Information Theory Tools for Scene Discretization

Miquel Feixas*, Esteve del Acebo*, Philippe Bekaert† and Mateu Sbert*
*Institut d'Informàtica i Aplicacions, Universitat de Girona
feixas|acebo|mateu@ima.udg.es
†Department of Computer Science, Katholieke Universiteit Leuven
Philippe.Bekaert@cs.kuleuven.ac.be

Abstract. Finding an optimal discretization of a scene is an important but difficult problem in radiosity. The efficiency of hierarchical radiosity for instance, depends entirely on the subdivision criterion and strategy that is used. We study the problem of adaptive scene discretization from the point of view of information theory. In previous work, we have introduced the concept of mutual information, which represents the information transfer or correlation in a scene, as a complexity measure and presented some intuitive arguments and preliminary results concerning the relation between mutual information and scene discretization. In this paper, we present a more general treatment supporting and extending our previous findings to the level that the development of practical information theory-based tools for optimal scene discretization becomes feasible.

Keywords: information theory, radiosity, adaptive scene discretization

1 Introduction

From the point of view of Information Theory (IT) [4, 7], the discretization of a scene into patches in the radiosity method can be understood as the encoding or compression of a continuous signal over a discrete channel with the consequent distortion or information loss. The continuous signal corresponds to the (continuous) radiosity function $B(x)$ on the surfaces of the scene. The discrete channel is represented by a Markov Chain, with states corresponding to the patches into which the scene is discretized. The transition probabilities essentially correspond to the form factors.

To each discretization can be associated a quantity, *mutual information*, which quantifies the information transfer or information gain in a system. From the IT point of view, the optimal discretization corresponds with the one with minimal loss of information, or vice versa, the one with the highest mutual information. The most common distortion measure is the mean square error (MSE) or equivalently the L_2 norm. We can expect that the optimal discretization should give a minimum MSE.

Mutual information was proposed as a scene complexity measure in [9]. In [9] we also presented some intuitive arguments and preliminary results concerning the relation between mutual information and scene discretization. In this paper, we present a theorem that solidifies this previous work and allows to extend it to the level that the development of practical IT-based tools for scene discretization becomes feasible. We derive the optimal discretization for some simple scenes and common patch configurations. A heuristic for the patch-to-patch case is given, which can be used as an oracle for subdivision. We give also some evidence that the optimal mutual information discretization corresponds well to an optimal mean square error discretization.

The organisation of the paper is as follows: In section 2 we present a brief overview of our previous work [1, 8, 9] on the subject. In section 3 we present a general setting with its application to visibility, radiosity and importance. In section 4 we develop heuristics for good discretizations. In section 5 a proposal for a new subdivision oracle is presented and finally in section 6 we present our conclusions and future work.

2 An IT framework for the analysis of scene complexity

In this section we briefly review some basic concepts introduced in [1, 8, 9].

Complexity. Complexity reflects "the difficulty of describing a system, the difficulty of reaching a goal, the difficulty of performing a task, and so on" [14]. Over the last twenty years several complexity measures have been proposed from different fields to quantify the degree of structure or correlation of a system [10]. We study this concept of complexity from the point of view of IT [4, 7].

Markov Chains. In general, a Markov Chain is a sequence of random variables (sets of values or events with associated probabilities) $X^k, k = 0 \ldots \infty$ in which each $X^k, k \geq 1$ depends only on the previous X^{k-1} and not on the ones before. A Markov Chain is often characterized by a set of *states* labelled $i = 1, \ldots, n$. 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. In each step, the imaginary particle makes a transition from its current state i to a new state j with *transition probability* P_{ij}. Under certain conditions (which are fulfilled in the context of this paper), the probabilities of finding the particle in each state i converge to a *stationary distribution* $w = (w_1, \ldots, w_n)$ after a number of steps. The stationary, or equilibrium probabilities w_i fulfil the relation $w_i = \sum_{j=1}^{n} w_j P_{ji}$. For the Markov Chains we deal with in this paper, the stationary distribution also satisfies another (balance or reciprocity) relation $w_i P_{ij} = w_j P_{ji}$.

Markov Chains for studying scene visibility complexity. In [9], we studied *discrete scene visibility* complexity by letting the states $i = 1, \ldots, n_p$ correspond to the patches of a scene and the transition probabilities P_{ij} with the form factors F_{ij}. n_p denotes the number of patches. It can be shown [9] that the stationary probabilities of the resulting Markov Chain are given by $w_i = A_i/A_T$, the relative area of the patches i of the scene (A_i is the area of patch i, A_T is the total scene surface area).

When the states form a countable set, as above, 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 $w(x) = 1/A_T$ results. We have used this Markov Chain to study *continuous scene visibility* complexity.

Shannon entropy. The Shannon entropy of the stationary distribution (w_i) of a discrete Markov Chain is defined as $H_p = -\sum_{i=1}^{n} w_i \log w_i$. In the case of discrete scene visibility:

$$H_p = -\sum_{i=1}^{n_p} \frac{A_i}{A_T} \log \frac{A_i}{A_T}. \quad (1)$$

The Shannon entropy, which we will call the (discrete) *positional entropy* in this case, reflects the uncertainty on the position (patch) of a particle travelling an infinite random walk with transition probabilities equal to the form factors. The logarithms are taken in base 2 and we take $0 \log 0 = 0$.

For a continuous Markov Chain, e.g. for studying continuous visibility complexity, the sum $\sum_{i=1}^{n}$ shall be replaced by an integral over the uncountable set of states.

Entropy rate. The entropy rate of a Markov Chain with transition probability matrix (P_{ij}) and stationary probability distribution (w_i) is defined as $H_s = -\sum_{i=1}^{n}\sum_{j=1}^{n} w_i P_{ij} \log P_{ij}$. Applied to scene visibility:

$$H_s = -\sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \frac{A_i}{A_T} F_{ij} \log F_{ij}. \tag{2}$$

The scene visibility entropy rate measures the average uncertainty that remains about the patch j visited next (*destination* patch) when an imaginary particle undergoing an infinite random walk, with the form factors as transition probabilities, is known to be on a given patch i (*source* patch).

Mutual information. Mutual information is defined as the difference of Shannon entropy and entropy rate: $I_s = H_p - H_s$. The *discrete scene visibility mutual information*

$$I_s = \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \frac{A_i F_{ij}}{A_T} \log \frac{F_{ij} A_T}{A_j} \tag{3}$$

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* [9].

Continuous versus discrete mutual information. By discretizing a scene into patches, a distortion or error is introduced. In a way, to discretize means to equalize. Obviously, the maximum accuracy of the discretization is obtained when the number of patches tends to infinity and the size of the patches tends to zero.

Mutual information between two continuous random variables X and Y is the limit of the mutual information between their discretized versions [7, 11]. In our case, discrete scene visibility mutual information I_s (3), converges to *continuous scene visibility mutual information I_s^c* when the maximum patch size tends to zero [9]:

$$I_s^c = \int_{x\in S}\int_{y\in S} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dx dy \tag{4}$$

I_s^c expresses with maximum accuracy the information transfer or correlation in a scene. This is an *absolute* measure of the complexity of scene visibility. On the other hand, discrete mutual information I_s expresses the complexity of a discretized scene, which is always lower than the corresponding I_s^c.

Scene radiosity complexity. Measures for the complexity of a scene, taking also diffuse illumination into account besides visibility, can be obtained by using a different pair of discrete and continuous Markov Chains [9] (see also §3.4 below).

3 Mutual information and patch subdivision

In [9] we presented intuitive arguments and preliminary results suggesting that between different discretizations of the same scene the most precise one will be the one that has the highest mutual information I_s, i.e., the one that best captures information transfer or has minimum information loss. We presented experiments for scene visibility and radiosity complexity.

In this section, we present a theorem (§3.1) that supports our preliminary findings and allows to derive more exact predictions of the gain in mutual information resulting from subdivision of scene patches. We first study the problem for a general scene complexity Markov Chain (§3.2) and next consider the application to scene visibility (§3.3), radiosity (§3.4) and importance (§3.5).

3.1 State refinement and continuous versus discrete mutual information

Theorem 1 *Consider a discrete Markov chain over a set of states labelled $i, j = 1, \ldots, n$, with transition probability matrix $P = (P_{ij})$ and stationary distribution $w = (w_1, w_2, \ldots, w_n)$ which satisfies the reciprocity relation $w_i P_{ij} = w_j P_{ji} \quad \forall i, j$. When a state i is refined into m sub-states $i_k, k = 1, \ldots, m$ such that*

(a) $w_{i_k} P_{i_k j} = w_j P_{j i_k} \quad \forall i_k, j$ *(reciprocity relation with the sub-states);*
(b) $P_{ji} = \sum_{k=1}^{m} P_{j i_k} \quad \forall j$ *(the sub-states i_k "cover" i),*

mutual information increases (or remains the same). (Proof in appendix A.)

Corollary 1 *Continuous mutual information I^c of a scene which fulfils the conditions of the above theorem is the least upper bound to discrete mutual information I.*

Proof: Continuous mutual information between two continuous random variables X and Y is the limit of the discrete mutual information between their discretized versions [7]. The statement that I^c is the least upper bound to I then immediately follows from the above theorem. □

3.2 Patch-to-patch increase in mutual information

If we consider a scene with planar patches, the increase in mutual information between two planar patches i and j when subdividing i into m sub-patches is

$$\begin{aligned}(\Delta I)_{ij} &= 2\left(\left(\sum_{k=1}^{m} w_{i_k} P_{i_k j} \log \frac{P_{i_k j}}{w_j}\right) - w_i P_{ij} \log \frac{P_{ij}}{w_j}\right) \\ &= 2\left(\left(\sum_{k=1}^{m} w_{i_k} P_{i_k j} \log P_{i_k j}\right) - w_i P_{ij} \log P_{ij}\right) \qquad (5)\end{aligned}$$

This can be obtained from (10) and (11) in appendix A , where the second half of these formulae is null, and from the conditions of the theorem. For a regular subdivision, $w_{i_k} = \frac{w_i}{m}$, we have

$$(\Delta I)_{ij} = 2\left(\left(\frac{w_i}{m} \sum_{k=1}^{m} P_{i_k j} \log P_{i_k j}\right) - w_i P_{ij} \log P_{ij}\right)$$

and it can be shown that the *theoretical* maximum possible increase in I happens when for all k except one $P_{i_k j} = 0$. The one not null can be shown to be equal to mP_{ij}. Thus the maximum possible increase in I is given by

$$max((\Delta I)_{ij}) = 2(w_i P_{ij} \log mP_{ij} - w_i P_{ij} \log P_{ij}) = 2w_i P_{ij} \log m \tag{6}$$

If we sum over j, we will obtain the maximum possible increase when dividing a given patch i

$$\sum_j max((\Delta I)_{ij}) = \sum_j 2w_i P_{ij} \log m = 2w_i \log m$$

Thus a heuristic to pick a patch to subdivide regularly, lacking any other knowledge, would be to take the one with maximum w.

3.3 Application to visibility

Taking $w_i = \frac{A_i}{A_T}$ and $P_{ij} = F_{ij}$, it is easy to see that the hypotheses of theorem 1 are fulfilled. Thus, from (5), the increment of mutual information is in this case

$$\begin{aligned}(\Delta I)_{ij} &= 2\left(\left(\sum_{k=1}^{m} \frac{A_{i_k}}{A_T} F_{i_k j} \log \frac{F_{i_k j} A_T}{A_j}\right) - \frac{A_i}{A_T} F_{ij} \log \frac{F_{ij} A_T}{A_j}\right) \\ &= 2\left(\left(\sum_{k=1}^{m} \frac{A_{i_k}}{A_T} F_{i_k j} \log F_{i_k j}\right) - \frac{A_i}{A_T} F_{ij} \log F_{ij}\right) \end{aligned} \tag{7}$$

Thus, the maximum increase upon a regular subdivision is

$$max((\Delta I)_{ij}) = 2\frac{A_i}{A_T} F_{ij} \log m$$

and the maximum possible increase when dividing a given patch i is $2\frac{A_i}{A_T} \log m$.

3.4 Application to radiosity

In the radiosity setting, we consider the following transition probabilities

$$P_{ij} = \frac{\int_{S_i} \int_{S_j} F(x_i, x_j) B(x_i) B(x_j) dx_i dx_j}{\int_{S_i} B(x_i) \frac{B(x_i) - E(x_i)}{R(x_i)} dx_i}$$

These are the extension to the continuous case of the discrete null variance probabilities (see below) and fulfil $\sum_j P_{ij} = 1$ due to the additivity of the integral over its domain S, where $S = \cup_j S_j$, and the fact that the radiosities fulfil the diffuse rendering equation

$$B(x_i) = E(x_i) + R(x_i) \int_S F(x_i, x) B(x) dx$$

It can be easily checked that the equilibrium probabilities are

$$w_i = \int_{S_i} B(x_i) \frac{B(x_i) - E(x_i)}{R(x_i)} dx_i$$

and the reciprocity relation is trivially fulfilled. The normalising factor of w_i is

$$\sum_j \int_{S_j} B(x_j) \frac{B(x_j) - E(x_j)}{R(x_j)} dx_j = \int_S B(x) \frac{B(x) - E(x)}{R(x)} dx$$

If we divide patch i into i_1 and i_2 it is easy to prove, due to the additivity of the integrand, that the hypotheses of the theorem 1 are fulfilled. The radiosity case reverts to the visibility case when $B(x) = k$, where k is a constant, and this happens whenever $\forall x$ $E(x) = k(1 - R(x))$.

Now let us suppose radiosities and reflectivities are constant along each patch. In this case $w_i = A_i B_i \frac{(B_i - E_i)}{R_i}$ and $P_{ij} = \frac{R_i F_{ij} B_j}{B_i - E_i}$. These quantities can be considered a kind of generalized area and form factor respectively, by analogy with the visibility case in section 3.3. The P_{ij} probabilities were found to be the null variance transition probabilities for a gathering random walk in [17].

3.5 Application to importance

The continuous importance $I(x)$, given initial importance $V(x)$, is the solution to the integral equation for importance on a point x [16]:

$$I(x) = V(x) + \int_S R(y) F(x, y) I(y) dy$$

Consider now the transition probability

$$P_{ij} = \frac{\int_{S_i} \int_{S_j} F(x_i, x_j) R(x_i) R(x_j) I(x_i) I(x_j) dx_i dx_j}{\int_{S_i} R(x_i) I(x_i)(I(x_i) - V(x_i)) dx_i}$$

Similarly to the radiosity case, we have $\sum_j P_{ij} = 1$ due to the additivity of the integral over its domain S, where $S = \cup_j S_j$, and the fact that importances fulfil the importance integral equation. The equilibrium probabilities w_i are (without normalising)

$$w_i = \int_{S_i} R(x_i) I(x_i)(I(x_i) - V(x_i)) dx_i$$

and the reciprocity relation is fulfilled. If we divide patch i into i_1 and i_2, similarly to radiosity, the hypotheses of the theorem are fulfilled. It can be seen that for $R(x)I(x) = k$ for all x, that happens when we take $V(x) = k(\frac{1}{R(x)} - 1)$, the importance case reverts to the visibility case.

Now let us suppose importances and reflectivities are constant along each patch. In this case $w_i = A_i R_i I_i (I_i - V_i)$ and $P_{ij} = \frac{R_j F_{ji} I_j}{I_i - V_i}$. When $V_i = \delta_{ik}$ we have the null variance transition probabilities for a shooting random walk [17].

4 Results and discussion

In this section we discuss how mutual information varies for some common patch configurations and some simple scenes (§4.1). We also provide some evidence that an optimal subdivision obtained by mutual information maximization corresponds well to an optimal subdivision in terms of mean square error (§4.2).

4.1 Maximal mutual information subdivision for some common configurations

The following results can be obtained from (7), form factor properties and closed form formulae for the unoccluded form factors [6, 18, 12]:

Partially occluded pair of patches (figure 1a). Consider the subdivision of patch i into two sub-patches:

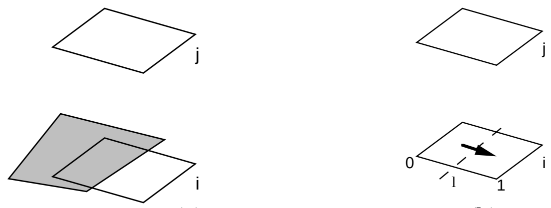
1) Of all subdivisions of i with one sub-patch totally occluded to j, the maximum mutual information increase corresponds to the discontinuity mesh (see appendix B.1).

2) When the point-to-point form factor $F(x, y)$ is approximately constant for x in the unoccluded part of i and y in j, the maximum increase in mutual information corresponds to the discontinuity mesh (see appendix B.2).

Two square patches with common edge (figure 2). Consistent with observations in [5], orthogonal splitting (figure 2(b)) leads to only a small gain in mutual information. Nothing is gained by orthogonal splitting in the middle. When splitting along a line parallel with the common edge (figure 2(a)) the maximum gain in mutual information results when splitting at a 40% relative distance from the edge (figure 2(c)).

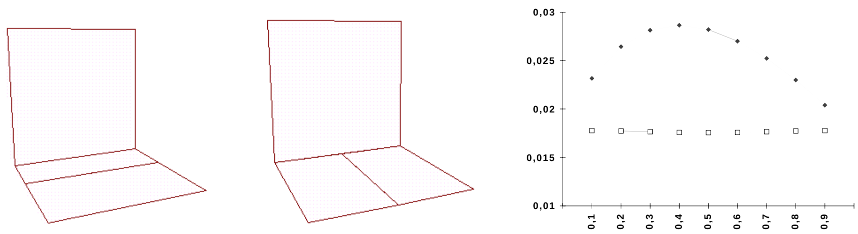
Three square patches with common edges (figure 3). The maximum gain is obtained at a distance 39% from the edge. The small displacement (from 40% to 39%) towards the edge with respect to the previous case is due to the small positive gradient of mutual information for the orthogonal subdivision (see figure 2(b), squares).

Empty cube (figures 4 and 5). The resulting maximum mutual information subdivision is a bit displaced towards the edges with respect to the regular one. Figure 5 shows an example with more subdivisions.



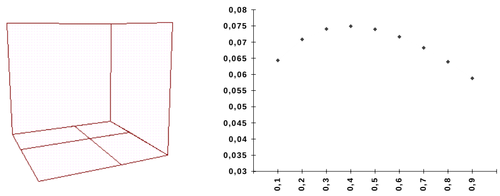
(a) (b)

Figure 1. (a) Partially occluded patch pair. (b) (§4.2) Subdivision of a patch perpendicular to the radiosity gradient. The position of the cutting line is parametrized by the relative distance $0 \le l \le 1$ to one edge.



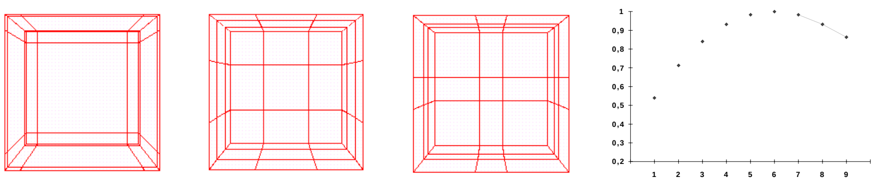
(a) (b) (c)

Figure 2. Mutual information on vertical axis (c) when dividing orthogonal (b,squares) and parallel (a,diamonds) to the common edge. Horizontal axis represents the displacement from the common edge (a) or one side edge (b).



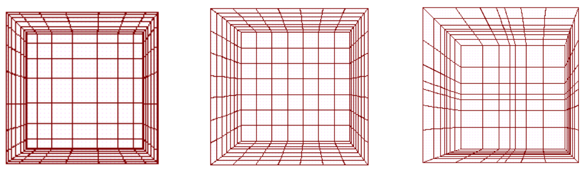
(a) (b)

Figure 3. Mutual information on vertical axis (b) when subdividing a patch in a corner (a). Horizontal axis represents the distance from the parallel division to one common edge.



(a) (b) (c) (d)

Figure 4. (d) Mutual information on vertical axis for an empty cube. Horizontal axis represents the relative displacement of a nearest subdivision to a common edge, ranging from 0 to 10. (b) corresponds to the optimal case, with value near to 6, (a) with value 2 and (c) with value 8.



(a) (b) (c)

Figure 5. An empty cube with (a) optimal ($I = 1.3569$), (b) regular ($I = 1.3331$) and (c) “bad” ($I = 1.2554$) subdivision.

4.2 Mutual information maximization and mean square error

Consider two square patches, i and j, with the following characteristics: B_j is constant over j, $F(x, y)$ is approximately constant for $x \in S_i$, $y \in S_j$ and the reflectivity is constant along each patch. Consider now that the radiosity on patch i varies along one axis parallel to one edge of i, $B(l) = l^n B$, for B constant and l between 0 and 1 parametrizes the patch (figure 1(b)). The increase in mutual information when dividing patch i into sub-patches i_1 and i_2 is given by (5)

$$2(w_{i_1} P_{i_1 j} \log P_{i_1 j} + w_{i_2} P_{i_2 j} \log P_{i_2 j} - w_i P_{ij} \log P_{ij})$$

where in the radiosity case quantities w and P have to be substituted by the values given in section 3.4. The maximum increase in mutual information results when splitting patch i perpendicular to the gradient (this is, across a line $l = k$). The optimal value for l is found by optimising the expression:

$$l^{n+1} \log \frac{1}{l} + (1 - l^{n+1}) \log \frac{1 - l^{n+1}}{1 - l^{n+2}}.$$

For $n = 1, 2, 3, 4$ the optimal values correspond to $l = 0.48, 0.61, 0.68, 0.74$.

Consider now the subdivision problem from the point of view of minimising the L_2 error (or MSE error) on patch i, when assuming constant values for the radiosities on the sub-patches (equal to the average of the continuous radiosity function $B(x)$). After some algebra again, it can be shown that the optimal solution satisfies:

$$2nl^n - l^{n-1} - \ldots - l^2 - l - 1 = 0.$$

For $n = 1, 2, 3, 4$ the optimal values are $l = 0.5, 0.64, 0.72, 0.77$.

We have seen with this example that, in the absence of a form factor gradient, the subdivision cuts along the radiosity gradient and the optimal value that corresponds to maximum increase in mutual information is very near to the minimum L_2 error.

5 Towards a mutual information based oracle for subdivision

It can be seen that for the constant radiosity case the increase in mutual information is given by

$$(\Delta I)_{ij} = 2B_i B_j ((\sum_{k=1}^{m} (A_{i_k} F_{i_k j} \log F_{i_k j}) - A_i F_{ij} \log F_{ij}))$$

Thus, for a regular subdivision and from (6), the maximum possible increase is

$$\begin{aligned} max((\Delta I)_{ij}) &= 2B_i B_j ((\frac{A_i}{m} m F_{ij} \log(m F_{ij})) - A_i F_{ij} \log F_{ij})) \\ &= 2B_i B_j A_i F_{ij} \log m \propto B_i B_j A_i F_{ij} \end{aligned}$$

Thus the quantity $B_i B_j A_i F_{ij}$ expresses the maximum potential gain of mutual information between two patches when subdividing one of them. However, this can be really far from the real gain obtained, when for instance the form factors are fairly equal in the subdivisions, as could be with two parallel patches at some distance and without occlusions. Thus, the use of this quantity as an oracle for subdividing is not recommended. Better, the full expression for ΔI should be used, or at least some information on form factor gradients along subdivisions should be taken into account, the larger the gradient the larger the increase in mutual information.

One could also consider which is the patch with more potential gain in mutual information with respect to all other patches. In this case we sum over j

$$\sum_j B_i B_j A_i F_{ij} = A_i B_i \frac{B_i - E_i}{R_i}$$

this is, the one with the larger generalized area. Thus, in lack of any other information, a heuristic would be to look for the largest generalized area patch to subdivide.

Summarising, an oracle proposal for hierarchical radiosity subdivision is the following:

- A patch of the pair (i, j) will be candidate to subdivide only when the quantity $B_i B_j A_i F_{ij} > \epsilon_1$. This discards subdivisions with small potential increment of mutual information.

- If a pair (i, j) is considered, we pick from both the one corresponding to

$$max(A_i B_i \frac{B_i - E_i}{R_i}, A_j B_j \frac{B_j - E_j}{R_j})$$

  this is, the one with the highest potential mutual information increase.
- A patch of the pair (i, j), say i, is finally subdivided only if the estimated gradient in form factors between the subdivisions and j is larger than a given threshold ϵ_2. This intends to guarantee a real increase in mutual information.

As a cheap gradient estimator we could use the differences between point-to-point form factors from the center of the subdivisions to the center of the j patch.

We remark here that the first step in the oracle is analogous to the power oracle $B_j R_i A_j F_{ji}$ [13, 3], the second step can be seen as an extension to the heuristics of dividing the patch with larger area, and the third step is analogous to the various gradient oracles used in hierarchical radiosity literature [5]. Obviously, the last step in the oracle can be refined to incorporate the exact form of the mutual information function, but at a much higher cost due to the numerical instabilities of the $\log$ function.

6 Conclusions and future research

In this paper we have taken one step further in our application of information theory to study the scene complexity. A general theorem on the increase of mutual information upon subdivision of the scene is presented, with application to visibility, radiosity and importances. The kind of subdivision driven by mutual information maximization is analysed, and shown that it has good properties, such as dividing across visibility and radiosity gradients. Also, evidence has been given on the minimization of MSE error by mutual information driven subdivision. Finally, a subdivision oracle based on the maximum mutual information increase has been presented and its potential utility in hierarchical radiosity justified.

Future research will be directed to analyse more precisely the interplay between MSE and mutual information. Also, the application to hierarchical radiosity of the proposed oracle will be undertaken, and the balance between cost and accuracy in subdivision for oracles reproducing the mutual information function more faithfully will be studied.

Acknowledgements

Many thanks to Roel Martínez and Joaquim Gelabertó. This project has been funded in part with grant numbers TIC 98-586-C03 and TIC-98-973-C03 of the Spanish Government and catalan-flemish joint action number ABM/acs/ACI98-19. The third author acknowledges financial support by a grant from the “Flemish Institute for the Promotion of Scientific-Technological Research in Industry” (IWT #941120) and the Flemish fund for scientific research (NWO-Vl #G.0105.96).

References

1. E. del Acebo, M. Feixas and M. Sbert. “Form Factors and Information Theory”. *Proceedings of 3IA’98*, Limoges, April 1998.
2. R. Badii and A. Politi. *Complexity. Hierarchical structures and scaling in physics.* Cambridge University Press, 1997.

3. Ph. Bekaert, L. Neumann, A. Neumann, M. Sbert and Y. D. Willems. “Hierarchical Monte Carlo Radiosity”. 9^{th} *Eurographics Workshop on Rendering*, pp. 259–268, Vienna, Austria, June 1998.
4. R. E. Blahut. *Principles and practice of Information Theory*. Addison-Wesley, 1987.
5. A. Campbell. “Modelling Global Diffuse Illumination for Image Synthesis”. PhD thesis, Dept. of Computer Science, University of Texas at Austin, 1991.
6. M. F. Cohen and J. R. Wallace *Radiosity and Realistic Image Synthesis*. Academic Press, 1993.
7. T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.
8. M. Feixas, E. del Acebo and M. Sbert. “Entropy of scene visibility”. *WSCG’99*, Plzen, 1999.
9. M. Feixas, E. del Acebo, Ph. Bekaert and M. Sbert. “An information theory framework for the analysis of scene complexity”. To appear in Eurographics’99 Proceedings. Available in *http://ima.udg.es/˜ mateu*.
10. D. P. Feldman and J. P. Crutchfield. “Statistical Measures of Complexity: Why?”. *Physics Letters A*, 238:244–252, 1998.
11. D. P. Feldman. “A brief introduction to: Information Theory, Excess Entropy and Computational Mechanics”. Department of Physics, University of California, July 1997. Available in *http://leopard.ucdavis.edu/dave/index.html*.
12. A. S. Glassner. *Principles of digital image synthesis*. Morgan Kaufmann Publishers, San Francisco, 1995.
13. P. Hanrahan, D. Salzman and L. Aupperle. “Rapid hierarchical radiosity algorithm”. *Computer Graphics (Proc.Siggraph’91)*, 25(4):197-206, 1991.
14. W. Li. “On the relationship between complexity and entropy for Markov chains and regular languages”. *Complex Systems*, 5(4):381-399, 1991.
15. D. Lischinski, F. Tampieri and D. Greenberg. “Combining Hierarchical Radiosity and Discontinuity Meshing”. SIGGRAPH’93 Proceedings, 1993.
16. S. Pattanaik and S. Mudur. “Adjoint Equations and Random Walks in Illumination Computation”. ACM Transactions on Graphics, Vol. 14, No. 1, January 1995.
17. M. Sbert, A. Brusi, R. Tobler and W. Purgathofer. “Random Walk radiosity with generalized transition probabilities”, Research Report IIiA-98-07-RR, Institut d’Informàtica i Aplicacions, Universitat de Girona, 1998. Available in *http://ima.udg.es/˜ mateu*.
18. F. X. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, San Francisco, 1994.

A Proof of theorem 1

Let us imagine a discrete random walk with discrete mutual information

$$I = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i P_{ij} \log \frac{P_{ij}}{w_j} \tag{8}$$

We must show that, if any state is discretized into m sub-states, the discrete mutual information I' of the new random walk fulfils $\Delta I = I' - I \geq 0$. Without loss of generality we divide the nth state into m sub-states $n_1, n_2, \ldots, n_m$. Thus, we have

$$I' = \sum_{i=1}^{n-1} \sum_{j=1}^{n-1} w'_i P'_{ij} \log \frac{P'_{ij}}{w'_j} + \sum_{i=1}^{n-1} \sum_{k=1}^{m} w'_i P'_{in_k} \log \frac{P'_{in_k}}{w_{n'_k}}$$
$$+ \sum_{k=1}^{m} \sum_{j=1}^{n-1} w'_{n_k} P'_{n_k j} \log \frac{P'_{n_k j}}{w'_j} + \sum_{k=1}^{m} \sum_{l=1}^{m} w'_{n_k} P'_{n_k n_l} \log \frac{P'_{n_k n_l}}{w'_{n_l}} \tag{9}$$

where $w_i = w'_i$ for $1 \leq i < n$, $w_n = \sum_{k=1}^{m} w'_{n_k}$, $P_{ij} = P'_{ij}$ for $1 \leq i,j < n$ and $P_{in} = \sum_{k=1}^{m} P'_{in_k}$ for $1 \leq i < n$. Because of $w_i P_{ij} \log \frac{P_{ij}}{w_j} = w_j P_{ji} \log \frac{P_{ji}}{w_i}, \forall i, j$, we have

$$I = 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} w_i P_{ij} \log \frac{P_{ij}}{w_j} + \sum_{i=1}^{n} w_i P_{ii} \log \frac{P_{ii}}{w_i} \tag{10}$$

Then

$$\begin{aligned} I' - I = {} & 2\sum_{i=1}^{n-1}\left(\sum_{k=1}^{m} w_i P_{in_k} \log \frac{P_{in_k}}{w_{n_k}} - w_i P_{in} \log \frac{P_{in}}{w_n}\right) \\ & + \sum_{k=1}^{m}\sum_{l=1}^{m} w_{n_k} P_{n_k n_l} \log \frac{P_{n_k n_l}}{w_{n_l}} - w_n P_{nn} \log \frac{P_{nn}}{w_n} \end{aligned} \tag{11}$$

where the coincident terms in I and I' have been deleted. Applying the above hypotheses and the concavity of the logarithm function for non-negative numbers

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i\right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{12}$$

we can conclude that

$$\Delta I = I' - I \geq 0$$

B Discontinuity meshing

B.1 Subdivision in occluded part

Patch i is divided into sub-patches i_a and i_b, where i_b is totally occluded and i_a has one part i_c unoccluded and one part i_d occluded. Then, from (7) and taking $A_T = 1$ without loss of generality,

$$\begin{aligned} (\Delta I)_{ij} &= 2(A_{i_a} F_{i_a j} \log \frac{F_{i_a j}}{A_j} + A_{i_b} F_{i_b j} \log \frac{F_{i_b j}}{A_j} - A_i F_{ij} \log \frac{F_{ij}}{A_j}) \\ &= 2((A_{i_c} F_{i_c j} + A_{i_d} F_{i_d j}) \log \frac{(F_{j i_c} + F_{j i_d})}{(A_{i_c} + A_{i_d})} - A_i F_{ij} \log \frac{F_{ij}}{A_j}) \\ &= 2(A_{i_c} F_{i_c j} \log \frac{F_{j i_c}}{(A_{i_c} + A_{i_d})} - A_i F_{ij} \log \frac{F_{ij}}{A_j}) \end{aligned} \tag{13}$$

The maximum is obtained when $A_{i_d} = 0$, i.e. when subdivision is made according to discontinuity meshing.

B.2 Subdivision in unoccluded part

Patch i is divided into sub-patches i_a and i_b, where i_a is totally unoccluded and i_b has one part i_c occluded and one part i_d unoccluded. Then, from (7) and taking $A_T = 1$,

$$\begin{aligned} (\Delta I)_{ij} &= 2(A_{i_a} F_{i_a j} \log \frac{F_{i_a j}}{A_j} + A_{i_b} F_{i_b j} \log \frac{F_{i_b j}}{A_j} - A_i F_{ij} \log \frac{F_{ij}}{A_j}) \\ &= 2(A_{i_a} F_{i_a j} \log \frac{F_{i_a j}}{A_j} + A_{i_d} F_{i_d j} \log \frac{F_{j i_d}}{(A_{i_d} + A_{i_c})} - A_i F_{ij} \log \frac{F_{ij}}{A_j}) \\ &\leq 2(A_{i_a} F_{i_a j} \log \frac{F_{i_a j}}{A_j} + A_{i_d} F_{i_d j} \log \frac{F_{j i_d}}{A_{i_d}} - A_i F_{ij} \log \frac{F_{ij}}{A_j}) \end{aligned} \tag{14}$$

where the right hand of the inequality corresponds to the mutual information increase in the discontinuity meshing case, where we have taken by hypothesis $F_{i_a j} = F_{i_d j}$.