpoint smoothly changes back to the characteristic viewpoint of the object in focus. User is then free to select another object to be focused onto. Focusing on a specific feature in the human hand dataset is shown in Figure 3. Viewpoint smoothly changes from characteristic viewpoint for the entire volume to viewpoint emphasizing the object of interest. As shown, parallel to the viewpoint change, the focus is discriminated from neighboring structures and the level of ghosting is continuously suppressing visual representation of occluding structures.

3.2. Semantics-Driven Guided Navigation

Viewpoint selection for intervention planning [KMP07] estimates visibility of objects from extracted iso-surfaces. Good viewpoints are estimated using many parameters with adjustable influence: object entropy, importance of occluders, size of unoccluded surface, preferred view region by surgeons, distance to viewpoint, and viewpoint stability. As example how tightly is viewpoint estimation bound to specific application is demonstrated by one of parameters: the distance to important feature defines importance of other features (e.g., neck muscles close to lymph node which is in focus). Furthermore guided navigation supports close zooming to the object of interest. Guided navigation through lymph nodes in the neck data set is shown in Figure 4.

Strong contribution of the domain knowledge parameters in the viewpoint selection computation shows that semantic-driven techniques outperform other existing automatic view selection techniques in a specific visualization scenario. This statement has been supported by evaluation of a user study that significantly included medical domain users.

4. Conclusion

Selection of a preferred viewpoint without targeting specific application is a very subjective choice. High knowledge about the data and the purpose of selecting particular viewpoint for a very specific application is moving viewpoint selection from subjective *taste* to more objective characteristics. Current trend in visualization research indicates that this is the direction for a good good viewpoint estimation. The work on viewpoint selection for intervention planning nicely demonstrates this trend.

References

[BS05] BORDOLOI U. D., SHEN H.-W.: Viewpoint evaluation for volume rendering. In *Visualization, IEEE 2005* (May 2005), pp. 62–62.

[JS06] JI G., SHEN H.-W.: Dynamic view selection for time-varying volumes. *IEEE Trans. Vis. Comput. Graph. 12*, 5 (2006), 1109–1116.

[KMP07]   Konrad Mühler Mathias Neugebauer
C. T., Preim B.:   Viewpoint selection for intervention
planning. In *Proceedingss of Eurographics/ IEEE-VGTC
Symposium on Visualization* (2007), pp. 267–274.

[TFTN05]   Takahashi S., Fujishiro I., Takeshima
Y., Nishita T.: Locating optimal viewpoints for volume
visualization. In *Visualization, IEEE 2005* (May 2005).

[VFSG06]   Viola I., Feixas M., Sbert M., Gröller
M. E.: Importance-driven focus of attention. *IEEE Trans.
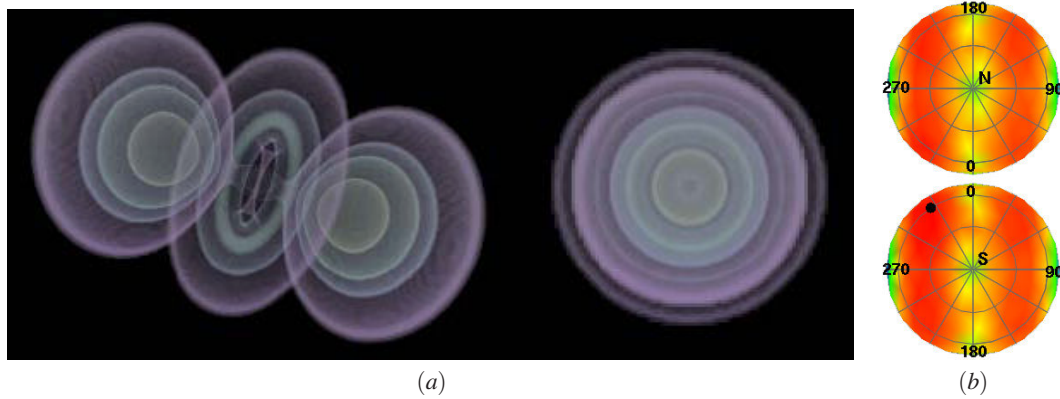Vis. Comput. Graph. 12*, 5 (2006), 933–940.

(*a*)   (*b*)

Figure 1: *Viewpoint estimation for interval volumes [TFTN05]: (a) best and worst views at interval volumes extracted from data set containing simulated electron density distribution in a hydrogen atom; (b) the sphere plot shows the viewpoint quality distribution where the green color encodes low quality viewpoints and the red encodes good and* informative *viewpoints. Best and worst views are indicated by a black and white dot respectively. Used by permission.*
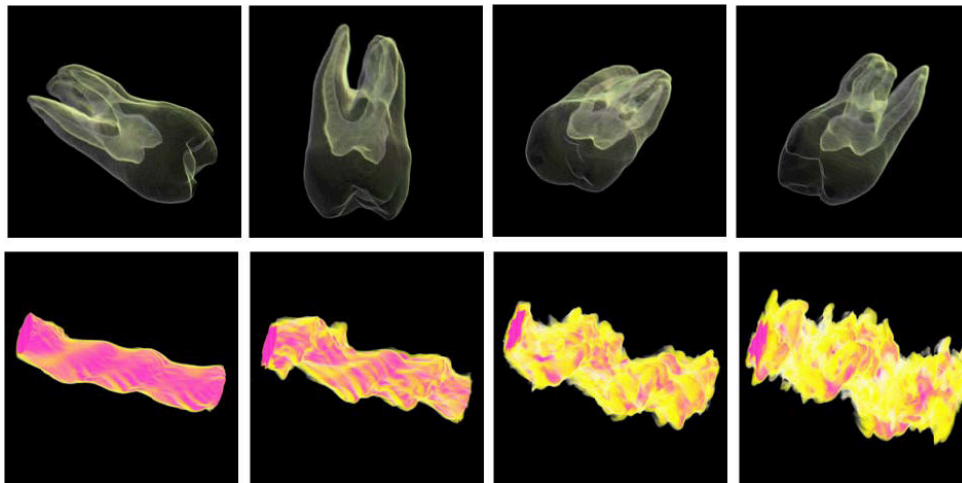

Figure 2: *View selection for static and time-varying volumes [BS05]: (upper row) four selected views on the tooth data set from four bounding sphere partitions; (bottom row) time-series of the shockwave data set from best temporal domain preserving viewpoint. Used by permission.*

Figure 3: *Importance-driven focusing [VFSG06]: Smooth guided navigation from overall characteristic viewpoint to focus on object of interest.*
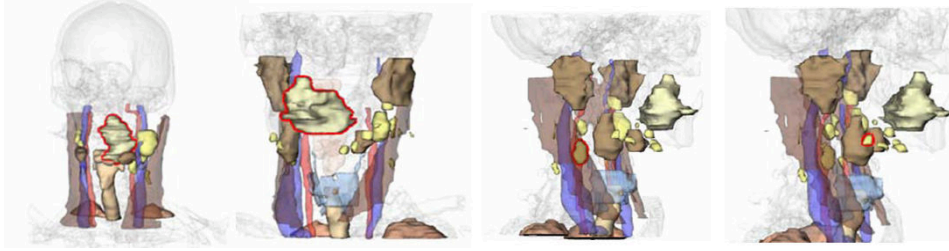
Figure 4: *Semantics-driven view selection [KMP07]: Guided navigation through features in the human neck data set assists studying the correspondence between focus objects, i.e., lymph nodes and surrounding tissue such as neck muscle. Used by permission.*