Universitat Politècnica de Catalunya

Departament de Llenguatges i Sistemes Informàtics

An Information-Theory Framework for the Study of the Complexity of Visibility and Radiosity in a Scene

Memòria presentada per **Miquel Feixas i Feixas** a la Universitat Politècnica de Catalunya amb la finalitat d'obtenir el grau de **doctor en Informàtica**

Director: Mateu Sbert i Casasayas

Barcelona, octubre 2002

Aquesta tesi ha estat defensada davant del tribunal format per

Dr. Xavier Pueyo, president
Dr. Àlvar Vinacua, secretari
Dr. Werner Purgathofer
Dr. László Neumann
Dr. Philippe Bekaert

Agraïments

Dono les gràcies a tothom que m'ha ajudat en aquesta tesi i, d'una manera especial, a la gent que hi ha tingut una implicació personal. Abans que res cal dir que sense en Mateu Sbert aquesta tesi no s'hagués dut a terme. La seva fe en el projecte des del primer moment i la seva capacitat de direcció i discussió han estat fonamentals per tirar endavant. En aquesta mateixa línia, la col.laboració, el treball i l'empenta d'en Jaume Rigau en l'exploració de les possibilitats de les mesures de la teoria de la informació ha estat un ajut inestimable. Fruit d'aquest treball conjunt han aparegut noves línies de recerca.

A tots els altres coautors, els he de donar també les gràcies. A l'Esteve del Acebo, perquè gràcies a ell va sorgir la idea de calcular l'entropia d'una escena! A en Philippe Bekaert, amb qui ha estat un plaer col.laborar i amb qui hem passat molt bones estones de discussió. Els seus comentaris i suggerències han estat sempre valuosos. A en Pere Pau Vàzquez, que des del primer moment va creure en el nostre treball i en possibles aplicacions a altres camps. A en Francesc Castro, pel seu interès i suport incondicional.

Als germans László i Attila Neumann que estan sempre disposats a oferir suggerències i obrir nous camins. A en László Szirmay-Kalos, a qui he d'agrair els seus consells. I a en Joaquim Gelabertó, jubilat de la seva feina a l'institut però no de la seva passiò per la ciència i el coneixement en general, que amb la seva àmplia i interdisciplinar saviesa, tan difícil de trobar avui, sempre ha estat disposat a llegir i comentar qualsevol dels nostres treballs.

Al grup de gràfics de Girona (GGG), pel seu suport i disponibilitat, i en especial al seu cap, en Xavier Pueyo. A la resta de companys de LSI i del departament, que sempre m'han ofert els seus ànims. A la Mercè i en Jordi (administratius) i a en Robert i en Joan (system managers), per fer-nos les coses més senzilles.

Permeteu-me també unes notes entranyables sobre les històries que han precedit el treball actual. Voldria deixar constància d'un agraïment molt especial per en Xavier Messeguer de LSI de la UPC, qui em va iniciar en l'aventura d'estudiar teoria de la informació i intentar aplicar-la a les estructures de dades. Li agraeixo la seva confiança. Paral.lelament, amb l'Albert Bramon i en Lluís Ametller, físics i amics banyolins des de fa molts anys, juntament amb en Joaquim Gelabertó, vàrem passar unes estones excel.lents aplicant la teoria de la informació a problemes d'informació quàntica, sempre amb un objectiu de fons, que ho és també d'aquesta tesi, maximitzar la transferència d'informació.

A en Jaume Rigau, en Joaquim Gelabertó i l'Imma Boada pels seus comentaris a algunes parts d'aquesta tesi. A la Corine Salnikov i al Servei de Llengües Modernes de la UdG pel seu ajut amb l'anglès.

També vull donar les gràcies a la meva tutora, la Isabel Navazo, que sempre m'ha ofert tot l'ajut necessari. I als doctors Xavier Pueyo, Àlvar Vinacua, Werner Purgathofer, László Neumann i Philippe Bekaert per haver acceptat ser membres del tribunal que ha de jutjar aquesta tesi.

Finalment, vull agrair el suport que en tot moment he tingut dels amics i de la família. I d'una manera especial de la Lluïsa, en Ferran i en Guillem, que han sofert directament les consequències de la meva dedicació a aquest treball i que sense el seu amor, comprensió i ajut no hagués pogut tirar endavant.

The work that has led to this thesis has been funded in part by grants from the Catalan Government (1999-SGR-00162, 2001-SGR-00296 and Catalan-Flemish joint action ABM/acs/ACI98-19) and from the Spanish Government (TIC98-0973-C03-03, TIC2001-2416-C03-01 and Spanish-Austrian joint actions HU1998-0015 and HU2000-0011).

A la Lluïsa, en Ferran i en Guillem

Contents

Chapter 1

Introduction

Light is the source of all excellence.
Antoni Gaudí (1852–1926)

All parts of creation are linked together and interchange their influences. The balanced rhythm of the universe is rooted in reciprocity.
Paramahansa Yogananda, "Autobiography of a Yogi", 1946

What does one do with information? What does one do with ignorance?
Murray Gell-Mann and Seth Lloyd [32]

In this thesis, *information theory* tools are applied to computer graphics in order to quantify the *complexity* of a *scene* from the points of view of *visibility* and *radiosity*. In this chapter, we discuss a possible relationship between radiosity, complexity and information theory. After that, our objectives and motivations are presented. Finally, we give an overview of this dissertation.

1.1 Radiosity, Complexity, and Information Theory

The three fundamental pillars of this thesis are radiosity, complexity, and information theory:

Radiosity

One of the most important topics in computer graphics is the *accurate computation of the global illumination in a closed virtual environment (scene)*, i.e. the intensities of light over all its surfaces. "The production of realistic images requires in particular a *precise* treatment of lighting effects that can be achieved by *simulating* the underlying physical phenomena of light emission, propagation, and reflection"[82]. This type of simulation is called global illumination and is represented by the *rendering equation* [43], which is a Fredholm integral equation of the second kind. However obtaining an exact representation of the illumination is an intractable problem. Many different techniques are used to obtain an *approximate* quantification of it [12, 82, 33].

In this thesis, we deal with global illumination using the *radiosity* method, which only considers diffuse surfaces, where reflected light does not depend on the outgoing direction. The radiosity method consists of the following steps: discretisation of the surfaces of the scene into a mesh of polygons called *patches*, computation of *form factors*, solution of the system of linear equations, and visualization of the solution [12, 82, 33]. The form factor F_{ij} between patches i and j expresses the fraction of energy leaving patch i which goes directly to patch j, and it can be interpreted as the *visibility* between patches. The kernel of the radiosity equation depends only on the visibility between the points or patches of a scene, i.e., the geometry of the scene [34, 11, 12, 82, 33]. The main problems of the radiosity method are *meshing* and *form factor computation*. Scene meshing not only has to accurately represent illumination variations, but it also has to avoid unnecessary subdivisions of the surfaces that would increase the number of form factors to be computed, and consequently the computational time. The best strategy tries to balance accuracy and computational cost. This dissertation looks at these problems from an information-theory approach.

Complexity

In the last two decades, the study of *complexity* has become a very active research area in many different fields (automata, information theory, computer science, physics, biology, neuroscience, etc.) [3]. It is generally accepted that "the problem of characterizing complexity in a *quantitative* way is a vast and rapidly developing subject. Although *various interpretations* of the term have been advanced in different disciplines, no comprehensive discussion has yet been attempted"[3]. But, what is complexity? The definition found in Webster's dictionary summarizes in a few words the notion of complexity: "A complex object is an arrangement of parts, so intricate as to be hard to understand or deal with" (Webster, 1986).

From the points of view of visibility and radiosity, a scene (or the simulation of the light propagation in it) shows typical characteristics of complex behaviour [3]: "simultaneous presence of elements of *order* and *disorder*, some degree of *unpredictability*, interactions between subsystems which change in *dependence* on how the system is subdivided". As we have seen, global illumination simulates the interreflection of light *between all the surfaces* in an environment. The difficulty in obtaining an accurate solution mainly depends on the degree of dependence between all these surfaces. This dissertation introduces a complexity measure called mutual information which quantifies the degree of interdependence or structure in a scene and which will be used to obtain new refinement criteria.

Information Theory

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. [15]. For instance, it is applied successfully in areas closely related to computer graphics, such as medical image processing (see for instance [95, 85, 64]) or computer vision and robot motion (see for instance [94, 2]). *Information* is simply the outcome of a selection among a finite number of possibilities [93]. The classical

measure of information, *Shannon entropy*, expresses the *information* content or *uncertainty* of a random variable. It is also a measure of the *variation*, *dispersion*, or *diversity* of a probability distribution of observed events. Another measure, *mutual information*, expresses the *information transfer* in a communication channel. It is also a measure of the *dependence* or *correlation* between two random variables. *Both measures capture different aspects of the complexity of a scene*: entropy measures its degree of randomness or uncertainty and mutual information quantifies its degree of structure or correlation [1].

1.2 Objectives and Motivations

This thesis aims to quantify the complexity of a scene and to obtain new criteria of refinement for visibility and radiosity. Mutual information and entropy are used to reach these objectives.

But, how can we apply information theory to the study of a scene? Basically, because a scene *contains information* which is *exchanged* between its different parts (points or patches), thus creating a *dependence* or *correlation* between them. "When a photon is emitted from a light source and then strikes an object, that photon has effected the transfer of some *information*. Minimally, it represents that a certain amount of energy of a specific quantity and quality has been transferred from one object to the other. But as we have seen, it also tells us something about the relative visibility of the two points, and the amount of impact that the light source will have on the final image" [33]. In this thesis, a scene will be analyzed from an information-based point of view, and we consider that information is a purely probabilistic concept. Obviously, the visibility and radiosity of a scene depends on its structure. Thus, for instance, the variation of the position of the objects of a scene changes the degree of interaction among all the parts of the environment, and consequently the information transfer between these parts also varies.

In relation to the requirements of the radiosity method this dissertation contributes in the following areas:

- Scene classification according to two different and complementary complexity measures, randomness (entropy) and correlation (mutual information), which quantify how difficult it is to compute the visibility and radiosity in a scene.
- Evaluation of the goodness of a particular discretisation and, consequently, also of the strategy used to obtain it.
- Obtaining new refinement criteria based on the information transfer in a scene.

In conclusion, this thesis studies scene visibility and radiosity from an information-theory standpoint and shows the feasibility of using information theory in dealing with a scene. To end this introduction, it is worth remembering Glassner's words in the last chapter of *Principles of Digital Image Synthesis*:

[1] Although we consider that entropy and mutual information express different perspectives of the complexity of a scene, entropy will be referred to as scene randomness and mutual information as scene complexity (see chapter 4).

"There has not been much attention paid to applying information theory to image synthesis. I think that it holds promise, though, and may help us design new types of efficient rendering algorithms"[33]. This work is a first step in the application of information theory to global illumination. In the same way that information theory has been applied successfully in many other areas, we sincerely hope that information theory in global illumination will also be of service.

1.3 Overview

This dissertation is organized into the following chapters:

- **Chapter 2: Previous Work**

  Basic concepts of the radiosity method, information theory, and complexity are reviewed.

- **Chapter 3: Scene Visibility Entropy**

  A discretised scene is represented by a discrete information channel. This fact enables us to apply the definitions of entropy and mutual information to the study of the scene visibility. The relationship between entropy and Monte Carlo error in the form factor computation is shown.

- **Chapter 4: Scene Visibility Complexity**

  We first analyze the concept of scene complexity. Then, the continuous mutual information is proposed as the scene visibility complexity, and the discrete mutual information as the visibility complexity in a discretised scene. Continuous mutual information integral is solved by Monte Carlo integration and it is computed efficiently by casting uniformly distributed global or local lines. The difference between continuous and discrete mutual information will enable us to evaluate the goodness of a given discretisation. Finally, we present a tentative scene classification in flatland.

- **Chapter 5: Scene Radiosity Entropy and Complexity**

  To study the complexity of a scene with illumination, we need to find an analog of the form factor matrix for the radiosity setting. An analogy appears naturally when the null variance probability transition matrix is considered. Similarly to the two previous chapters, the entropy and mutual information of a scene are defined by using both discrete and continuous Markov chains. A general proposition, which can be applied to visibility, radiosity, and importance, provides us with the increase in mutual information transfer when a patch is refined.

- **Chapter 6: Refinement Criteria for Visibility and Radiosity**

  From an information theory point of view, we know that between different discretisations of the same scene the most precise will be the one that has a minimum loss of information transfer. This proposal provides the basis for dealing with the discretisation error and for introducing some criteria of refinement applied to visibility and radiosity.

- **Chapter 7: Conclusions and Future Work**

  The conclusions and the main contributions of this thesis are presented, as well as some indications about our current and future research.

Chapter 2

Previous work

As stated in the introduction, this thesis deals with the application of information-theory concepts to the study of the complexity of visibility and radiosity in a scene. We review next the radiosity method, the basic definitions of information theory, and the different perspectives of the complexity of a system.

2.1 Radiosity Method

The radiosity method, first introduced in [34, 56, 11], 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.

2.1.1 Rendering Equation

(a) (b)

Figure 2.1: Two different illuminated scenes.

The light transport in a virtual closed environment or scene (Figure 2.1) is described by the *rendering equation* [43], 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:

$$L(x, \omega_x) = L_e(x, \omega_x) + \int_S \rho_{bd}(x, \omega_x, -\omega_{y \to x}) L(y, \omega_y) G(x, y) dA_y \tag{2.1}$$

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 2.2)

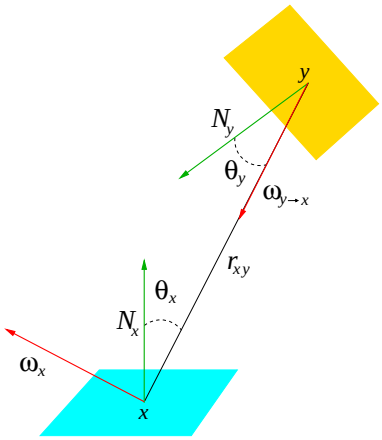
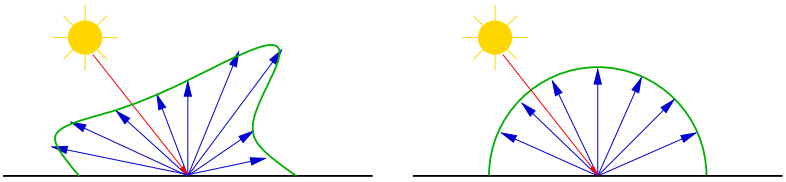


Figure 2.2: Outgoing and incoming directions at point x.

- $L(x, \omega_x)$ 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 2.3a)
- $G(x, y)$ 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



(a) (b)

Figure 2.3: (a) Bidirectional reflectance distribution function. (b) Diffuse reflectance.

This equation can be presented in slightly different forms (global illumination equation, radiance equation)[82, 33]. 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"[33].

2.1.2 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 2.3b). From this simplification, the *rendering equation for diffuse surfaces* can be expressed as

$$L(x) = L_e(x) + \int_S \rho_{bd}(x) L(y) G(x, y) dA_y \tag{2.2}$$

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) [82, 33]:

$$B(x) = \int_{\Omega_x} L(x) \cos\theta_x d\omega_x = \pi L(x) \tag{2.3}$$

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 2.3b).

The *radiosity equation* is then obtained by multiplying both sides of equation (2.2) by π:

$$B(x) = E(x) + \frac{\rho(x)}{\pi} \int_S B(y) G(x, y) dA_y \tag{2.4}$$

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)

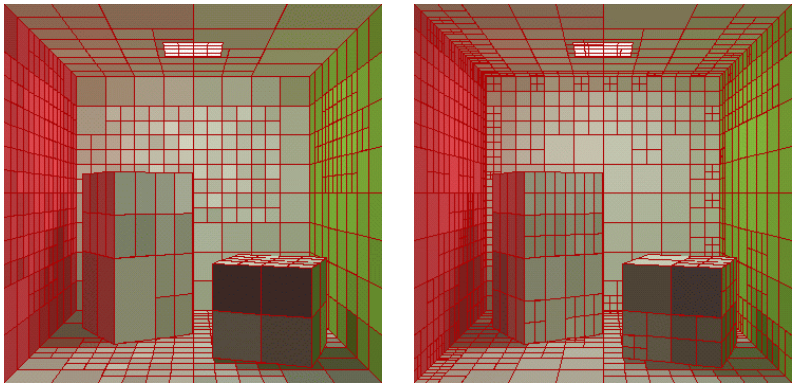
The radiosity equation can also be written in a directional form [82]:

$$B(x) = E(x) + \frac{\rho(x)}{\pi} \int_{\Omega_x} B(y) \cos\theta_x d\omega_x \tag{2.5}$$

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

2.1.3 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 2.4).



(a) (b)

Figure 2.4: A scene with two different discretisations.

With these assumptions, the integral equation (2.4) becomes the system of radiosity equations [34]:

$$B_i = E_i + \rho_i \sum_{j=1}^{n_p} F_{ij} B_j \tag{2.6}$$

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:

$$\begin{aligned} F_{ij} &= \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{G(x,y)}{\pi}dA_y dA_x \\ &= \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2}V(x,y)dA_y dA_x \\ &= \frac{1}{A_i}\int_{A_i}\int_{A_j}F(x,y)dA_y dA_x \\ &= \frac{1}{A_i}\int_{A_i}\int_{\Omega_{x\to j}}\frac{\cos\theta_x}{\pi}V(x,y)d\omega_x dA_x \end{aligned} \tag{2.7}$$

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
$$A_i F_{ij} = A_j F_{ji} \quad \forall i,j \tag{2.8}$$
- Energy conservation
$$\sum_{j=1}^{n_p} F_{ij} = 1 \quad \forall i \tag{2.9}$$
- Additivity
$$F_{i(k\cup l)} = F_{ik} + F_{il} \tag{2.10}$$
where i, k, and l are three disjoint patches. In general the reverse is not true
$$F_{(k\cup l)i} \neq F_{ki} + F_{li} \tag{2.11}$$
In fact, if the patch i is divided into n_i subpatches, we obtain
$$\sum_{k=1}^{n_i} A_{i_k} F_{i_k j} = A_i F_{ij} \tag{2.12}$$
As a direct consequence of this equation, if patch i is divided into n_i subpatches of equal area, we have
$$n_i F_{ij} = \sum_{k=1}^{n_i} F_{i_k j} \tag{2.13}$$
or
$$F_{ij} = \frac{1}{n_i}\sum_{k=1}^{n_i} F_{i_k j} \tag{2.14}$$
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 (2.7) can be considered as an average over the area of patch i of the inner integral. Thus, we have

$$F_{ij} = \frac{1}{A_i} \int_{A_i} F_{dA_x,A_j} dA_x \tag{2.15}$$

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

$$\begin{aligned} F_{dA_x,A_j} &= \int_{A_j} \frac{\cos\theta_x \cos\theta_y}{\pi r_{xy}^2} V(x,y) dA_y \\ &= \int_{\Omega_{x\to j}} \frac{\cos\theta_x}{\pi} V(x,y) d\omega_x \end{aligned} \tag{2.16}$$

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

$$\begin{aligned} F_{ij} &= \frac{1}{A_i} \int_{x\in A_i} F_{dA_x,A_j} dA_x \\ &\approx F_{dA_c,A_j} \frac{1}{A_i} \int_{x\in A_i} dA_x = F_{dA_c,A_j} \end{aligned} \tag{2.17}$$

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 [82]. We define $F_j(x) \equiv F_{dA_x,A_j}$.

2.1.4 Radiosity Equation in Flatland

Visibility and radiosity in flatland have been studied in [40, 57]. The *2D continuous radiosity equation* for the illumination in a diffuse environment can be written in the form

$$B(x) = E(x) + \rho(x) \int_L B(y) \frac{\cos\theta_x \cos\theta_y}{2r_{xy}} V(x,y) dL_y \tag{2.18}$$

where

- $B(x)$, $E(x)$, and $\rho(x)$ at point x are, respectively, the radiosity, the emittance, and the reflectance
- L is the set of segments that form the environment
- x and y are points on L
- dL_y is a differential length at point y
- r_{xy} is the distance between x and y
- $V(x,y)$ is the visibility flag between x and y
- θ_x and θ_y are the angles which the normals at x and y form with the line joining them

- $F(x,y) = \frac{\cos\theta_x \cos\theta_y}{2r_{xy}} V(x,y)$ is the point-to-point form factor between x and y

The discrete radiosity equation in flatland has the same form as the one for the 3D case (2.6):

$$B_i = E_i + \rho_i \sum_{j=1}^{n_p} F_{ij} B_j \tag{2.19}$$

The form factor F_{ij} is given by

$$F_{ij} = \frac{1}{L_i} \int_{L_i} \int_{L_j} \frac{\cos\theta_x \cos\theta_y}{2r_{xy}} V(x,y) dL_x dL_y \tag{2.20}$$

where L_i and L_j represent, respectively, the segments i and j and also their respective lengths, and x and y are, respectively, points on L_i and L_j. The form factors also fulfil the following properties:

$$L_i F_{ij} = L_j F_{ji} \quad \forall i,j \tag{2.21}$$

$$\sum_{j=1}^{n_p} F_{ij} = 1 \quad \forall i \tag{2.22}$$

2.1.5 Power Equation

Another form of the radiosity equation is the *power equation*, obtained by multiplying both sides of equation (2.6) by A_i and applying the reciprocity relation (2.8):

$$P_i = \Phi_i + \rho_i \sum_j F_{ji} P_j \tag{2.23}$$

where

- $P_i = B_i A_i$ and $P_j = B_j A_j$ are, respectively, the total powers emitted by patches i and j (W)
- $\Phi_i = E_i A_i$ is the self-emitted power of patch i (W)

From this power equation, the form factor can also be physically interpreted as the *fraction of power leaving patch i which goes directly to patch j*. Moreover, considering that light propagates in straight lines, we could give another definition: the form factor is the *fraction of lines that, exiting patch i, arrive at patch j* (see the next section).

2.1.6 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 [76] solved the polygon-to-polygon case. In [81, 33] 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 [1].

From Figure 2.5, 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

$$\begin{aligned} F_{ij} &= \frac{1}{A_i}\int_{A_i}\int_{A_j}\frac{\cos\theta_x\cos\theta_y}{\pi r_{xy}^2}V(x,y)dA_x dA_y \\ &= \frac{1}{\pi A_i}\int_{A_i}\int_{A_j}\frac{\cos\theta_x}{r_{xy}}\frac{\cos\theta_y}{r_{xy}}V(x,y)dA_x dA_y \\ &= \frac{1}{4\pi R^2 A_i}\int_{A_i}\int_{A_j}dA_x dA_y = \frac{A_j}{A_S} \end{aligned} \tag{2.24}$$

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.

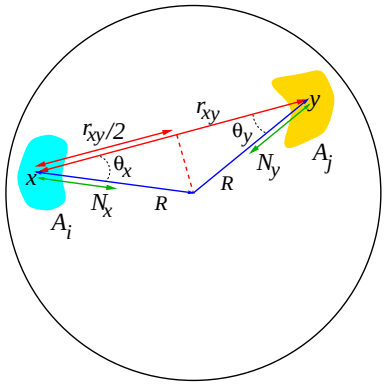


Figure 2.5: Geometry for the form factor between two spherical patches.

When occlusions between patches exist, we can use deterministic numerical approximations to compute the form factors. Different methods can be found in the literature [82, 33]: Wallace's method, Nusselt's analogy, hemi-cube method.

Monte Carlo integration

In this section, the form factor integral (2.7) will be evaluated by the Monte Carlo method. So, we will give a brief overview of this technique. For a more detailed description, see [44]. Also [82, 33, 4] review it and give different techniques to sample a random variable.

[1] We will see in the next chapter that the interior of a sphere presents very special properties from an information-theory point of view.

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

$$I = \int_D g(x)dx = \int_D \frac{g(x)}{f(x)} f(x)dx \tag{2.25}$$

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 (2.25) can be read as the expected value of the random variable $\frac{g(X)}{f(X)}$ with respect to the pdf $f(x)$:

$$I = E_f \left[\frac{g(X)}{f(X)} \right] \tag{2.26}$$

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:

$$I \approx \widehat{I} = \frac{g(x_1)}{f(x_1)} \tag{2.27}$$

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

$$V[\widehat{I}] = E\left[\left(\frac{g(X)}{f(X)} \right)^2 \right] - \left(E\left[\frac{g(X)}{f(X)} \right] \right)^2 = \int_D \frac{g(x)^2}{f(x)} dx - I^2 \tag{2.28}$$

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 $\widehat{I_N}$

$$I \approx \widehat{I}_N = \frac{1}{N} \sum_{k=1}^{N} \frac{g(x_k)}{f(x_k)} \tag{2.29}$$

with variance

$$V[\widehat{I_N}] = \frac{1}{N} V[\widehat{I}] = \frac{1}{N} \left(\int_D \frac{g(x)^2}{f(x)} dx - I^2 \right) \tag{2.30}$$

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 [44].

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 (2.30) 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

$$MSE(\widehat{\theta}) = E\left[\left(\widehat{\theta} - \theta \right)^2 \right] \tag{2.31}$$

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 [4]). 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 (2.7)

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

  we take random points x and y on patches i and j respectively (Figure 2.6). 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

  $$\widehat{F_{ij}^1} = \frac{1}{A_i}\frac{F(x,y)}{f(x,y)} = \frac{1}{A_i}\frac{F(x,y)}{\frac{1}{A_iA_j}} = A_jF(x,y) \tag{2.32}$$

  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

  $$V[\widehat{F_{ij}^1}] = \int_{A_i}\int_{A_j}(A_jF(x,y))^2\frac{1}{A_iA_j}dA_xdA_y - F_{ij}^2 \tag{2.33}$$

  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 [4].

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

  $$\widehat{F_{ij}^1} = A_j\frac{1}{N}\sum_{k=1}^{N}F(x_k,y_k) \tag{2.34}$$

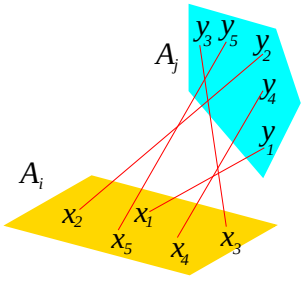


Figure 2.6: Form factor F_{ij} can be computed by taking random points on patches i and j.

- **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 (2.7)

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

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 $V[X] = p(1-p)$ [58]. Thus,

$$V[\widehat{F_{ij}^2}] \quad = \quad F_{ij}(1 - F_{ij}) \tag{2.35}$$

A secondary estimator for F_{ij} is given by

$$\widehat{F_{ij}^2} = \frac{N_{ij}}{N_i} \tag{2.36}$$

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 2.7a).

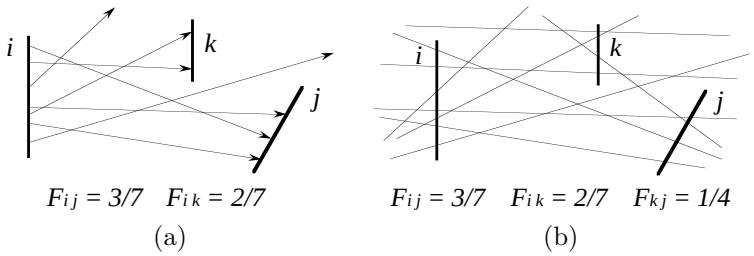


Figure 2.7: Form factors can be computed with local (a) and global lines (b).

  - **Global lines**

Global lines [72] 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 2.7b). It can be shown that each line can contribute to the computation of several form factors (Figure 2.8).

Also, it is important to note that, from integral geometry [71, 73], 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

$$\widehat{F_{ij}^3} = \frac{N_{ij}}{N_i} \tag{2.37}$$

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

$$V[\widehat{F_{ij}^3}] = \frac{1}{N_i} F_{ij}(1 - F_{ij}) \tag{2.38}$$

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 [73].

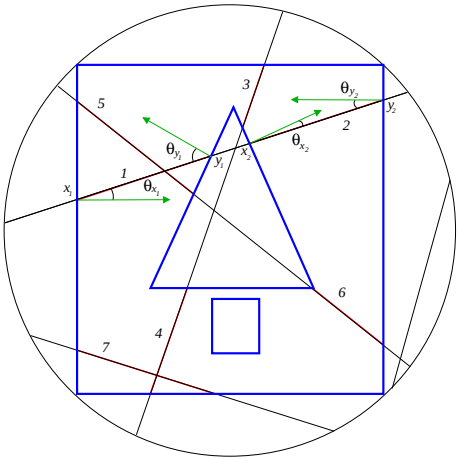


Figure 2.8: Each segment of a global line contributes to the computation of two form factors. Thus, the depicted line is used in four form factor computations.

2.1.7 Solution to the Radiosity Equation

The classic radiosity method consists of the following steps [12, 82, 33]:

- 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 [82].

The radiosity equation (2.6), which refers to a single patch, can be expressed as a system of n_p linear equations:

$$\begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_{n_p} \end{pmatrix} = \begin{pmatrix} E_1 \\ E_2 \\ \vdots \\ E_{n_p} \end{pmatrix} + \begin{pmatrix} \rho_1 F_{11} & \rho_1 F_{12} & \dots & \rho_1 F_{1n_p} \\ \rho_2 F_{21} & \rho_2 F_{22} & \dots & \rho_2 F_{2n_p} \\ \vdots & \vdots & \vdots & \vdots \\ \rho_{n_p} F_{n_p 1} & \rho_{n_p} F_{n_p 2} & \dots & \rho_{n_p} F_{n_p n_p} \end{pmatrix} \begin{pmatrix} B_1 \\ B_2 \\ \vdots \\ B_{n_p} \end{pmatrix}$$

This linear system can be written in the form

$$\mathbf{B} = \mathbf{E} + \mathbf{R}\mathbf{B} \tag{2.39}$$

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 [2] 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:

$$\mathbf{B} = \mathbf{E} + \mathbf{R}\mathbf{E} + \mathbf{R}^2\mathbf{E} + \cdots + \mathbf{R}^n\mathbf{E} + \cdots \tag{2.40}$$

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 [82, 33] are available to solve the radiosity and power systems: Jacobi relaxation, Gauss Seidel relaxation, Southwell relaxation, and also their respective stochastic versions [80, 82, 54, 53, 55, 4].

2.1.8 Random Walks and Markov Chains

A discrete *random walk* [44, 70] is a Monte Carlo technique used to solve linear systems of equations like (2.39). A *random walk* in a scene can be considered as a *Markov chain* [13, 15, 52]. This is a discrete stochastic process defined over a set of states $S = 1, 2, \dots, n$ 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 $X_k, k = 0 \dots \infty$, 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*

[2] The norm of $\mathbf{R}$ can be defined by $\|\mathbf{R}\| = \max_i(\sum_j |r_{ij}|)$.

$\mathbf{w} = \{w_1, \ldots, w_n\}$ 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 [13]. 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 explicitely [80, 60, 19, 74].

For the purpose of this thesis we are mainly interested in the following two properties [73]:

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

$$\lim_{m\to\infty} (\mathbf{F}^m)_{ij} \to \frac{A_j}{A_T} = a_j \tag{2.41}$$

   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.

2.1.9 Importance Equations

Given a system of linear equations

$$\mathbf{x} = \mathbf{H}\mathbf{x} + \mathbf{a} \tag{2.42}$$

we can always associate with it a family of adjoint systems

$$\mathbf{y} = \mathbf{H}^T\mathbf{y} + \mathbf{b} \tag{2.43}$$

where $\mathbf{H}^T$ is the transpose of $\mathbf{H}$.

The solutions $\mathbf{x} = (1-\mathbf{H})^{-1}\mathbf{a}$ and $\mathbf{y} = (1-\mathbf{H}^T)^{-1}\mathbf{b}$ fulfil $< \mathbf{x}, \mathbf{b} >=< \mathbf{y}, \mathbf{a} >$, where $<,>$ represents a scalar product.

The adjoint system corresponding to the radiosity equation (2.6) is given by

$$J_i = W_i + \sum_{j=1}^{n_p} \rho_j F_{ji} J_j \tag{2.44}$$

This system can be solved with a *shooting* method, such as Southwell relaxation.

The adjoint system corresponding to the power equation (2.23) is given by

$$I_i = V_i + \sum_{j=1}^{n_p} \rho_j F_{ij} I_j \tag{2.45}$$

This system can be solved with a *gathering* method, such as Jacobi or Gauss-Seidel relaxations. Multiplying both sides of this equation by A_i we obtain (2.44), where $W_i = A_i V_i$ and $J_i = A_i I_i$. The quantities W_i, J_i, V_i, and I_i are called *importances* (see [83, 61, 62]).

2.1.10 Random Walk Radiosity

Monte Carlo literature describes methods to solve a system of linear equations using a random walk [38, 70]. The basis of these methods is to consider each unknown as an state and to build a random walk between the different states, with transition probabilities related to the coefficients of the matrix of the system. Two kinds of solutions are used: *direct* solutions, which build random paths from the states of interest and only serve to evaluate those states, and *adjoint* solutions, which begin at the states with the corresponding independent term different to zero. In the radiosity setting, the *direct* solutions correspond to *gathering* (from the patches) solutions, and the *adjoint* solutions correspond to *shooting* (from the sources) solutions. As probability transition matrix we usually take the form factor matrix [74, 4]. Other importance solutions are obtained using different transition matrices than the form factor.

For instance, it can be found [75, 4] that in gathering random walk radiosity the transition probabilities with null variance are

$$P_{ij} = \frac{\rho_i F_{ij} B_j}{B_i - E_i} \tag{2.46}$$

and in shooting random walk radiosity the null variance transition probabilities are given by

$$P_{ij} = \frac{\rho_j F_{ij} I_j}{I_i - V_i} \tag{2.47}$$

2.1.11 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 [82].

The hierarchical refinement algorithm was first introduced in [39] by Hanrahan and Salzman. Additional information can be found in [82, 33]. 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?"[33]. 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" [33]. 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 [33].

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

2.1.12 Refinement Criteria

In this section, we review some refinement criteria for hierarchical radiosity [3]. The cheapest and most widely-used oracle has been the power-based oracle [39]. 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 [83, 35, 50, 59, 49, 6, 84, 41]. 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 [4].

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.

Oracle based on transported power

Hierarchical refinement radiosity was initially presented for constant radiosity approximations by Hanrahan et al. [39]. 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. [39] 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

$$\rho_i A_i F_{ij} B_j < \epsilon \tag{2.48}$$

Other strategies [87, 30] 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 (2.48) 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 [83], the refinement criterion is given by

$$\rho_i (F_{ij}^{max} - F_{ij}^{min}) A_j B_j < \epsilon \tag{2.49}$$

[3] This section follows closely the discussion in [4]

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 [4] 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 [35] in order to drive hierarchical refinement with higher-order approximations. When applied to constant approximations, the refinement criterion is given by

$$\rho_i max(F_{ij}^{max} - F_{ij}^{av}, F_{ij}^{av} - F_{ij}^{min}) A_j B_j < \epsilon \tag{2.50}$$

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 [4].

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. [50] for constant approximations. Pattanaik and Bouatouch [59] proposed a similar strategy for linear basis functions. Other approaches are given in [49, 6, 84, 41]. The computation cost of kernel and radiosity smoothness-based oracles has not yet been found to compensate for the gain in mesh quality [4].

2.2 Information Theory

In 1948, Claude Shannon published "A mathematical theory of communication" [78] which marks the beginning of information theory. In this paper, he defined measures such as entropy and mutual information [5], and introduced the fundamental laws of data compression and transmission.

In this section, we present some basic concepts of information theory. A very good reference is the text by Cover and Thomas [15]. Other main references used in this thesis are Blahut [8] and Lubbe [88].

2.2.1 Entropy

In [78], 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?".

[4] In this thesis, point-to-point form factor $F(x,y)$ is also referred to as the radiosity kernel value.

[5] In Shannon's paper, the mutual information is called rate of transmission.

His answer is: "Suppose we have a *set of possible* events whose *probabilities of occurrence* are $p_1, p_2, \ldots, p_n$. 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 2.9.

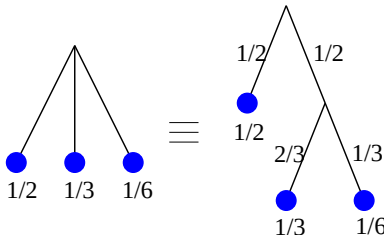


Figure 2.9: Grouping property of the entropy.

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:

$$H = -K \sum_{i=1}^{n} p_i \log p_i \tag{2.51}$$

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 [15].

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

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{2.52}$$

where $n = |S|$, $p_i = Pr[X = x_i]$ 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 $H(\mathbf{p})$ for the entropy, where $\mathbf{p}$ 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 [78] 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 $S' = \{y_1, y_2, \ldots, y_m\}$, the *joint entropy* of X and Y is defined as

$$H(X, Y) = -\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} \log p_{ij} \tag{2.53}$$

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

When $n = 2$, the *binary* entropy (Figure 2.10) is given by

$$H(X) = -p \log p - (1 - p) \log(1 - p) \tag{2.54}$$

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

Also, the *conditional entropy* is defined as

$$H(X|Y) = -\sum_{j=1}^{m} \sum_{i=1}^{n} p_{ij} \log p_{i|j} \tag{2.55}$$

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

The Bayes theorem expresses the relation between the different probabilities:

$$p_{ij} = p_i p_{j|i} = q_j p_{i|j} \tag{2.56}$$

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:

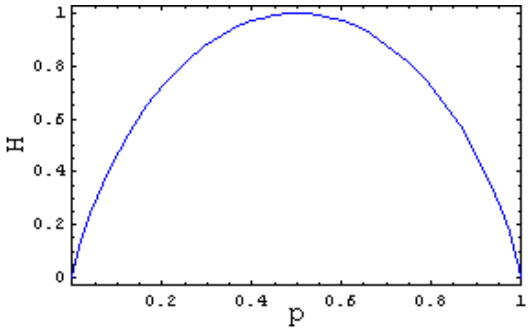


Figure 2.10: Binary entropy.

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

2.2.2 Mutual Information

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

$$\begin{aligned} I(X,Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= -\sum_{i=1}^{n} p_i \log p_i + \sum_{j=1}^{m}\sum_{i=1}^{n} p_{ij} \log p_{i|j} \\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} p_{ij} \log \frac{p_{ij}}{p_i q_j} \end{aligned} \tag{2.57}$$

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 2.11.

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

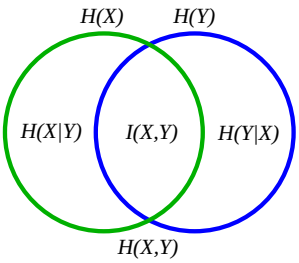


Figure 2.11: Venn diagram of a discrete channel.

as

$$D_{KL}(\mathbf{p}\|\mathbf{q}) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i} \tag{2.58}$$

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 $\mathbf{q}$ when the true distribution is $\mathbf{p}$" [15].

The relative entropy satisfies the *information inequality* $D_{KL}(\mathbf{p}\|\mathbf{q}) \geq 0$, with equality only if $\mathbf{p} = \mathbf{q}$. 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

$$I(X, Y) = D_{KL}(\{p_{ij}\}\|\{p_i q_j\}) \tag{2.59}$$

2.2.3 Entropy Rate of a Markov Chain

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

$$H(X_1, \ldots, X_n) = H(X_1) + H(X_2|X_1) + \ldots + H(X_n|X_{n-1}, \ldots, X_1) \tag{2.60}$$

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

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1}, \ldots, X_1) \end{aligned} \tag{2.61}$$

representing the *average information content* per output symbol [6] [15]. It is the "uncertainty associated with a 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 [28].

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

$$H(X_n|X_1, X_2, \ldots, X_{n-1}) = H(X_n|X_{n-1}) \tag{2.62}$$

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

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

$$\begin{aligned} h &= \lim_{n\to\infty} \frac{1}{n} H(X_1, X_2, \ldots, X_n) \\ &= \lim_{n\to\infty} H(X_n|X_{n-1}) \\ &= H(X_2|X_1) = -\sum_{i=1}^{n} w_i \sum_{j=1}^{n} P_{ij} \log P_{ij} \end{aligned} \tag{2.63}$$

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* [16, 36, 79, 86] of an infinite chain is defined by

$$E = \lim_{n\to\infty} (H(X_1, X_2, \ldots, X_n) - nh) \tag{2.64}$$

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" [27].

2.2.4 Important Inequalities

Some of the above properties can be deduced from the following inequalities [15]. In addition, these will also play an important role in obtaining fundamental results in this thesis.

Jensen's inequality

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

$$f(\lambda x_1 + (1-\lambda)x_2) \leq \lambda f(x_1) + (1-\lambda) f(x_2) \tag{2.65}$$

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 [15].

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

$$f(E[X]) \leq E[f(X)] \tag{2.66}$$

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 $D_{KL}(\mathbf{p}\|\mathbf{q}) \geq 0$. 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.

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

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq (\sum_{i=1}^{n} a_i) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{2.67}$$

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. $D_{KL}(\mathbf{p}\|\mathbf{q})$ is convex in the pair (p, q)
2. $H(X)$ is a concave function of $\mathbf{p}$
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

$$I(X, Y) \geq I(X, Z) \tag{2.68}$$

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 $H(X|Y)$ [15, 20]. 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 [15].

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

$$P_e = Pr[\widehat{X} \neq X] \tag{2.69}$$

Fano's inequality:

$$H(P_e) + P_e \log n \geq H(X|Y) \tag{2.70}$$

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

This inequality can be weakened to

$$P_e \geq \frac{H(X|Y) - 1}{\log n} \tag{2.71}$$

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

2.2.5 Entropy and Coding

Other ways of interpreting the Shannon entropy are possible:

- As we have seen in section 2.2.1, $-\log p_i$ 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 [27].

- Entropy is also related to the difficulty in guessing the outcome of a random variable. Thus, it can be seen [15, 27] that

  $$H(X) \leq \overline{questions} < H(X) + 1 \tag{2.72}$$

  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 [15, 27]. "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" [15]. For instance, Huffman instantaneous coding [7] is optimal and fulfils the following theorems:

  - Similarly to (2.72), we have

    $$H(X) \leq \overline{\ell} < H(X) + 1 \tag{2.73}$$

    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:

    $$H(X) \leq \overline{\ell_n} < H(X) + \frac{1}{n} \tag{2.74}$$

    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 [15].

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

– For a stationary stochastic process, we have

$$\frac{H(X_1, X_2, \ldots, X_n)}{n} \leq \overline{\ell_n} < \frac{H(X_1, X_2, \ldots, X_n)}{n} + 1 \tag{2.75}$$

and thus, by definition of entropy rate h (2.61),

$$\lim_{n\to\infty} \overline{\ell_n} \to h \tag{2.76}$$

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.

2.2.6 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 $\mathbf{p}$ has been defined (2.52) as

$$H(X) = -\sum_{i=1}^{n} p_i \log p_i \tag{2.77}$$

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

$$H^c(X) = -\int_S p(x) \log p(x) dx \tag{2.78}$$

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

$$H^c(X|Y) = -\int_S \int_S p(x, y) \log p(x|y) dx dy \tag{2.79}$$

and the continuous mutual information is defined as

$$I^c(X, Y) = \int_S \int_S p(x, y) \log \frac{p(x, y)}{p(x)p(y)} dx dy \tag{2.80}$$

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 [15]), 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 [78, 15, 27]:

$$\lim_{\Delta\to 0} H(X^\Delta) = H^c(X) - \log \Delta \tag{2.81}$$

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:

$$\lim_{\Delta \to 0} I(X^{\Delta}, Y^{\Delta}) = I^{c}(X, Y) \tag{2.82}$$

In addition, Kolmogorov [45] and Pinsker [63] 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* [37]. This last property can also be deduced from the data processing inequality (2.68) [37].

2.3 Complexity

The study of complexity has multiple directions or objectives, and also many fields of application, which reflect the great activity in this area. In this section, we review the concept of complexity and different ways to quantify it. This summary is extracted from the works of Badii and Politi [3], Grassberger [36], James Crutchfield's group from the Santa Fe Institute (with David Feldman and Cosma Rohilla Shalizi) [29, 28, 77], Murray Gell-Mann [32], Seth Lloyd [51], and Wentian Li [48]. Other references are [97, 46].

2.3.1 What is Complexity?

Complexity is an active research area in many different fields. But, what is a complexity measure? W. Li's answer is: "The meaning of this quantity should be very close to certain measures of *difficulty* concerning the object or the system in question: the difficulty in constructing an object, the difficulty in describing a system, the difficulty in reaching a goal, the difficulty in performing a task, and so on" [48]. There are many definitions of complexity [3, 36, 48, 29, 1, 51] corresponding to the different ways of quantifying these difficulties and "there is not yet a consensus on a precise definition" [1].

"The concept of complexity is closely related to that of *understanding*, in so far as the latter is based upon the accuracy of *model* descriptions of the system obtained using condensed information about it" [3]. When defining complexity, three fundamental points ought to be considered [3]:

- "Understanding implies the presence of a *subject* having the task of describing the *object*, usually by means of model predictions."
- "The object, or a suitable representation of it, must be conveniently divided into *parts* which, in turn, may be further split into subelements, thus yielding a *hierarchy*. Notice that the hierarchy need not be manifest in the object but may arise in the construction of a model. Hence, the presence of an actual hierarchical structure is not an infallible indicator of complexity."
- "Having individuated a hierarchical encoding of the object, the subject is faced with the problem of studying the *interactions* among the subsystems

> and of incorporating them into a model. Considerations of the interactions at different levels of resolution brings in the concept of *scaling*. Does the increased resolution eventually lead to a stable picture of the interactions or do they escape any recognizable plan? And if so, can a different model reveal a simpler underlying scheme?"

In this same direction, we recall that among the typical characteristics of complex behaviour there are "simultaneous presence of elements of *order* and *disorder*, some degree of *unpredictability*, interactions between subsystems which change in *dependence* on how the system is subdivided"[3]. Note that all these requirements can be considered in a scene.

2.3.2 Complexity Measures

In this section we present a "non-exhaustive list" of complexity measures provided by Seth Lloyd [51]. This list is grouped under "three questions that researchers frequently ask to quantify the complexity of the thing (house, bacterium, problem, process, investment scheme) under study". In fact, "many measures of complexity represent variations on a few underlying themes":

- How hard is it to describe?[8]

  Information, entropy, algorithmic complexity or algorithmic information content, minimum description length, Fisher information, code length (prefix-free, Huffman, Shannon-Fano, error-correcting, Hamming), Chernoff information, dimension, fractal dimension, Lempel-Ziv complexity.

- How hard is it to create?[9]

  Computational complexity, time computational complexity, space computational complexity, information-based complexity, logical depth, thermodynamic depth, cost, cripticity.

- What is its *degree of organization*? This may be divided up into two quantities: *difficulty of describing organizational structure* and *amount of information shared between the parts of a system* as a result of this organizational structure.

  - *Effective complexity*

    Metric entropy, fractal dimension, excess entropy, stochastic complexity, sophistication, effective measure complexity, true measure complexity, topological epsilon-machine size, conditional information, conditional algorithmic information content, schema length, ideal complexity, hierarchical complexity, tree subgraph diversity, homogeneous complexity, grammatical complexity.

  - *Mutual information*

    Algorithmic mutual information, channel capacity, correlation, stored information, organization.

[8] Typically measured in bits
[9] Typically measured in time, energy, etc.

Subtle differences distinguish measures in the same group because they are closely related.

In the first group, entropy and algorithmic complexity are the most representative. Entropy is widely applicable for indicating randomness. It also measures uncertainty, ignorance, surprise, information, etc. Entropy is strongly related to statistical entropy (Boltzmann, Gibbs) and also to algorithmic complexity (also called algorithmic randomness and Kolmogorov-Chaitin complexity) which can be used to measure disorder or randomness without any recourse to probabilities [46, 96]. The algorithmic complexity of an object is defined as the length of the minimal universal Turing machine program needed to reproduce it. A basic theorem states that the entropy of a random variable X taking values in S is equal, except for an additive constant, to the expected value of algorithmic complexity of elements in S.

In the second group, computational complexity is the amount of computational resources (usually time or space) needed to solve a problem [42]. Logical depth, introduced by Bennett [7], measures the computational resources taken to calculate the results of a program of minimal length. And thermodynamic depth, introduced by Pagels and Lloyd [51], is a measure of how hard it is to put something together.

Finally, the third group expresses the concept of complexity that we have adopted in this thesis. A short review of this concept now follows.

2.3.3 Statistical Complexity

In the two last decades, diverse complexity measures, such as, for instance, excess entropy or effective measure complexity and mutual information, have been proposed from different fields to quantify the degree of structure or correlation of a system [36, 48, 29, 28]. To avoid confusion, Feldman and Crutchfield [29] proposed calling them "measures of *statistical complexity*".

James Crutchfield, David Feldman, and Murray Gell-Mann present a good introduction about their complexity concept in [29, 31] [10] . We summarize the most basic ideas below:

- Information is important in the study of complex structures.
- Randomness and unpredictability of a system (entropy) does not completely capture the correlational structure [11] in its behaviour.
- The larger and more intricate the correlations between the system's constituents, the more structured is the underlying distribution.
- Structure and correlation, however, are not completely independent of randomness.
- Both maximally random and perfectly ordered systems possess no structure.
- Statistical complexity measures provide a measure of the regularities present in an object above and beyond pure randomness.

[10] This point of view provided us with the basis of our concept of scene complexity.
[11] Structure is referred to as the relationship between the components of a system.

Diverse approaches to measuring statistical complexity have been taken. One line is based on information theory and the quantities used are related to various forms of mutual information. Another line appeals to computation theory's classification of devices that recognize different classes of formal languages. On the other hand, other researchers of the Santa Fe Institute define the statistical complexity of a system "so as to capture the degree to which the system consists of regularities versus randomness"[31].

Our complexity approach will be based on information theory and the complexity (correlation-structure-dependence) of a scene will be quantified by the mutual information.

2.4 Summary

In this chapter, we have reviewed the most basic concepts of the radiosity method, information theory, and complexity needed in this dissertation.

First, we focused attention on the radiosity equation, the properties of the form factors and their computation using local and global lines, random walks in a scene, and refinement criteria for hierarchical radiosity. Second, entropy and mutual information were reviewed for discrete and continuous channels, and the most basic inequalities were introduced. Finally, the notion of complexity and some possible ways of quantifying it were presented.

Chapter 3

Scene Visibility Entropy

In this chapter, a discretised scene is interpreted as a discrete information channel. This fact enables us to introduce the notions of entropy and mutual information applied to the visibility of a scene, where only the geometry is taken into account. We also discuss the concept of scene randomness, associated with entropy, and we study its relationship with Monte Carlo error in the form factor computation. Most of the contents of this chapter have been presented in [17, 23, 25].

3.1 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 2.1.8) 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 3.1).

- *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.

In the previous chapter (sections 2.1.8 and 2.2), 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}

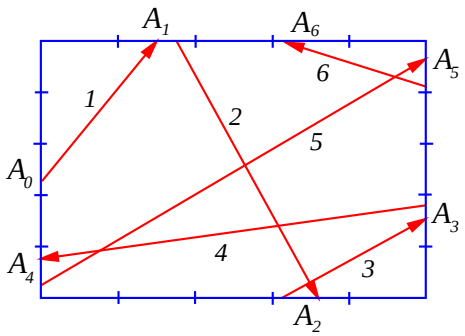


Figure 3.1: Discrete random walk in a scene.

  - 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}

3.1.1 Discrete Scene Visibility Entropy

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

$$H_S = H(Y|X) = -\sum_{i=1}^{n_p} a_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij} \tag{3.1}$$

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

$$H_i = H(Y|X = i) = -\sum_{j=1}^{n_p} F_{ij} \log F_{ij} \tag{3.2}$$

and thus (3.1) can be written as

$$H_S = \sum_{i=1}^{n_p} a_i H_i \tag{3.3}$$

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 (2.56) can be now expressed by the *reciprocity* property of the form factors (2.8)

$$p_{ij} = a_i F_{ij} = a_j F_{ji} \quad \forall i, j \tag{3.4}$$

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

$$H_P = H(X) = H(Y) = -\sum_{i=1}^{n_p} a_i \log a_i \tag{3.5}$$

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

$$H_J = H(X,Y) = -\sum_{i=1}^{n_p}\sum_{j=1}^{n_p} a_i F_{ij} \log a_i F_{ij} \tag{3.6}$$

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 $\{a_i F_{ij}\}$.

3.1.2 Discrete Scene Visibility Mutual Information

The *discrete scene visibility mutual information* is defined as

$$\begin{aligned} I_S &= I(X,Y) = H(Y) - H(Y|X) \\ &= -\sum_{i=1}^{n_p} a_i \log a_i + \sum_{i=1}^{n_p} a_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij} \end{aligned} \tag{3.7}$$

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 (section 2.2.2): $I(X,Y) = D_{KL}(\{p_{ij}\}\|\{p_i q_j\})$. 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

$$\begin{aligned} I_S &= \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} a_i F_{ij} \log \frac{a_i F_{ij}}{a_i a_j} \\ &= \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} a_i F_{ij} \log \frac{F_{ij}}{a_j} \end{aligned} \tag{3.8}$$

3.1.3 Properties

In a discretised scene the following properties are met:

- From the reciprocity property of the form factors (2.8), 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 3.2)

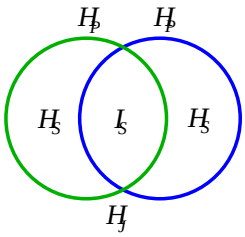


Figure 3.2: Venn diagram of a scene.

- 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

$$H_S = -\frac{1}{n_p} \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} F_{ij} \log F_{ij} = \frac{1}{n_p} \sum_{i=1}^{n_p} H_i \tag{3.9}$$

and

$$H_P = \log n_p \tag{3.10}$$

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: $H_S \approx 0$.

3.2 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?" [27]. 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 [47], "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".

3.2.1 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 [1].

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:

$$H_S = H_P = \log n_p \tag{3.11}$$

  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.

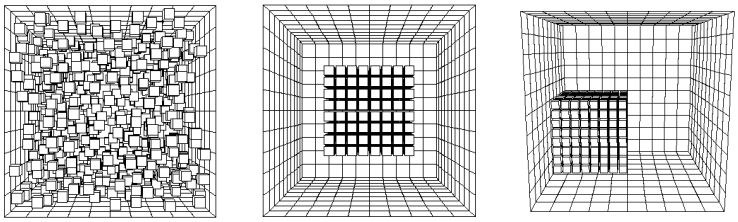
3.2.2 Empirical Results

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 2.1.6).

Different scenes but the same discretisation of the environment

In scenes with the same discretisation (as in Figure 3.3/Table 3.1, 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 3.3a), less randomness means more correlation (Figure 3.3c). The Venn diagrams in Figure 3.4 illustrate this behaviour.

[1] We have to be careful with the concept of maximum and minimum scene entropy, because this is strongly dependent on the scene discretisation, and particularly on the number of patches. For instance, it is very difficult to obtain a low value of the entropy if the number of patches is large.



(a) (b) (c)

Figure 3.3: A random configuration and two clustered configurations with 512 cubes.

Scene	H_S	I_S
Fig.3.3a	6.761	4.779
Fig.3.3b	5.271	6.270
Fig.3.3c	4.849	6.692

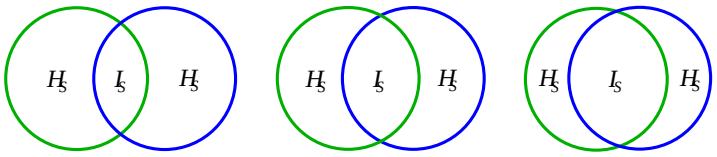
Table 3.1: Entropy and mutual information for Figures 3.3a, 3.3b and 3.3c. For each scene, $H_P = 11.541$ and 10^7 global lines have been used to compute the form factors.

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 4.4). 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 3.2, corresponding to Figure 3.5, 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 3.6 illustrate the behaviour of the entropy and mutual information of these scenes.

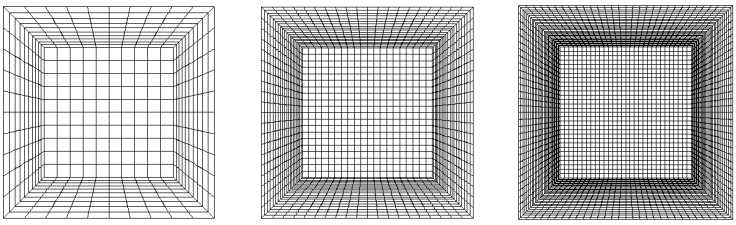
Normalized measures

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



(a) (b) (c)

Figure 3.4: 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 3.5: 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.

Scene	H_S	I_S	H_P
Fig.3.5a	7.838	1.391	9.229
Fig.3.5b	9.731	1.498	11.229
Fig.3.5c	10.852	1.547	12.399

Table 3.2: Results for a cubical enclosure with different discretisations of its surfaces (Figure 3.5). 10^9 global lines have been used to compute the form factors.

positional entropy H_P. So, they range from 0 to 1.

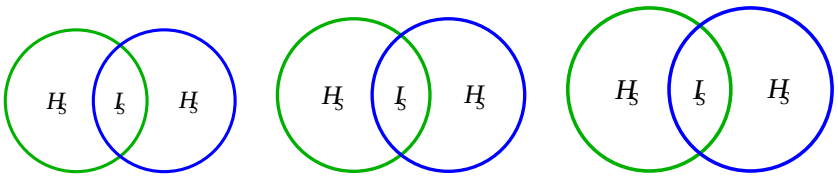
Normalized scene entropy can be defined as

$$\overline{H_S} = \frac{H_S}{H_P} \tag{3.12}$$

and *normalized scene mutual information* as

$$\overline{I_S} = \frac{I_S}{H_P} = 1 - \frac{H_S}{H_P} \tag{3.13}$$

In the literature, normalized mutual information is considered as a measure of *correlation* [15]. 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

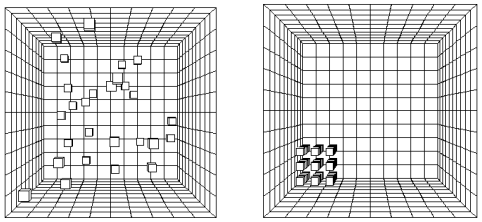


(a) (b) (c)

Figure 3.6: Venn diagrams corresponding to the scenes of Figure 3.5 where we have a cubical enclosure with successive refinements. The size of the circles (H_P) increases with the number of patches.

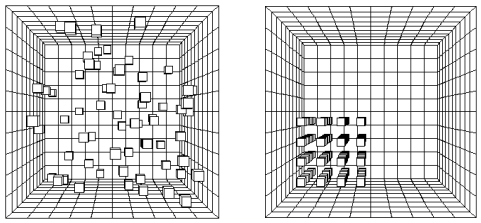
0 to 1. A similar approach has been used as a measure of 3D medical image alignment [85].

In the sequence of results of Table 3.3, where we start with an empty cubical enclosure (Figure 3.5a) and then we add small interior cubes (Figures 3.7, 3.8, 3.3a, and 3.3c), 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 3.7: (a) Random and (b) clustered configurations with 27 cubes.



(a) (b)

Figure 3.8: (a) Random and (b) clustered configurations with 64 cubes.

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.

3.3 Entropy and Mutual Information of a Scene in Flatland

In this section, discrete entropy and discrete mutual information are adapted to flatland by simply substituting the area of each patch with the length of each segment or "patch" [2].

Similarly to section 3.1, a random walk in a discretised 2D scene can be considered as a Markov chain where $P_{ij} = F_{ij}$, $n = n_p$, and $\{w_i\} = \{\frac{L_i}{L_T}\} = \{l_i\}$

[2]The exploration of flatland has facilitated our understanding of the information theory measures applied to a scene and has opened new lines of research. 2D studies have applications in areas such as robot motion and architectural design.

Scene	H_S	I_S	$\overline{H_S}$
Fig.3.5a	7.821	1.408	0.847
Fig.3.7a	7.780	1.669	0.823
Fig.3.7b	7.589	1.861	0.803
Fig.3.8a	7.709	2.009	0.793
Fig.3.8b	7.606	2.112	0.783
Fig.3.3a	6.761	4.779	0.586
Fig.3.3c	4.849	6.692	0.420

Table 3.3: Results for the empty scene of Figure 3.5a and the scenes with 27, 64 and 512 small cubes of Figures 3.5a, 3.7, 3.8, 3.3a, and 3.3c. For each scene, 10^7 global lines have been cast.

(l_i is the relative length of patch i). In a 2D scene, the Bayes theorem can be expressed by the following property of the form factors

$$p_{ij} = l_i F_{ij} = l_j F_{ji} \quad \forall i, j \tag{3.14}$$

3.3.1 Definitions

From the above assumptions, we obtain the following definitions:

- *Discrete scene visibility entropy rate* or simply *scene visibility entropy*

$$H_S = -\sum_{i=1}^{n_p} l_i \sum_{j=1}^{n_p} F_{ij} \log F_{ij} \tag{3.15}$$

- *Scene visibility positional entropy*

$$H_P = -\sum_{i=1}^{n_p} l_i \log l_i \tag{3.16}$$

- *Scene visibility joint entropy*

$$H_J = -\sum_{i=1}^{n_p} \sum_{j=1}^{n_p} l_i F_{ij} \log(l_i F_{ij}) \tag{3.17}$$

- *Discrete scene visibility mutual information*

$$I_S = H_P - H_S = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} l_i F_{ij} \log \frac{F_{ij}}{l_j} \tag{3.18}$$

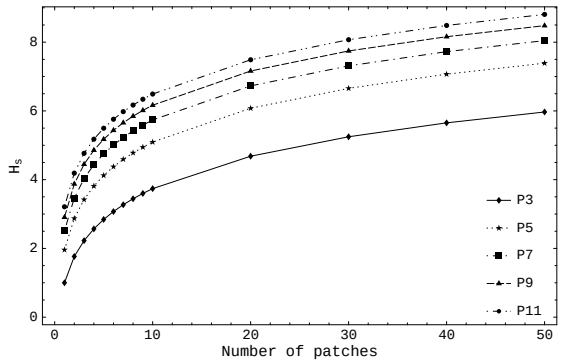


Figure 3.9: H_S values (in vertical axis) for regular polygons of 3, 5, 7, 9, and 11 sides with different regular discretisations for each side. The horizontal axis is labeled with the number of patches for each side.

3.3.2 Empirical Results

In order to illustrate the above definitions, we study two sets of scenes: the regular polygons and three scenes with 64 squares in each (Figure 3.12). In the first case, the form factors have been computed exactly by the string rule [57] and, in the second, 10^5 global lines have been cast to obtain a Monte Carlo solution for the form factors (see section 2.1.6).

For regular polygons, Figure 3.9 shows that the minimum scene visibility entropy corresponds to an equilateral triangle and the maximum to an 11-sided polygon. This fact can also be tested in Table 3.4. For instance, the entropy H_S of an equilateral triangle with 150 patches ($H_S \approx 5.969$) is less than the entropy H_S of a pentagon with the same number of patches ($H_S \approx 6.657$). In fact, continuing the sequence of regular polygons, maximum entropy should correspond to a circle.

number	H_S		
of sides	*10*	*30*	*50*
3	3.739	5.248	5.969
4	4.590	6.138	6.867
5	5.094	6.657	7.389
6	5.455	7.025	7.758
7	5.737	7.311	8.045
8	5.969	7.545	8.281
9	6.166	7.745	8.480
10	6.338	7.918	8.654
11	6.491	8.072	8.808
12	6.628	8.210	8.946

Table 3.4: H_S for regular polygons from 3 to 12 sides with different regular discretisations (10, 30, and 50 patches for each side).

In Figure 3.10/Table 3.5 we observe that maximum mutual information cor-

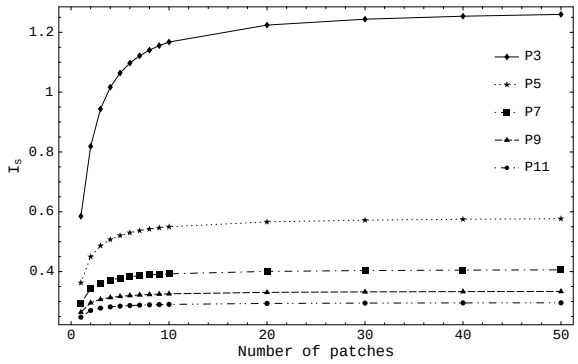


Figure 3.10: I_S values (in vertical axis) for regular polygons of 3, 5, 7, 9, and 11 sides with different regular discretisations for each side. The horizontal axis is labeled with the number of patches for each side.

responds to an equilateral triangle ($I_S \approx 1.260$) and minimum to a polygon of 11 sides ($I_S \approx 0.296$): there is a higher correlation between the sides of a triangle than between the sides of an 11-sided polygon. Thus, continuing this sequence of regular polygons, minimum mutual information has to correspond to a circle.

number of sides	I_S		
	10	*30*	*50*
3	1.168	1.244	1.260
4	0.732	0.769	0.777
5	0.550	0.572	0.577
6	0.452	0.467	0.470
7	0.393	0.404	0.406
8	0.353	0.362	0.363
9	0.326	0.332	0.333
10	0.306	0.311	0.312
11	0.290	0.295	0.296
12	0.279	0.282	0.283

Table 3.5: I_S for regular polygons from 3 to 12 sides with different regular discretisations (10, 30, and 50 patches for each side).

Figure 3.11 shows the behaviour of the entropy and mutual information of an equilateral triangle when the number of patches grows. In accordance with the theory, while entropy increases clearly with the number of patches, mutual information appears to converge to a determined value (see the next chapter, section 4.4).

For the scenes with 64 squares (Figure 3.12/Table 3.6) we observe that maximum mutual information, or correlation, is obtained in Figure 3.12a, and maximum entropy in Figure 3.12c. The mutual information of an empty square scene ($I_S \approx 0.760$ for 20 patches on each side) increases outstandingly when

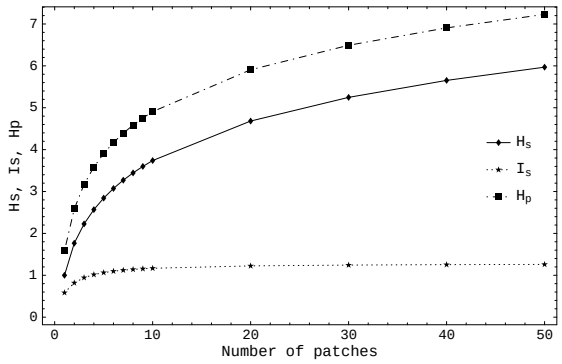
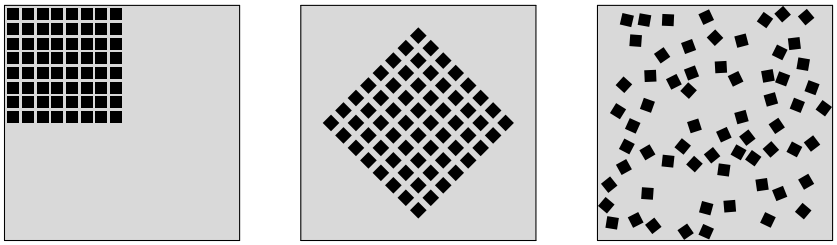


Figure 3.11: H_P, H_S and I_S values (in vertical axis) for an equilateral triangle with different regular discretisations for each side. The horizontal axis is labeled with the number of patches for each side.



(a) (b) (c)

Figure 3.12: Three scenes with 64 squares in each. The enclosure is regularly discretised into 80 patches.

we add 64 little squares in its interior (from $I_S \approx 4.990$ to $I_S \approx 6.148$). In Table 3.6, H_P should be equal in the three cases, but the difference is due to the computational error.

3.4 Scene Entropy and Monte Carlo Error in Form Factor Computation

We now explore the relationship between the scene entropy and the Monte Carlo error in form factor computation. We give some evidence that entropy is deeply related to the computational error.

3.4.1 Local and Global Lines from an Information-Theory Point of View

As we have seen in section 3.1, a scene can be modeled by a random walk or an information channel. A random walk in a scene can also be thought of as a sequence of local lines: starting in a given patch, a local line is sampled to

Scene	H_S	I_S	H_P
Fig.3.12a	2.244	6.148	8.392
Fig.3.12b	3.001	5.391	8.392
Fig.3.12c	3.443	4.990	8.433

Table 3.6: H_P, H_S, and I_S values for the scenes in Figure 3.12. 10^6 global lines have been cast for each scene.

obtain a destination patch which turns into the origin of a new local line, and so, recursively, we get a random walk.

Now, uniformly distributed local and global lines are interpreted from an information-theory point of view:

- The *entropy H_i of patch i* can be interpreted as the uncertainty about the destination patch of a local line cast from this patch.
- The *scene entropy rate H_S* represents the average uncertainty of a local line, cast from any patch, about the destination patch.
- The *joint entropy H_J* expresses the average uncertainty of a segment of a global line traversing a scene about the two patches connected.

Observe that the average uncertainty or unpredictability of a segment of a global line is higher than the uncertainty of a local line because $H_J = 2H_S + I_S$ (Figure 3.2). Thus, at least H_J is twice H_S. H_S ranges from 0 to H_P and H_J from H_P to $2H_P$. Let us also remember that, for a given discretisation, the higher the correlation, the smaller the uncertainty.

It is of interest to examine two extremal cases:

- A highly correlated scene (Figure 3.13a)

  In this case, the uncertainty of a local line can be very low and it is much smaller than the uncertainty of a global line segment. But in general, both uncertainties are relatively small compared with those of a spherical scene.

- A spherical scene (Figure 3.13b)

  The uncertainty of a segment of a global line is twice the uncertainty of a local line:

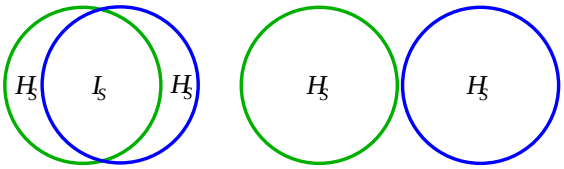
$$H_J = 2H_S = 2\log n_p \tag{3.19}$$

3.4.2 Scene Entropy and Variance of the Form Factor Estimators

In this section, we study the relationship between the entropy and the mean square error of all form factors.

Using local or global lines, the variance for the form factor estimator (section 2.1.6) is given by (2.38)

$$V[F_{ij}] = \frac{F_{ij}(1 - F_{ij})}{N_i}$$



(a) (b)

Figure 3.13: (a) A highly correlated scene. (b) A spherical scene.

where N_i is the number of local lines cast from patch i or the number of global lines which cross i.

Let us consider now an importance vector α_i, which gives us the importance we assign to each patch. For each patch i, we cast a number of lines, N_i, proportional to this importance, that is $N_i = \alpha_i N$, where N is the total number of lines cast. For instance, α_i could be $a_i = \frac{A_i}{A_T}$.

Now we consider the expected value of the mean square error for all patches (weighted with the importance α_i). Since for each patch the expected value of the square error is equal to the variance, we obtain:

$$\begin{aligned} E(MSE) &= \sum_i \alpha_i \sum_j V[F_{ij}] \\ &= \sum_i \alpha_i \sum_j \frac{F_{ij}(1-F_{ij})}{N_i} \\ &= \sum_i \alpha_i \sum_j \frac{F_{ij}(1-F_{ij})}{\alpha_i N} \\ &= \frac{1}{N}(\sum_i \sum_j F_{ij} - \sum_i \sum_j F_{ij}^2) \\ &= \frac{1}{N}(n_p - \sum_i \sum_j F_{ij}^2) \end{aligned} \tag{3.20}$$

This expression has a minimum value when for all i in the sum there exists a j such that $F_{ij} \approx 1$ and ≈ 0 otherwise. This ideal situation gives us a null error, and obviously corresponds to the minimum scene entropy and maximum mutual information. On the other hand, the maximum is obtained when all the form factors are equal (with value $\frac{1}{n_p}$), which corresponds to the maximum scene entropy.

The sets of scenes analyzed in this chapter can be used to study the relationship between scene entropy and the expected value of the mean square error for all form factors.

Tables 3.7, 3.8 and 3.9 show that the greater the entropy the greater the error in form factor computation. This means that, *for a given error, we need to cast more lines for a scene with more entropy.* Also, by increasing the number of lines, we can see how the error in entropy estimation increases with the scene entropy. Tables 3.7 and 3.8 also show how the increase in the number of lines increases the estimated value of the entropy and, logically, decreases the estimated value of the mutual information. The reason for this entropy increase is that, in the algorithm used, the form factors have an initial value of zero

and thus, with few lines, many form factors retain this value. This leads to an apparently larger correlation in a scene, i.e., a lower value in entropy estimation and a higher value in mutual information estimation.

Scene	*Lines* (10^6)	H_S	I_S	$E(MSE)$
Fig.3.3a	1	6.370	5.171	1284.6
Fig.3.3a	10	6.761	4.779	1286.7
Fig.3.3b	1	5.072	6.469	949.3
Fig.3.3b	10	5.271	6.270	950.1
Fig.3.3c	1	4.674	6.867	898.8
Fig.3.3c	10	4.849	6.692	900.9

Table 3.7: Results for three cubical enclosures with 512 small cubes (Figures 3.3a, 3.3b and 3.3c). The $E(MSE)$ is normalized to a single line. For each scene, $H_P = 11.541$ and 10^6 and 10^7 lines have been cast.

By increasing the number of patches, how does the error behave? Table 3.8 shows how the error increases with the number of patches, in the same way as scene entropy increases.

Scene	*Lines* (10^6)	H_S	I_S	$E(MSE)$
Fig.3.5a	10	7.821	1.408	497.7
Fig.3.5a	100	7.837	1.392	497.7
Fig.3.5a	1000	7.838	1.391	497.7
Fig.3.5b	10	9.420	1.809	2012.7
Fig.3.5b	100	9.705	1.524	2013.0
Fig.3.5b	1000	9.731	1.498	2013.0
Fig.3.5c	10	9.684	2.715	4543.9
Fig.3.5c	100	10.708	1.691	4545.8
Fig.3.5c	1000	10.852	1.547	4546.1

Table 3.8: Results for the cubical enclosure with different discretisations of its surfaces (Figure 3.5a: 600 patches, Figure 3.5b: 2400 patches and Figure 3.5c: 5400 patches). The $E(MSE)$ is normalized to a single line. For each scene, 10^7, 10^8 and 10^9 lines have been cast.

The relationship between scene entropy and computational error is reinforced by the following observations [3] :

- H_S expresses uncertainty or *equivocation* [78].

[3] A theoretical relationship between scene entropy and the variance of the form factor estimates remains to be investigated.

Scene	H_S	I_S	H_P	$E(MSE)$
Fig.3.12a	2.244	6.148	8.392	23.315
Fig.3.12b	3.001	5.391	8.392	31.274
Fig.3.12c	3.443	4.990	8.433	37.121

Table 3.9: H_P, H_S, I_S and $E(MSE)$ values for the scenes of Figure 3.12. $E(MSE)$ is normalized to a single line segment. 10^6 global lines have been cast for each scene.

- H_S, like variance, increases with refinement and is zero when there is no choice.
- H_S represents the average difficulty in hitting a patch.
- "Although both entropy and variance are measures of dispersion and uncertainty, the lack of a simple relationship between orderings of a distribution by the two measures emanates from quite substancial and subtle differences. Both measures reflect *concentration* but their respective metrics for concentration are different. Unlike variance which measures concentration only around the mean, entropy measures diffuseness of the density irrespective of the location(s) of concentration" [18].
- Suppose we wish to measure the probability of error P_e in guessing the origin patch (X) of a global line segment from the knowledge of the patch hit (Y). From Fano's inequality (2.71) we obtain a tight lower bound on this error probability in terms of the scene entropy H_S:

$$P_e \geq \frac{H_S - 1}{\log n_p} \tag{3.21}$$

  As H_S 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 $H_S = 0$. Intuitively, we expect to be able to estimate X with a low probability of error if and only if H_S is small.

3.5 Summary

In this chapter, we have presented a scene as a discrete information channel where the input and output variables take values over a set of patches and the channel transition matrix as the form factor matrix. This gives us a new way of looking at the visibility of a scene.

We applied the most basic information-theory measures (Shannon entropy, conditional entropy, joint entropy, and mutual information) to scene visibility and we interpreted entropy and mutual information of a scene, respectively, in the following way:

- Scene visibility entropy provides the average uncertainty that remains about the source patch when the destination patch is known, and vice versa. It also expresses the randomness or unpredictability in a scene.

- Discrete scene visibility mutual information expresses the average information transfer in a scene. It can also be interpreted as the correlation or dependence in a scene.

We described different experiments with 3D and 2D scenes to show the complementarity of these measures.

We developed a new interpretation of the entropy from local and global lines. We have shown that the uncertainty of a global line segment is at least twice the uncertainty of a local line.

Finally, we demonstrated that there is a close relationship between the scene entropy and the Monte Carlo error in form factor computation. We have seen that the bigger the entropy, the bigger the error in form factor computation. This means that, for a given error, we need to cast more lines for a scene with more entropy.

Chapter 4

Scene Visibility Complexity

This chapter is dedicated to scene visibility complexity, which will be interpreted as the difficulty in achieving a precise discretisation. We begin by proposing the continuous scene mutual information, independent of any discretisation, as a measure for the scene complexity. Continuous scene visibility mutual information is computed using both global or local lines. The relationship between the discrete and continuous mutual information is analysed and the difference between them is defined as the discretisation error. Finally, we describe a possible scene classification in flatland. Most of the content of this chapter has been discussed in [21, 25].

4.1 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. As we have mentioned in section 1.1, 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 thesis, the word *complexity* will be reserved for a measure of *statistical complexity* and *entropy* will be 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 becomes the mutual information.*

Proof. From (2.60–2.63) and the definitions in section 3.1, we have

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

Thus, from (2.64), $E=\lim_{n\to\infty}(H_P+(n-1)H_S-nH_S)=H_P-H_S=I_S$. □

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." [28]
- Entropy and mutual information are orthogonal or complementary: "Complexity and randomness each capture a useful property to describe how a process manipulates information" [77].

In this thesis, 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 (2.57) relates entropy to mutual information:

$$H(X)=H(X|Y)+I(X,Y) \tag{4.1}$$

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.

4.2 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 4.1) or by a continuous information channel. As we have seen in section 2.2.6, 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.

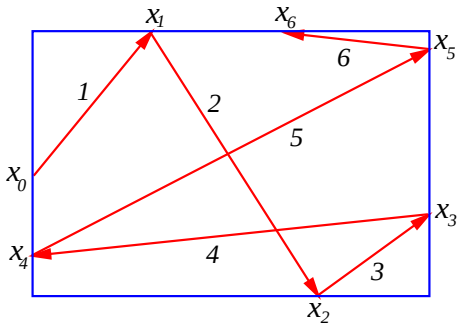


Figure 4.1: Continuous random walk in a scene.

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:

$$I_S^c = \lim_{n_p \to \infty} I_S \tag{4.2}$$

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:

$$\lim_{n_p \to \infty} H_S = \infty \tag{4.3}$$

As we have seen (section 2.1.8), 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 2.1.3).

In the same way, the continuous formulae for a scene can also be obtained from the continuous formulae of the entropy and mutual information (section 2.2.6) 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*

$$H_P^c = -\int_S \frac{1}{A_T}\log\frac{1}{A_T}dA_x = \log A_T \tag{4.4}$$

- *Continuous scene visibility entropy*

$$H_S^c = -\int_S\int_S \frac{1}{A_T}F(x,y)\log F(x,y)dA_x dA_y \tag{4.5}$$

- *Continuous scene visibility mutual information*

$$\begin{aligned} I_S^c &= \log A_T + \int_S\int_S \frac{1}{A_T}F(x,y)\log F(x,y)dA_x dA_y \\ &= \int_S\int_S \frac{1}{A_T}F(x,y)\log(A_T F(x,y))dA_x dA_y \end{aligned} \tag{4.6}$$

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.

4.3 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.

4.3.1 Monte Carlo Integration

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

$$\begin{aligned} I_S^c &= \int_S\int_S \frac{1}{A_T}F(x,y)\log(A_T F(x,y))dA_x dA_y \\ &= \int_S\int_{\Omega_x} \frac{1}{A_T}\frac{\cos\theta_x}{\pi}\log(A_T F(x,y(x,\omega_x)))dA_x d\omega_x \end{aligned} \tag{4.7}$$

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

$$\begin{aligned} I_S^c &\approx \frac{1}{N}\sum_{k=1}^{N} \log(A_T F(x_k, y_k(x_k, \omega_{x_k})) \\ &= \frac{1}{N}\sum_{k=1}^{N} \log\Big(\frac{A_T \cos\theta_{x_k} \cos\theta_{y_k}}{\pi r_{x_k y_k}^2}\Big) \end{aligned} \tag{4.8}$$

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 2.8). 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.

4.3.2 Empirical Results

We begin by computing the complexity of platonic solids and the Cornell box (Figure 2.1a) [1]. In Table 4.1, 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.

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

In addition, in Table 4.2, we show the complexity for the scenes of Figure 4.2. In Figure 4.2a, an object, made up of a table and four chairs, is situated in

[1] In our version of the Cornell box scene, prism and cube are slightly separated from the floor. This fact increases the scene visibility complexity, as in the narrow spaces the correlation is very high.

the middle of a room. In Figures 4.2b and 4.2c, arrays of 4 or 16 objects have been situated in the middle of the same room. In Figures 4.2d, 4.2e and 4.2f, 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 (4.2c, 4.2d, 4.2e and 4.2f) 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.

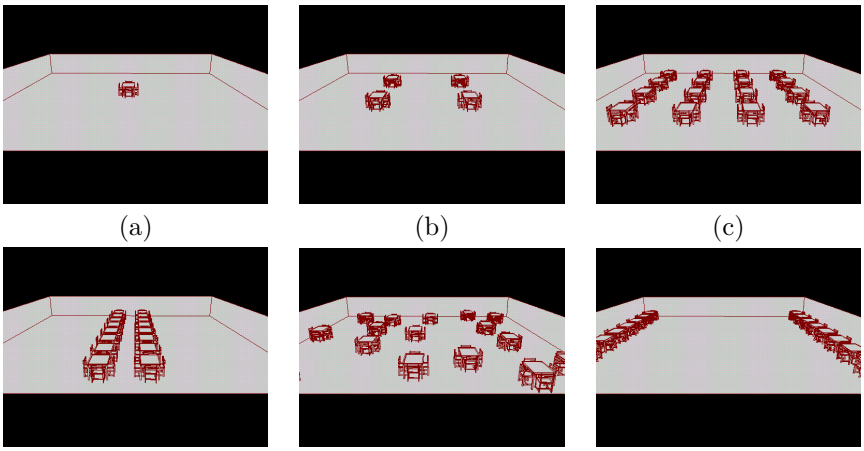


Figure 4.2: (a) An object, composed of a table and four chairs, and (b) an array of 4 objects with the same composition, have been situated in the middle of a room. (c, d, e, f) The same 16 objects have been distributed in four different ways.

Scenes	Fig.4.2a	Fig.4.2b	Fig.4.2c	Fig.4.2d	Fig.4.2e	Fig.4.2f
I_S^c	3.837	4.102	5.023	5.043	5.044	5.089

Table 4.2: Complexity of the scenes of Figure 4.2. For each scene, 10^6 global lines have been cast.

4.4 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" [3]. 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" [48].

4.4.1 Continuous versus Discrete Mutual Information

From section 2.2.6, 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:

- 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:

$$I_S^c - I_S \geq 0 \tag{4.9}$$

Initially, these results were proved in [26, 21]. In the next chapter, we give a general proposition proving these properties for visibility, radiosity and importance. In chapter 6, we will study in more detail the relationship between I_S^c and I_S.

Scene	*Lines* (10^6)	H_S	I_S	I_S^c
Fig.3.5a	0.1	6.482	2.747	1.610
Fig.3.5a	10	7.821	1.408	1.612
Fig.3.5a	1000	7.838	1.391	1.610
Fig.3.5b	0.1	5.342	5.887	1.608
Fig.3.5b	10	9.420	1.809	1.611
Fig.3.5b	1000	9.731	1.498	1.610
Fig.3.5c	0.1	4.313	8.086	1.610
Fig.3.5c	10	9.684	2.715	1.611
Fig.3.5c	1000	10.852	1.547	1.610

Table 4.3: Results for the cubical enclosure of Figure 3.5 with different discretisations of its surfaces. For each scene, 10^5, 10^7, and 10^9 global lines have been cast.

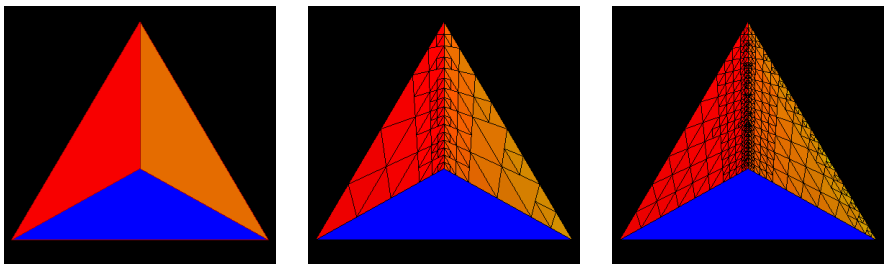
As we can see in Tables 4.3 and 4.4, corresponding to Figures 3.5 and 3.3(a), respectively, the computational cost of I_S^c is much lower than the cost of I_S: with few lines I_S^c can be computed with enough precision, unlike I_S which needs a lot of lines to get a precise measurement. Observe that I_S increases with the

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 4.4: Discrete and continuous mutual information for a scene with 512 cubes (Figure 3.3a).

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 4.5 and 4.6, corresponding to Figures 4.3 and 4.4, we also show how discrete mutual information I_S increases with the mesh refinement.



(a) (b) (c)

Figure 4.3: Three different discretisations for a tetrahedron. The total number of patches is, respectively, 4, 151, and 793.

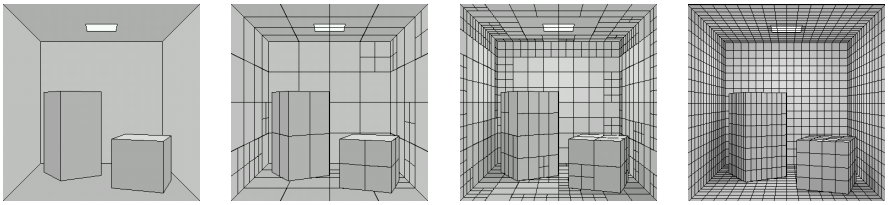
Scene	*Patches*	I_S	I_S^c
Fig.4.3a	4	0.415	2.626
Fig.4.3b	151	1.217	2.626
Fig.4.3c	793	1.445	2.626

Table 4.5: I_S^c and I_S for the scenes in Figure 4.3. For each scene, 10^7 local lines have been cast.

4.4.2 Discretisation Accuracy

We know that the difference $I_S^c - I_S$ (4.9), always positive, expresses the *loss of information transfer* due to the discretisation [2]. From this assumption, we can now take a leap forward with two fundamental proposals. The first appears

[2]This idea will be discussed further in chapter 6, where we will propose a refinement criterion based on this difference.



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

Figure 4.4: Four different discretisations for the Cornell box. The total number of patches is, respectively, 19, 196, 826, and 1924.

Scene	*Patches*	*Lines* (10^6)	I_S	I_S^c
Fig.4.4a	19	10	0.690	3.273
Fig.4.4b	196	10	2.199	3.273
Fig.4.4c	826	10	2.558	3.273
Fig.4.4d	1924	100	2.752	3.273

Table 4.6: I_S^c and I_S for the scenes in Figure 4.4.

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

$$\delta^v = I_S^c - I_S \tag{4.10}$$

and the *relative discretisation error* as the quotient

$$\overline{\delta}^v = \frac{I_S^c - I_S}{I_S^c} \tag{4.11}$$

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.

In the following experiments we find that, for a regular discretisation, the relative discretisation error is greater in the most complex scenes, i.e., a regular discretisation obtains better results in less complex scenes.

These proposals can be analysed from the results shown in Tables 4.7 and 4.8 and on the graphics of Figures 4.7 and 4.8, which have been obtained from Figures 4.5 and 4.6. Initially, 64 cubes are grouped very closely together in the center of the cubical enclosure. Little by little they are separated and moved outwards until they almost touch the walls. Complexity has been calculated for this sequence of scenes. Figure 4.7, showing continuous versus discrete mutual information, indicates that, according to our definitions, the more complex scenes are those that have surfaces closer to each another (Figures a,f and b).

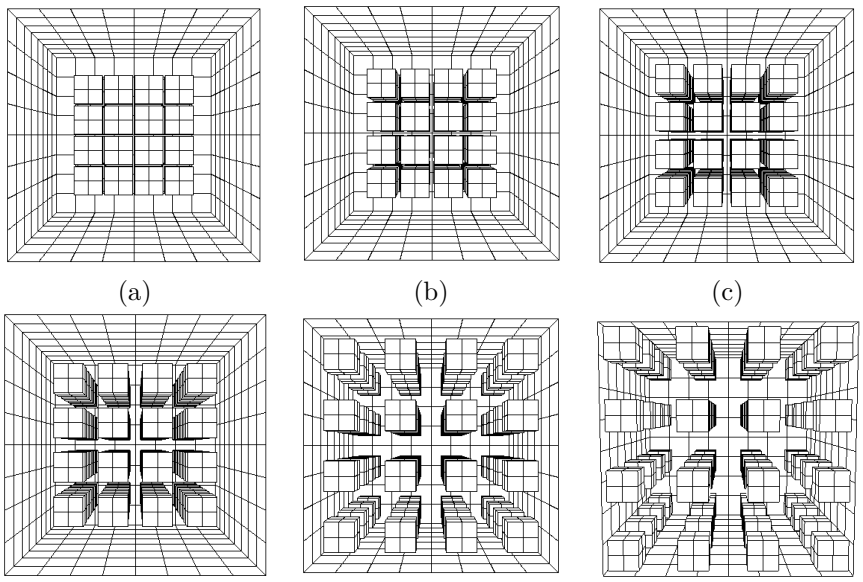


Figure 4.5: 64 cubes are grouped very closely together in the center of the cubical enclosure and then are separated and moved outwards until they almost touch the walls. The discretisation of the cubes (1536 patches) is finer than the discretisation of the walls (384 patches).

Scene	Fig.4.5a	Fig.4.5b	Fig.4.5c	Fig.4.5d	Fig.4.5e	Fig.4.5f
I_S	5.492	5.054	4.672	4.395	4.356	4.775
I_S^c	6.430	5.678	5.177	4.867	5.015	6.055
$\frac{I_S}{I_S^c}$	0.854	0.890	0.902	0.903	0.869	0.789

Table 4.7: Results for the scenes of Figure 4.5. For each scene, 10^8 global lines have been cast.

This sequence of scenes has been discretised in two different ways with the same number of patches. In the first sequence (Figure 4.5), the discretisation of the cubes is finer whereas in the second (Figure 4.6) the discretisation of the walls is finer. The accuracy of the discretisation appears to be higher in the "middle" scenes (b, c, d and e) and lower in the "extremal" scenes (a and f).

Scenes a and f are the most complex scenes. Consequently, these scenes should have a finer discretisation in order to obtain greater accuracy. In other words, these scenes are the ones most difficult to discretise.

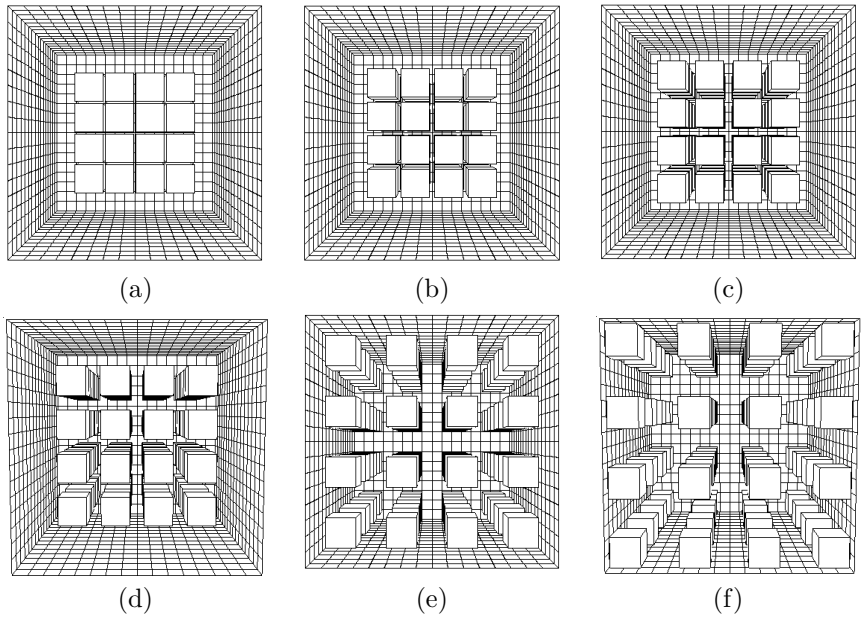


Figure 4.6: 64 cubes are grouped very closely together in the center of the cubical enclosure and then are separated and moved outwards until they almost touch the walls. The discretisation of the walls (1536 patches) is finer than the discretisation of the cubes (384 patches).

Scene	Fig.4.6a	Fig.4.6b	Fig.4.6c	Fig.4.6d	Fig.4.6e	Fig.4.6f
I_S	5.110	4.809	4.543	4.348	4.483	4.932
I_S^c	6.430	5.678	5.177	4.867	5.015	6.055
$\frac{I_S}{I_S^c}$	0.795	0.847	0.878	0.893	0.894	0.814

Table 4.8: Results for the scenes of Figure 4.6. For each scene, 10^8 global lines have been cast.

In Figures 4.7 and 4.8 we compare alternative discretisations. For instance, from Figures 4.5a and 4.6a we can see that the best discretisation appears to correspond to Figure 4.5a because the discretisation is finer in the narrow spaces between the cubes. In contrast, when the cubes are near the walls, greater precision is obtained when the discretisation of the walls is finer (Figure 4.6f).

In Table 4.9, corresponding to Figure 4.9, we also observe that the relative discretisation error is higher for more complex scenes.

These experiments suggest that discretisation error may be used to choose a better discretisation from several alternatives and, while computational error is deeply related to entropy, discretisation error is related to mutual information.

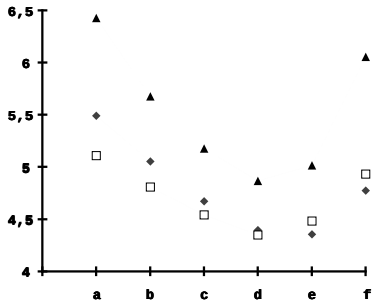


Figure 4.7: The vertical axis shows the continuous mutual information (triangles) and the discrete mutual information for the scenes a-to-f in Figures 4.5 (diamonds) and 4.6 (squares).

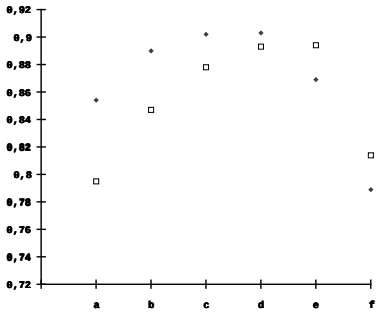
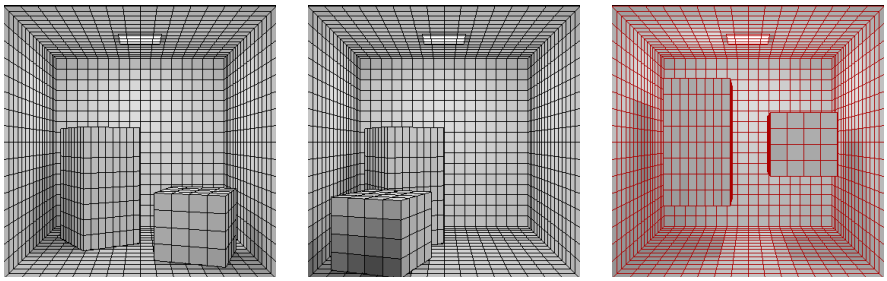


Figure 4.8: Discretisation accuracy in vertical axis for the scenes a-to-f in Figures 4.5 (diamonds) and 4.6 (squares).



(a) (b) (c)

Figure 4.9: Three different scenes with a regular discretisation of their surfaces and the same number of patches (1924).

Scene	Fig.4.9a	Fig.4.9b	Fig.4.9c
I_S	2.752	2.823	2.459
I_S^c	3.274	3.375	2.613
$\overline{\delta}^v$	0.159	0.164	0.059

Table 4.9: I_S, I_S^c, and $\overline{\delta}^v$ for the scenes in Figure 4.9. For each scene, 10^8 local lines have been cast.

4.5 Scene Visibility Complexity in Flatland

In this section, continuous scene visibility mutual information is defined and computed in flatland. An extensive analysis is undertaken to study the reasons for the growth in complexity.

4.5.1 Continuous Scene Visibility Mutual Information

Similarly to section 4.2, the continuous mutual information in flatland can be obtained from the discrete mutual information (3.18) by substituting

- Each state by an infinitesimal length and each summatory by an integral.
- $l_i \Longrightarrow \frac{1}{L_T}$.
- $F_{ij} \Longrightarrow F(x,y)$. Remember that the value of $F(x,y)$ is $\frac{\cos\theta_x \cos\theta_y}{2r_{xy}}$ for mutually visible points and zero if not (see section 2.1.3).

Thus, continuous mutual information is defined by

$$\begin{aligned} I_S^c &= \log L_T + \int_L \int_L \frac{1}{L_T} F(x,y) \log F(x,y) dL_x dL_y \\ &= \int_L \int_L \frac{1}{L_T} F(x,y) \log(L_T F(x,y)) dL_x dL_y \end{aligned} \tag{4.12}$$

As in 4.3, this integral can be solved by Monte Carlo integration and the computation can be done efficiently by casting uniformly distributed global lines upon segments [10] (see Figure 2.8). Hence, continuous mutual information can be approximated by

$$\begin{aligned} I_S^c &\approx \frac{1}{N} \sum_{k=1}^{N} \log(L_T F(x_k, y_k)) \\ &= \frac{1}{N} \sum_{k=1}^{N} \log(\frac{L_T cos\theta_{x_k} cos\theta_{y_k}}{2r_{x_k y_k}}) \end{aligned} \tag{4.13}$$

where N is the total number of pairs of points considered, which is the total number of intersections divided by two.

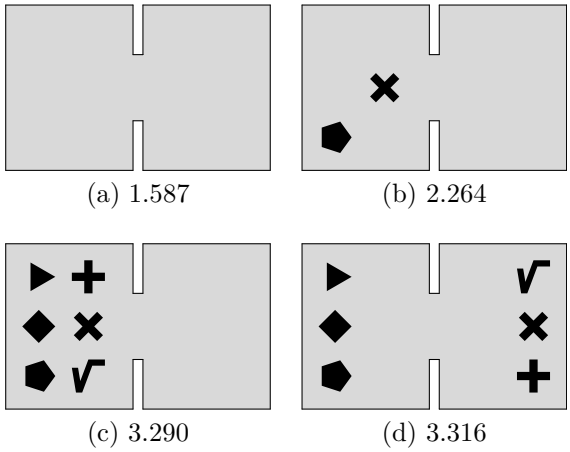
Results and discussion

As we have seen in section 4.4, I_S^c is the least upper bound to I_S. This fact, which is essential in the study of discretisation, is illustrated by the results obtained in this section. Continuous mutual information is computed for two sets of scenes: regular polygons and a room with diverse objects. In all cases, we have solved the Monte Carlo integral by casting 10^5 global lines, but in the first case we have also calculated the closed form of I_S^c for a circle, a hexagon, a square, and an equilateral triangle.

For regular polygons, Table 4.10 groups the discrete and continuous results. We can see that I_S^c is higher than I_S and that the complexity of regular polygons is very low. In the other scenes (Figure 4.10), we can see once again how complexity grows with the introduction of objects in the scene.

number of sides	I_S			I_S^c
	10	*30*	*50*	
3	1.168	1.244	1.260	1.284
4	0.732	0.769	0.777	0.788
5	0.550	0.572	0.577	0.583
6	0.452	0.467	0.470	0.475
7	0.393	0.404	0.406	0.408
8	0.353	0.362	0.363	0.366
9	0.326	0.332	0.333	0.336
10	0.306	0.311	0.312	0.314
11	0.290	0.295	0.296	0.297
12	0.279	0.282	0.283	0.284

Table 4.10: I_S and I_S^c values for regular polygons, from 3 to 12 sides, with different regular discretisations (10, 30, and 50 patches for each side).



(a) 1.587 (b) 2.264

(c) 3.290 (d) 3.316

Figure 4.10: I_S^c value for a room with several objects.

Scene	*exact value*			*MC*
circle	$\log \frac{\pi}{e}$	$\simeq$	0.209	0.209
hexagon	$\log \frac{e^{\sqrt{3}-4} 324(7+4\sqrt{3})}{168+97\sqrt{3}}$	$\simeq$	0.475	0.475
square	$\log \frac{8(1+\sqrt{2})}{e^{1+\sqrt{2}}}$	$\simeq$	0.789	0.788
equilateral triangle	$\log \frac{18}{e^2}$	$\simeq$	1.285	1.284

Table 4.11: Exact I_S^c values for a circle and three regular polygons compared with results obtained by Monte Carlo simulation (MC) with 10^5 global lines.

The closed-form solution of the continuous mutual information integral for some regular polygons is shown in Table 4.11.

The circle requires special attention. Observe that its complexity is different from zero: $I_S^c = \log \frac{\pi}{e}$ [3]. Since a sphere has zero complexity ($I_S^c = 0$), we could expect the same for a circle. But null complexity for the sphere is due to the fact that, for two spherical patches i and j, form factor F_{ij} is equal to $\frac{A_j}{A_T}$ (A_j is the area of patch j and A_T is the total area of the sphere). For the same reason, a uniform line can be generated by selecting two random points on its surface [74]. However, in the case of a circle, in general $F_{ij} \neq \frac{L_j}{L_T}$ (L_j is the length of patch j, L_T is the total length of the circle) and selecting pairs of random points on its perimeter will not yield a uniform density [10]. In flatland, we can not imagine a scene with less complexity than a circle. In this sense, there is a significant difference between the 3D and 2D worlds.

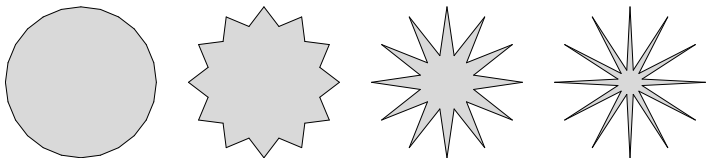
4.5.2 Scene Classification in Flatland

In this section we want to present a simple scene classification based on complexity and study the main reasons for the growth in complexity. First we compute the complexity of some empty scenes and, after that, we analyse other scenes with objects placed inside.

Some study cases

First, we compute the complexity of four sequences of scenes: the formation of a 12-pointed star, the Von Koch fractal, some different triangles and an L-shaped room. After that, we study two particular sequences of scenes: a scene with three objects which increase in size and an expansion of 16 squares from the center of a scene to its walls.

If we start with a polygon of 24 edges, with a complexity very similar to the one of a circle, and we continue closing the edges as shown in Figure 4.11, the complexity increases noticeably, due to the growth of the interaction between the edges. In the Von Koch fractal (Figure 4.12), a similar thing happens:

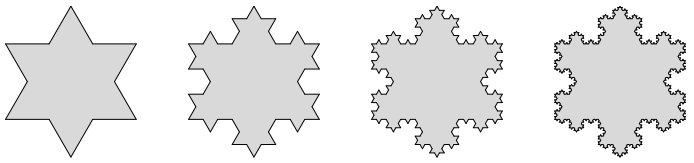


(a) 0.232 (b) 1.966 (c) 5.236 (d) 6.915

Figure 4.11: I_S^c value for a 24-sided regular polygon and three 12-pointed stars.

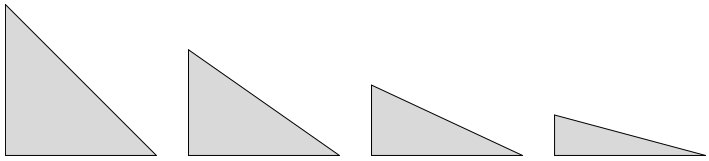
by increasing the number of corners, the correlation increases. In the case of Figures 4.13 and 4.14, we show how complexity increases when the scene becomes less regular. Going from an equilateral triangle (Table 4.10) to the triangle in Figure 4.13(d), complexity increases from 1.284 to 2.311. The same thing happens when we convert a square ($I_S^c \approx 0.788$) to the L-shaped figure in Figure 4.14(d) ($I_S^c \approx 1.319$). The importance of the size of the objects is obvious in Figure 4.15. We have seen that the introduction of objects increases

[3] N.B. the *beauty* of this result.



(a) 1.910 (b) 2.950 (c) 4.258 (d) 5.726

Figure 4.12: I_S^c value for Von Koch fractals.



(a) 1.493 (b) 1.530 (c) 1.787 (d) 2.311

Figure 4.13: I_S^c value for different triangles.

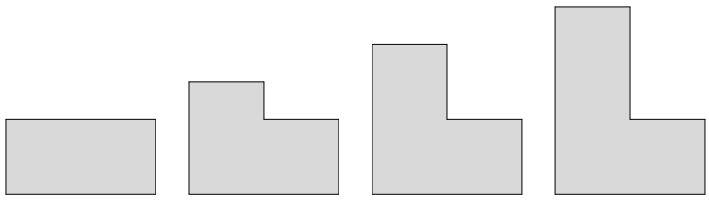
complexity, but the degree of interaction between the objects themselves and the enclosure depends clearly on their relative size: in general, the bigger the objects, the greater the complexity. In Figure 4.16, the scene begins with 16 squares almost touching each other and ends with the 16 squares almost touching the edges of the enclosure. We can see (Figure 4.17) that complexity is maximal in the first and last scenes and minimal in the middle scene. After these sequences of scenes, we can confirm that the increase in the number of corners or the creation of closer corners or spaces produces an increase in complexity.

Increase in complexity near singularities

In this section, we study the evolution of three groups of scenes. In the first case, a pentagonal star grows until it almost touches the vertexes of a pentagonal enclosure. In the second case, an interior square rotates in a square enclosure from a position with parallel sides to a position where the vertexes of the interior square almost touch the enclosure. In the third case, an interior square advances until almost touching the walls of a corner in a square enclosure.

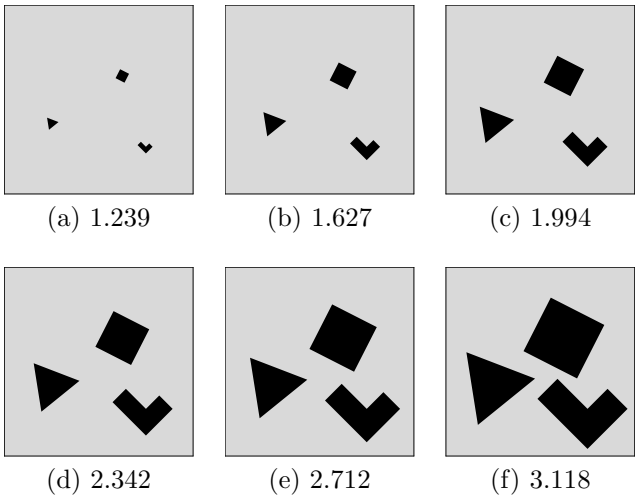
In Figure 4.18, a singularity occurs when the points of the star touch the vertexes of the pentagon. We obtain, in this particular case, 5 independent scenes. A similar thing happens with a rotating square (Figure 4.19).

When the vertexes of the pentagonal star (Figure 4.18) or the square (Fig-



(a) 0.945 (b) 1.106 (c) 1.218 (d) 1.319

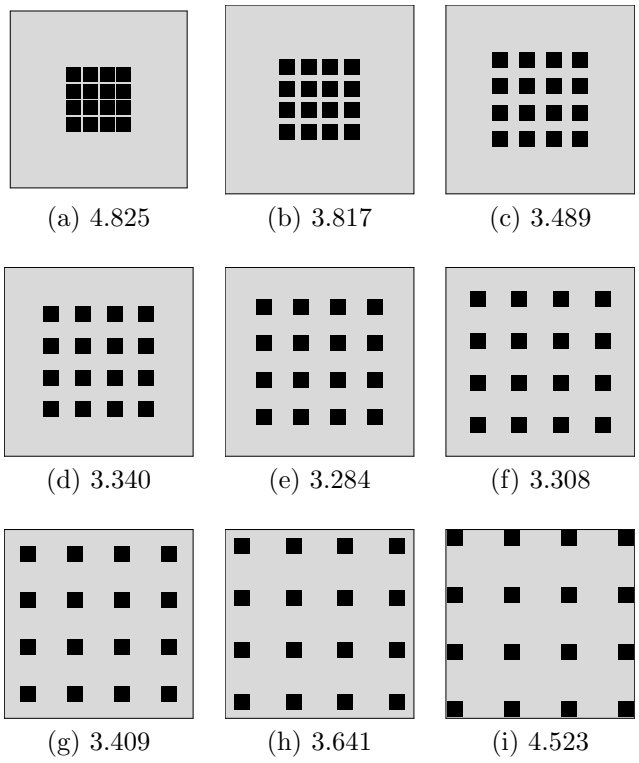
Figure 4.14: I_S^c value for a rectangle and three L-shaped rooms.



(a) 1.239 (b) 1.627 (c) 1.994

(d) 2.342 (e) 2.712 (f) 3.118

Figure 4.15: I_S^c value for six scenes with three objects increasing in size.



(a) 4.825 (b) 3.817 (c) 3.489

(d) 3.340 (e) 3.284 (f) 3.308

(g) 3.409 (h) 3.641 (i) 4.523

Figure 4.16: I_S^c value for nine scenes which represent an expansion of 16 squares from the center of a square to its walls.

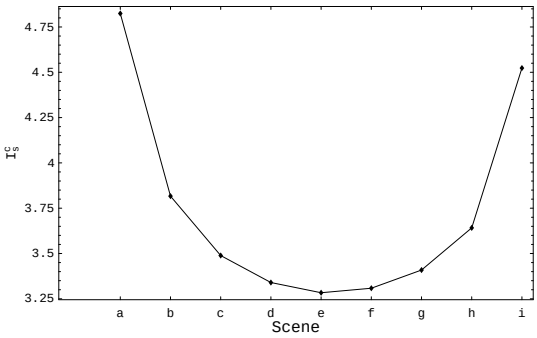
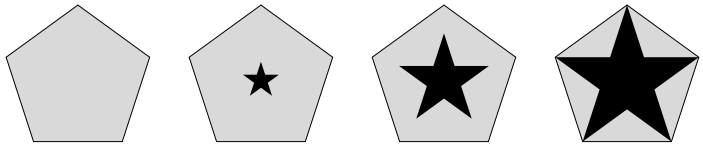
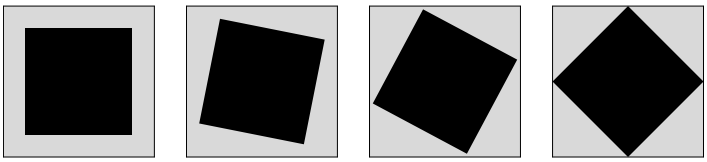


Figure 4.17: Representation of I_S^c value corresponding to the expansion in Figure 4.16.



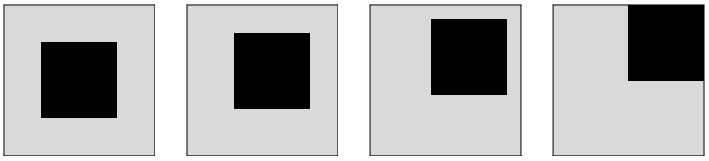
(a) 0.583 (c) 1.450 (f) 2.481 (i) 4.067

Figure 4.18: I_S^c value for a pentagon and three scenes with a pentagonal star which grows until it almost touches the vertexes of the enclosure.



(a) 3.216 (c) 3.235 (f) 3.340 (i) 3.477

Figure 4.19: I_S^c value for a scene with an interior square which rotates until its vertexes almost touch the enclosure.



(a) 2.272 (c) 2.308 (f) 2.520 (i) 3.304

Figure 4.20: I_S^c value for a scene with an interior square which advances until almost touching the walls of a corner.

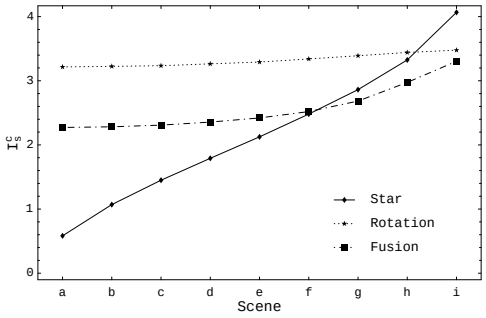


Figure 4.21: Representation of I_S^c corresponding to the scenes in Figures 4.18 (star), 4.19 (rotation) and 4.20 (fusion).

ure 4.19) *almost* touch their respective enclosures, the following proposition is fulfilled:

Proposition 2 *The complexity I_S of a scene composed of the union of N identical subscenes with complexity $I_S{}'$ is equal to $I_S{}'$ plus the logarithm of N.*

Proof. Let S be a scene composed of N equal subscenes S'_k with $k \in \{1..N\}$. Thus, we have that $S = \bigcup_{k\in\{1..N\}} S'_k$. Let us also suppose that all the subscenes have the same discretisation and P and P' represent the set of patches of S and S'_k respectively. Thus, if l'_i and $\{F'_{ij}\}$ represent, respectively, the relative length of a segment in a subscene and the form factor matrix in a subscene, then $l_i = \frac{l'_i}{N}$ and $F_{ij} = F'_{ij}$ and it can be shown that

$$\begin{aligned}
I_S &= \sum_{i\in P}\sum_{j\in P} l_i F_{ij} \log \frac{F_{ij}}{l_j} \\
&= N \sum_{i\in P'}\sum_{j\in P'} l_i F_{ij} \log \frac{F_{ij}}{l_j} \\
&= N \sum_{i\in P'}\sum_{j\in P'} \frac{l'_i F'_{ij}}{N} \log(\frac{N F'_{ij}}{l'_j}) \\
&= \sum_{i\in P'}\sum_{j\in P'} l'_i F'_{ij} \log \frac{F'_{ij}}{l'_j} + \sum_{i\in P'}\sum_{j\in P'} l'_i F'_{ij} \log N \\
&= I'^d + \log N \sum_{i\in P'} l'_i = I'^d + \log N \qquad (4.14)
\end{aligned}$$

□

Results illustrating this proposition are shown in Table 4.12.

In Figure 4.20, a singularity occurs when the internal square adheres to the right upper hand corner of the square, and thus recreates the L-shaped room as in Figure 4.14c. So, we can see that when we convert a square scene with a square object inside it to an empty L-shaped scene, this produces a collapse in complexity

$$I_S^c(\text{Fig.4.20i}) \approx 3.304 \Longrightarrow I_S^c(\text{Fig.4.14c}) \approx 1.218$$

	square rotation	*pentagonal star*
$I^c_{subscene}$	1.474	1.750
number of subscenes (N)	4	5
$\log N$	2	2.322
$I^c_{subscene} + \log N$	3.474	4.072
I^c_S	3.477	4.067

Table 4.12: Analysis of I^c_S for Figures 4.19i and 4.18i.

Figure 4.21 shows the evolution of the three sequences of scenes which we have just discussed. We can observe that the scene with the rotating square has the most stable complexity. In the star scene, as the size of the star increases, the correlation of the scene changes dramatically.

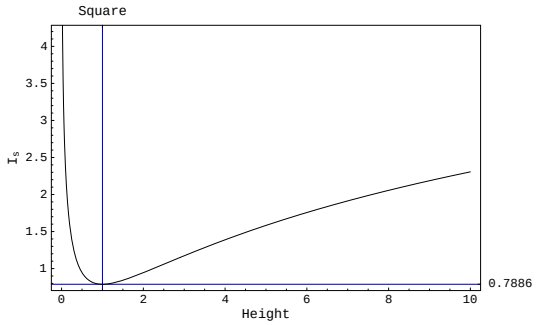


Figure 4.22: Analysis of I^c_S for a rectangle. The horizontal axis represents the number of times that the height is greater than the base.

The strongest singularity in a scene is produced when the space between the edges disappears. Figure 4.22 shows that when the height of a rectangle tends to zero, complexity tends to infinity. The same happens with concentric circles, concentric squares, and so on. In conclusion, complexity grows as singularities are approached.

Scene classification

From the previous results, we present a tentative scene classification:

- *Low complexity.* Simple empty scenes (without objects), like regular polygons, or scenes with few objects of low relative size, as in Figure 4.15a-c.
- *Medium complexity.* Scenes with few big objects, as in Figure 4.15d-f, scenes with more objects but with low relative size, as in Figure 4.16, or scenes with edges not too close to each other, as in Figures 4.19 and 4.20.
- *High complexity.* Scenes with a lot of objects, as in Figure 3.12, or scenes with very narrow spaces, as in Figure 4.11.

- *Very high complexity.* When all the edges of the scene are very close to each other, like very close concentric circles, a rectangle that is so long and narrow that it looks like a line, or a fractal scene whose walls form infinite almost closed cavities.

4.6 Summary

After analysing the notion of complexity, we have proposed in this chapter continuous scene mutual information (which expresses the degree of correlation between all the surfaces of a scene and measures with maximum accuracy the information transfer) as the main measure of the scene visibility complexity and discrete scene mutual information as the measure of the visibility complexity of a discretised scene.

We have solved the continuous mutual information integral by Monte Carlo integration, and the computation has been carried out efficiently by casting uniformly distributed global and local lines. We have also seen that discrete and continuous mutual information fulfil the two following properties:

- 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 I_S^c is the least upper bound to the discrete scene visibility mutual information I_S.

We have also described how continuous mutual information expresses the difficulty in obtaining a precise discretisation. So, the higher the scene complexity, the more difficult it is to obtain an accurate discretisation and probably more refinements will be necessary. The difference between continuous and discrete mutual information can be interpreted as the discretisation error, representing the loss of information transfer due to the discretisation.

We have calculated the scene complexity for many different cases and some experiments have illustrated the main reasons for the growth in complexity. Finally, a tentative scene classification has been proposed in flatland.

Chapter 5

Scene Radiosity Entropy and Complexity

In this chapter, we extend our previous results to the radiosity setting. Thus, measures for the entropy and complexity of a scene, taking diffuse illumination into account, are obtained by using a different pair of discrete and continuous Markov chains. A general proposition for visibility, radiosity, and importance enables us to predict the gain in mutual information resulting from refinement. Most of the contents of this chapter can be found in [21, 22].

5.1 From Visibility to Radiosity

In the previous chapters, we have only considered the visibility of a scene. In this section we make a leap forward and set the basis for the study of radiosity complexity.

Our work on visibility (chapters 3 and 4) has been based on the existence of a Markov chain and the knowledge of its stationary distribution. Thus, to study the complexity of a *scene with illumination*, we need to find an analog of the form factor matrix for the radiosity setting. This analog appears naturally when the null variance *probability transition matrix* (2.46)

$$P_{ij} = p_{j|i} = \frac{\rho_i F_{ij} B_j}{B_i - E_i} \tag{5.1}$$

is considered (section 2.1.10). This matrix corresponds to the transition probabilities that lead to null variance estimators [75]. The null variance matrix must have a preferred position between the different possible transition matrices.

The left eigenvalue property [13] is used to obtain (without normalization) the *stationary distribution*

$$p_i = A_i \frac{B_i - E_i}{\rho_i} B_i = A_i B_i^{in} B_i^{out} \tag{5.2}$$

where $B_i^{in} = \frac{B_i - E_i}{\rho_i}$ is the incoming radiosity and $B_i^{out} = B_i$ is the outgoing radiosity. It is easy to check that these probabilities fulfill the Bayes theorem

$p_{ij} = p_i p_{j|i} = q_j p_{i|j}$:

$$A_i \frac{B_i - E_i}{\rho_i} B_i \frac{\rho_i F_{ij} B_j}{B_i - E_i} = A_j \frac{B_j - E_j}{\rho_j} B_j \frac{\rho_j F_{ji} B_i}{B_j - E_j} \tag{5.3}$$

$$p_{ij} = A_i B_i F_{ij} B_j = A_j B_j F_{ji} B_i \tag{5.4}$$

This expression is an *extended reciprocity relation* (without normalization). Thus, the analogy is complete.

5.2 Discrete Channel

From the above assumptions, the entropy and mutual information can be defined straightforwardly for the radiosity setting using the following analogy:

- $A_i \Longrightarrow \mathsf{A}_i = A_i \frac{B_i - E_i}{\rho_i} B_i$
- $A_T \Longrightarrow \mathsf{A}_T = \sum_i \mathsf{A}_i$
- $a_i = \frac{A_i}{A_T} \Longrightarrow \mathsf{a}_i = \frac{\mathsf{A}_i}{\mathsf{A}_T}$
- $F_{ij} \Longrightarrow \mathsf{F}_{ij} = \frac{\rho_i F_{ij} B_j}{B_i - E_i}$

This analogy can be interpreted as a mapping of a given scene into a new (imaginary) scene, transforming the areas and the transition probabilities according to the above formulae. Studying the radiosity complexity of the original scene corresponds to studying the visibility complexity of the new scene.

5.2.1 Definitions

Now we can give the following definitions:

- *Discrete scene radiosity entropy*

$$\mathsf{H}_S = -\sum_{i=1}^{n_p} \mathsf{a}_i \sum_{j=1}^{n_p} \mathsf{F}_{ij} \log \mathsf{F}_{ij} \tag{5.5}$$

- *Positional entropy*

$$\mathsf{H}_P = -\sum_{i=1}^{n_p} \mathsf{a}_i \log \mathsf{a}_i \tag{5.6}$$

- *Discrete scene radiosity mutual information*

$$\mathsf{I}_S = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} \mathsf{a}_i \mathsf{F}_{ij} \log \frac{\mathsf{F}_{ij}}{\mathsf{a}_j} \tag{5.7}$$

This mapping from visibility to radiosity can also be extended to the concept of complexity. Thus, the discrete scene radiosity mutual information I_S represents the complexity of a discretised scene with diffuse illumination.

5.2.2 Particular Cases

To illustrate the behaviour of the above mapping, we study two particular cases:

1. **Constant radiosity**

   Let us consider the case where the resulting radiosity is constant for all patches. This will avoid introducing any additional complexity with regard to the visibility case and therefore the transformation (or analogy) considered above will become the identity transformation.

   To obtain constant radiosity B everywhere, we can easily find from the radiosity equation (2.6) that we must have $B = \rho_i B + E_i$, for all i. Then $B_i^{in} = \frac{B - E_i}{\rho_i} = \frac{\rho_i B}{\rho_i} = B$ and this means that

$$\frac{\mathsf{A}_i}{\mathsf{A}_T} = \frac{A_i B^2}{A_T B^2} = \frac{A_i}{A_T} \tag{5.8}$$

$$\mathsf{F}_{ij} = \frac{\rho_i F_{ij} B_j}{B_i - E_i} = F_{ij} \tag{5.9}$$

   We have computed the visibility and radiosity complexity for the labyrinth scene (Figure 5.1), with $\rho_i + E_i = 1$ for each patch. In this case, the radiosity is equal to 1 everywhere. The results shown in Table 5.1 confirm the theoretical prediction.

Setting	$H_P, \mathsf{H_P}$	$H_S, \mathsf{H_S}$	$I_S, \mathsf{I_S}$
Visibility	10.6883	6.47804	4.21022
Radiosity	10.6842	6.47833	4.20591

Table 5.1: Results for the labyrinth scene of Figure 5.1, with constant radiosity everywhere. For this case, both visibility and radiosity complexity are the same.

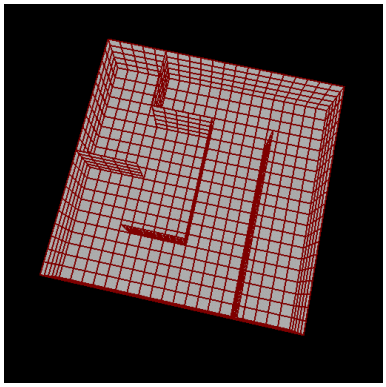


Figure 5.1: Labyrinth scene used to show the equivalence of visibility and radiosity complexity when radiosity is constant everywhere (see Table 5.1).

2. **Empty sphere with one source patch and constant reflectance**

   Let us consider the interior of an empty sphere with area A_S and constant reflectivity ρ, and emissivity E_s along one source patch with area A_s. From the fact that $F_{ij} = \frac{A_j}{A_S}$, the solution for any patch i is

$$B_i = \frac{\rho}{1-\rho}\frac{A_s E_s}{A_S} + E_i \tag{5.10}$$

   Consequently, for any patch, if we take

$$B = B_i - E_i = \frac{\rho}{1-\rho}\frac{A_s E_s}{A_S} \tag{5.11}$$

   we have

$$B_i^{in} = \frac{B_i - E_i}{\rho} = \frac{B}{\rho} \tag{5.12}$$

   After some algebra we find

$$\mathsf{A}_T = \sum_i \mathsf{A}_i = \frac{B^2}{\rho^2} A_S \tag{5.13}$$

   Thus, for any patch (but not the source), we have

$$\mathsf{a}_i = \frac{\frac{B^2}{\rho}A_i}{\frac{B^2}{\rho^2}A_S} = \frac{\rho A_i}{A_S} \tag{5.14}$$

   and for the source patch

$$\mathsf{a}_s = \frac{\frac{B}{\rho}A_s(B+E_s)}{\frac{B^2}{\rho^2}A_S} = \frac{\rho A_s(B+E_s)}{A_S B} = \frac{\rho A_s}{A_S} + (1-\rho) \tag{5.15}$$

   On the other hand, for a patch j which is not the source

$$\mathsf{F}_{ij} = \frac{\rho_i F_{ij} B_j}{B_i - E_i} = \frac{\rho A_j}{A_S} = \mathsf{a}_j \tag{5.16}$$

   and

$$\mathsf{F}_{is} = \frac{\rho_i F_{is} B_s}{B_i - E_i} = \frac{\rho A_s(B+E_s)}{A_S B} = \frac{\rho A_s}{A_S} + (1-\rho) = \mathsf{a}_s \tag{5.17}$$

   Hence, the net result is the transformation of the sphere into a sphere in which the relative area of each non-source patch has shrunk by a factor equal to the reflectivity, and in which the source patch has expanded to fill the gap. Thus, $\mathsf{I}_S = 0$.

5.2.3 Empirical Results

We have computed mutual information I_S and entropy H_S for the Cornell box scene shown in Figure 5.2 and for the tetrahedron in Figure 4.3 . Six different

discretisations have been generated for the Cornell box scene [1] and three for the tetrahedron scene. Observe that for I_S and H_S we give three independent color components (R,G,B).

In Tables 5.2 and 5.3, we present the results corresponding to the meshes in Figures 5.2 and 4.3. In both figures, the discrete mutual information reflects that a finer mesh is indeed a better mesh. Among equally fine meshes, the regular mesh is quantified to be the worst. The erratical behaviour of I_S, with few patches, is due to the inaccuracy of the radiosity solution obtained with so a coarse discretisation. In Figure 5.2, mutual information is very similar for all the channels, whereas entropy in the red channel is clearly the highest. In Figure 4.3, mutual information of the red channel is the highest, whereas the mutual information of the blue channel is zero (there is no information transfer), due to the fact that there is only one surface in the tetrahedron with a blue component. On the other hand, entropy is maximum in the blue channel and minimum in the green one.

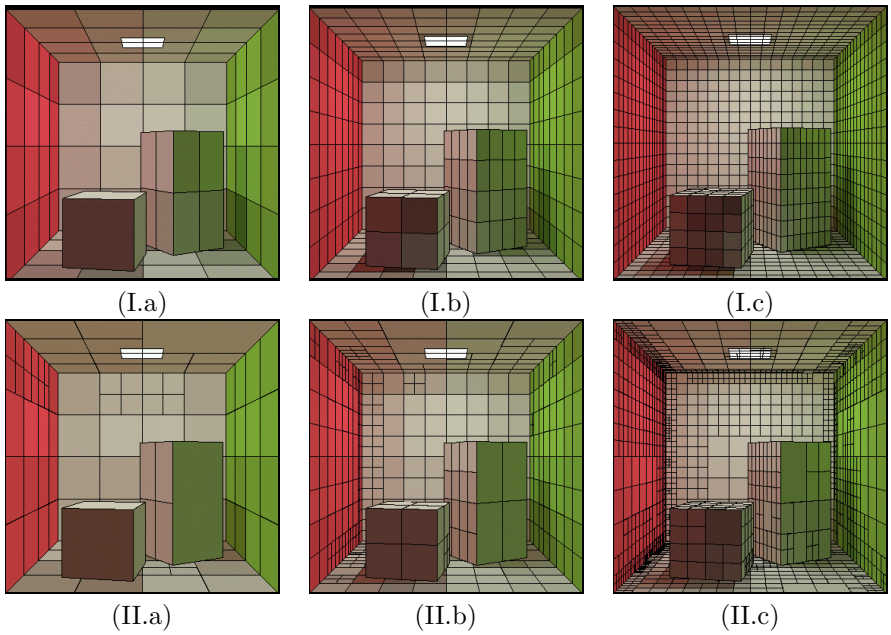


Figure 5.2: (a)Three regular and (b) three non-regular discretisations for the Cornell box scene.

5.3 Continuous Radiosity Mutual Information

In section 4.2, continuous visibility mutual information I_S^c has been derived from discrete visibility mutual information I_S. In a similar way, continuous radiosity mutual information I_S^c can be obtained from I_S with the following substitutions:

- $\mathsf{A}_i = A_i \frac{B_i - E_i}{\rho_i} B_i \Longrightarrow B(x) \frac{B(x) - E(x)}{\rho(x)}$

[1] The non-regular mesh has been generated with a mutual-information-based oracle which will be introduced in the following chapter.

Scene	*Patches*	*Lines* (10^6)	$\mathsf{I}_S(R,G,B)$	$\mathsf{H}_S(R,G,B)$
Fig.5.2I.a	121	10	(0.935, 0.965, 0.960)	(3.138, 2.636, 2.285)
Fig.5.2I.b	481	10	(0.958, 0.974, 0.959)	(4.210, 3.658, 3.278)
Fig.5.2I.c	1924	100	(1.004, 1.014, 0.993)	(6.170, 5.616, 5.243)
Fig.5.2II.a	121	10	(1.009, 1.028, 1.014)	(3.407, 2.989, 2.694)
Fig.5.2II.b	493	10	(1.017, 1.017, 0.996)	(5.077, 4.538, 4.214)
Fig.5.2II.c	1933	100	(1.052, 1.046, 1.020)	(6.849, 6.339, 6.059)

Table 5.2: I_S and H_S for the Cornell box scene (Figure 5.2) with six different discretisations.

Scene	*Patches*	*Lines* (10^6)	$\mathsf{I}_S(R,G,B)$	$\mathsf{H}_S(R,G,B)$
Fig.4.3a	4	10	(1.000,1.000,0.000)	(0.000,0.000,1.362)
Fig.4.3b	151	10	(2.912,2.158,0.000)	(4.003,3.077,6.117)
Fig.4.3c	793	10	(3.432,2.765,0.000)	(5.696,4.854,8.259)

Table 5.3: I_S and H_S for the tetrahedron (Figure 4.3) with three different discretisations.

- $\mathsf{A}_T = \sum_i A_i \frac{B_i - E_i}{\rho_i} B_i \Longrightarrow \mathsf{A}_T^c = \int_S B(x) \frac{B(x)-E(x)}{\rho(x)} dA_x$
- $\mathsf{F}_{ij} = \frac{\rho_i F_{ij} B_j}{B_i - E_i} \Longrightarrow \mathsf{F}(x,y) = \frac{\rho(x) F(x,y) B(y)}{B(x)-E(x)}$
- $\mathsf{a}_i \mathsf{F}_{ij} \Longrightarrow \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c}$

where S represents the total surface of the scene, $F(x,y)$ is the point-to-point form factor, $B(x)$, $E(x)$, and $\rho(x)$ are, respectively, the radiosity, self-emitted radiosity and reflectivity at $x \in S$.

From the continuous radiosity equation (2.4) we can see that

$$\begin{aligned} \int_S \mathsf{F}(x,y) dA_y &= \int_S \frac{\rho(x) F(x,y) B(y)}{B(x) - E(x)} dA_y \\ &= \frac{B(x) - E(x)}{B(x) - E(x)} = 1 \end{aligned} \tag{5.18}$$

Thus, the continuous radiosity mutual information is given by

$$\begin{aligned} \mathsf{I}_S^c &= \int_S \int_S \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c} \log \frac{\mathsf{A}_T^c F(x,y)}{\frac{B(x)-E(x)}{\rho(x)} \frac{B(y)-E(y)}{\rho(y)}} dA_x dA_y \\ &= \int_S \int_S \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c} \log \frac{\mathsf{A}_T^c F(x,y)}{B^{in}(x) B^{in}(y)} dA_x dA_y \end{aligned} \tag{5.19}$$

where $B^{in}(x) = \frac{B(x)-E(x)}{\rho(x)}$ represents the incoming radiosity at point x.

Similarly to I_S^c, which represents the scene complexity in the visibility setting, I_S^c quantifies the scene radiosity complexity. It expresses with maximum accuracy the information transfer in a scene with illumination and also the difficulty in obtaining an accurate discretisation. Note that I_S^c depends on the geometry, radiosity, emittance and reflectance of the scene.

Expression (5.19) could be computed using the same approach as for the continuous visibility mutual information whenever we know the exact radiosity

distribution over the scene and the value of A_T^c. Generally we know neither, so we have to make do with approximate values, computed using a piecewise constant function over all patches. The accuracy of the value found for I_S^c will depend on the quality of this distribution. Thus, contrary to the visibility case, we can not compute a continuous radiosity mutual information independent of any discretisation.

Applying the following substitutions: A_T^c by A_T, $B(x)$ by B_i, $B(y)$ by B_j, $E(x)$ by E_i, $E(y)$ by E_j, and $\rho(x)$ by ρ_i, $\rho(y)$ by ρ_j , and proceeding as in the visibility case (section 4.3), we obtain by Monte Carlo integration

$$\mathsf{I}_S^c \approx \frac{A_T}{N} \sum_{k=1}^{N} \frac{B_i B_j}{\mathsf{A}_T} \log\Big(\frac{\mathsf{A}_T cos\theta_{x_k} cos\theta_{y_k}}{\pi r_{x_k y_k}^2 B_i^{in} B_j^{in}}\Big) \tag{5.20}$$

where N represents the number of local lines or the number of segments of global lines.

As in the case of visibility, we have that

- The difference between I_S^c and I_S represents the loss of information transfer due to the discretisation. Thus, the global discretisation error for radiosity is given by
$$\delta^r = \mathsf{I}_S^c - \mathsf{I}_S \tag{5.21}$$
- Between different discretisations of the same scene, the most precise will be the one that has the highest discrete mutual information I_S.

5.3.1 Particular Cases

From a continuous perspective, we can now analyze the same cases we dealt with in section 5.2.2:

1. **Constant radiosity**

   When radiosity B is constant everywhere, we have
   $$\frac{B(x) - E(x)}{\rho(x)} = B(x) = B \tag{5.22}$$
   and thus
   $$\mathsf{A}_T^c = \int_S \frac{B(x) - E(x)}{\rho(x)} B(x) dA_x = B^2 A_T \tag{5.23}$$
   Thus continuous radiosity mutual information becomes
   $$\begin{aligned} \mathsf{I}_S^c &= \int_S \int_S \frac{F(x,y)B^2}{B^2 A_T} \log \frac{B^2 A_T F(x,y)}{B^2} \\ &= \int_S \int_S \frac{F(x,y)}{A_T} \log A_T F(x,y) \end{aligned} \tag{5.24}$$
   which is the continuous visibility mutual information.

2. **Empty sphere with one source patch and constant reflectance**

   For the discrete case, we know that $\mathsf{I}_S = 0$. Thus, the continuous mutual information should also be zero.

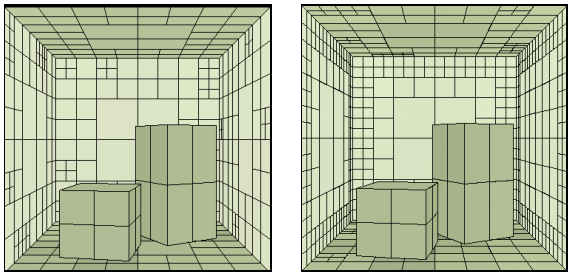
It can be proved that the radiosity solution for the continuous case is the same as for the discrete one. Thus, using the result of section 5.2.2, $\mathsf{A}_T^c = \mathsf{A}_T = \frac{B^2}{\rho^2} A_S$, continuous mutual information becomes

$$\begin{aligned} \mathsf{I}_S^c &= \int_S \int_S \frac{F(x,y)B^2}{\frac{B^2}{\rho^2}A_S} \log \frac{A_S \frac{B^2}{\rho^2} F(x,y)}{\frac{B^2}{\rho^2}} \\ &= \int_S \int_S \frac{F(x,y)\rho^2}{A_S} \log A_S F(x,y) = 0 \end{aligned} \tag{5.25}$$

because in a sphere $F(x,y) = \frac{1}{A_S}$.

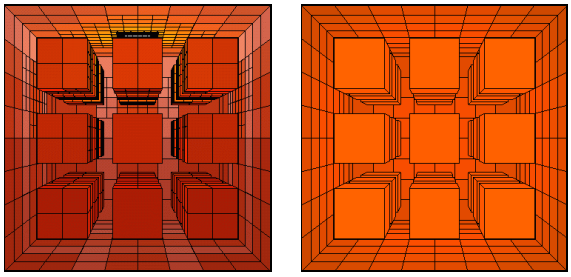
5.3.2 Empirical Results

In Table 5.4, we present the scene complexity I_S^c for the scenes in Figures 4.3, 4.4, 4.9, 5.3 and 5.4. In this table, we can observe how I_S^c varies with the number of patches. This drawback is due to the fact that we take constant radiosity on each patch. Obviously, the precision of I_S^c increases with the number of patches. From the results corresponding to Figure 5.4, we see that the scene complexity increases with the number of emitter patches. Thus, as expected, the collocation and the number of the light sources changes the correlation in a scene dramatically.



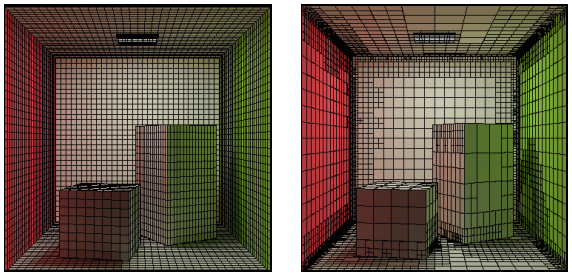
(a) (b)

Figure 5.3: Two different discretisations for the same scene. The walls are light sources.



(a) (b)

Figure 5.4: (a) The scene has only one light source on the ceiling. (b) All the surfaces of the cubes, in the center of the scene, are light sources.



(a) (b)

Figure 5.5: Two different discretisations of the Cornell box scene.

In Table 5.5, corresponding to Figures 5.2 and 5.5, we show how I_S depends on the strategy (oracle) used in the refinement process. Observe that I_S is larger in the scenes obtained with a mutual-information-based oracle (Figure 5.2II) which will be introduced in the next chapter. Thus, from an information-theory point of view, the best results, compared with the ones obtained with a regular mesh, are found using an oracle based on the variation of the radiosity kernel. In most of the experiments, we can also note that, for finer meshes, I_S increases and I^c_S decreases. Observe that, in Figure 5.5b, the values of I_S are higher than the values of I^c_S. This fact is due to an insufficient number of lines cast.

Scene	*Patches*	*Lines* (10^6)	I^c_S*(R,G,B)*
Fig.4.4a	19	10	(1.676,1.676,1.676)
Fig.4.4b	196	10	(1.654,1.654,1.654)
Fig.4.4c	826	10	(1.286,1.286,1.286)
Fig.4.9a	1924	100	(1.163,1.163,1.163)
Fig.4.9b	1924	100	(1.129,1.129,1.129)
Fig.4.9c	1924	100	(1.375,1.375,1.375)
Fig.5.3a	444	50	(2.443,2.443,2.443)
Fig.5.3b	759	100	(2.455,2.455,2.455)
Fig.5.4a	2029	100	(2.149,1.481,1.980)
Fig.5.4b	546	100	(5.087,5.300,0.541)
Fig.4.3a	4	10	(3.219,3.197,0.000)
Fig.4.3b	151	10	(3.419,3.102,0.000)
Fig.4.3c	793	10	(3.629,3.158,0.000)

Table 5.4: I^c_S for different scenes.

As we can see in Table 5.6, corresponding to Figure 5.4a, similar to the computation of I_S, the value of I_S decreases with the number of local or global lines used to compute the form factors. Observe also that the computational cost of I^c_S is much lower than the cost of I_S.

5.4 Patch Refinement

In this section, we present a proposition that supports our preliminary results (section 4.4) and allows us to calculate the gain in mutual information resulting

Scene	*Patches*	*Lines* (10^6)	I_S*(R,G,B)*	I^c_S*(R,G,B)*
Fig.5.2I.a	121	10	(0.935, 0.965, 0.960)	(1.132, 1.144, 1.107)
Fig.5.2I.b	481	10	(0.958, 0.974, 0.959)	(1.055, 1.066, 1.037)
Fig.5.2I.c	1924	100	(1.004, 1.014, 0.993)	(1.044, 1.051, 1.024)
Fig.5.5a	7744	1000	(1.027, 1.028, 1.013)	(1.032, 1.032, 1.016)
Fig.5.2II.a	121	10	(1.009, 1.028, 1.014)	(1.194, 1.194, 1.138)
Fig.5.2II.b	493	10	(1.017, 1.017, 0.996)	(1.082, 1.080, 1.043)
Fig.5.2II.c	1933	100	(1.052, 1.046, 1.020)	(1.070, 1.064, 1.032)
Fig.5.5b	6595	500	(1.078, 1.064, 1.046)	(1.056, 1.048, 1.027)

Table 5.5: I^c_S and I_S for different discretisations of the same scene.

Lines (10^6)	I_S*(R,G,B)*	I^c_S*(R,G,B)*
1	(3.563, 3.006, 5.366)	(2.141, 1.472, 1.996)
10	(2.269, 1.631, 2.519)	(2.145, 1.479, 1.948)
100	(2.094, 1.456, 1.605)	(2.149, 1.481, 1.980)

Table 5.6: I^c_S and I_S for the scene in Figure 5.4a.

from subdivision of scene patches. We first study the problem for a general Markov Chain and next consider the application to scene visibility, radiosity, and importance.

5.4.1 State Refinement and Continuous versus Discrete Mutual Information

Proposition 3 *Consider a discrete Markov chain over a set of states labeled* $i, j = 1, \ldots, n$*, with transition probability matrix* $\mathbf{P} = (P_{ij})$ *and stationary distribution* $\mathbf{w} = (w_1, w_2, \ldots, w_n)$ *which satisfies the reciprocity relation* $w_i P_{ij} = w_j P_{ji}$ $(\forall i, j)$*. 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. Let us imagine a discrete random walk with discrete mutual information

$$I = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i P_{ij} \log \frac{P_{ij}}{w_j} \tag{5.26}$$

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

$$\begin{aligned} I' = &\sum_{i=1}^{n-1} \sum_{j=1}^{n-1} w'_i P'_{ij} \log \frac{P'_{ij}}{w'_j} + \sum_{i=1}^{n-1} \sum_{k=1}^{m} w'_i P'_{i n_k} \log \frac{P'_{i n_k}}{w_{n'_k}} \\ + &\sum_{k=1}^{m} \sum_{j=1}^{n-1} w'_{n_k} P'_{n_k j} \log \frac{P'_{n_k j}}{w'_j} + \sum_{k=1}^{m} \sum_{l=1}^{m} w'_{n_k} P'_{n_k n_l} \log \frac{P'_{n_k n_l}}{w'_{n_l}} \end{aligned} \tag{5.27}$$

where

- $w_i = w'_i$ for $1 \leq i < n$
- $w_n = \sum_{k=1}^{m} w'_{n_k}$
- $P_{ij} = P'_{ij}$ for $1 \leq i, j < n$
- $P_{in} = \sum_{k=1}^{m} P'_{in_k}$ for $1 \leq i < n$

Since $w_i P_{ij} \log \frac{P_{ij}}{w_j} = w_j P_{ji} \log \frac{P_{ji}}{w_i}, \forall i, j$, we have

$$I = 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} w_i P_{ij} \log \frac{P_{ij}}{w_j} + \sum_{i=1}^{n} w_i P_{ii} \log \frac{P_{ii}}{w_i} \tag{5.28}$$

Then

$$\begin{aligned} I' - I = 2 \sum_{i=1}^{n-1} \left(\sum_{k=1}^{m} w_i P_{in_k} \log \frac{P_{in_k}}{w_{n_k}} - w_i P_{in} \log \frac{P_{in}}{w_n} \right) \\ + \sum_{k=1}^{m} \sum_{l=1}^{m} w_{n_k} P_{n_k n_l} \log \frac{P_{n_k n_l}}{w_{n_l}} - w_n P_{nn} \log \frac{P_{nn}}{w_n} \end{aligned} \tag{5.29}$$

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 (2.67)

$$\sum_{i=1}^{n} a_i \log \frac{a_i}{b_i} \geq \left(\sum_{i=1}^{n} a_i \right) \log \frac{\sum_{i=1}^{n} a_i}{\sum_{i=1}^{n} b_i} \tag{5.30}$$

we can conclude that

$$\Delta I = I' - I \geq 0 \tag{5.31}$$

□

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 [15]. Hence, the statement that I^c_S is the least upper bound to I_S immediately follows from the above proposition. □

5.4.2 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

$$\begin{aligned} (\Delta I)_{ij} &= 2 \left(\left(\sum_{k=1}^{m} w_{i_k} P_{i_k j} \log \frac{P_{i_k j}}{w_j} \right) - w_i P_{ij} \log \frac{P_{ij}}{w_j} \right) \\ &= 2 \left(\left(\sum_{k=1}^{m} w_{i_k} P_{i_k j} \log P_{i_k j} \right) - w_i P_{ij} \log P_{ij} \right) \end{aligned} \tag{5.32}$$

This can be obtained from (5.28) and (5.29), where the second half of these formulae is zero, and from the conditions of the proposition. For a regular subdivision, $w_{i_k} = \frac{w_i}{m}$, we have

$$(\Delta I)_{ij} = 2\left(\left(\frac{w_i}{m}\sum_{k=1}^{m} P_{i_k j}\log P_{i_k j}\right) - w_i P_{ij}\log P_{ij}\right) \tag{5.33}$$

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 that is not zero can be shown to be equal to mP_{ij}. Thus the maximum possible increase in I is given by

$$max((\Delta I)_{ij}) = 2(w_i P_{ij}\log mP_{ij} - w_i P_{ij}\log P_{ij}) = 2w_i P_{ij}\log m \tag{5.34}$$

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

$$\sum_j max((\Delta I)_{ij}) = \sum_j 2w_i P_{ij}\log m = 2w_i \log m \tag{5.35}$$

5.4.3 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 proposition 3 are fulfilled. Thus, from (5.32), the increase in mutual information is, in this case,

$$\begin{aligned}(\Delta I)^v_{ij} &= 2\left(\left(\sum_{k=1}^{m} a_{i_k} F_{i_k j}\log\frac{F_{i_k j}}{a_j}\right) - a_i F_{ij}\log\frac{F_{ij}}{a_j}\right)\\ &= 2\left(\left(\sum_{k=1}^{m} a_{i_k} F_{i_k j}\log F_{i_k j}\right) - a_i F_{ij}\log F_{ij}\right)\end{aligned} \tag{5.36}$$

Thus, the maximum increase upon a regular subdivision is

$$max((\Delta I)^v_{ij}) = 2a_i F_{ij}\log m \tag{5.37}$$

and the maximum possible increase when dividing a given patch i is $2a_i \log m$.

5.4.4 Application to Radiosity

In the radiosity setting, we consider the following transition probabilities

$$P_{ij} = \frac{\int_{A_i}\int_{A_j} F(x,y)B(x)B(y)dA_x dA_y}{\int_{A_i} B(x)\frac{B(x)-E(x)}{\rho(x)}dA_x} \tag{5.38}$$

These are an extension to the continuous case of the discrete null variance probabilities (2.46) and fulfil $\sum_j P_{ij} = 1$ due to the additivity of the integral over its domain S, where $S = \cup_i A_i$, and the fact that the radiosities fulfil the continuous radiosity equation (2.4)

$$B(x) = E(x) + \rho(x)\int_S F(x,y)B(y)dA_y \tag{5.39}$$

It can be easily checked that the stationary probabilities (without normalization) are

$$w_i = \int_{A_i} B(x) \frac{B(x) - E(x)}{\rho(x)} dA_x \tag{5.40}$$

and the reciprocity relation $w_i P_{ij} = w_j P_{ji}$ is fulfilled. The normalising factor of w_i is

$$\sum_i \int_{A_i} B(x) \frac{B(x) - E(x)}{\rho(x)} dA_x = \int_S B(x) \frac{B(x) - E(x)}{\rho(x)} dA_x \tag{5.41}$$

If we divide patch i into i_1 and i_2, it is easy to prove that the hypotheses of proposition 3 are fulfilled. The radiosity case reverts to the visibility case when $B(x) = k$, where k is a constant, and this happens whenever $E(x) = k(1-\rho(x))$, $\forall x$.

Now, let us suppose that radiosities and reflectivities are constant along each patch. From section (5.1), we have $w_i = A_i B_i \frac{(B_i - E_i)}{\rho_i}$ and $P_{ij} = \frac{\rho_i F_{ij} B_j}{B_i - E_i}$. These quantities can be considered a kind of generalized area and form factor, respectively, by analogy with the visibility case. The P_{ij} probabilities were found to be the null variance transition probabilities for a gathering random walk (section 2.1.10).

Since the hypotheses of proposition 3 are fulfilled, from (5.32) we obtain the increase in mutual information:

$$(\Delta I)^r_{ij} = 2 \frac{B_i B_j}{\mathsf{A}_T} \left(\left(\sum_{k=1}^{m} A_{i_k} F_{i_k j} \log F_{i_k j} \right) - A_i F_{ij} \log F_{ij} \right) \tag{5.42}$$

5.4.5 Application to Importance

The continuous importance $I(x)$, given initial importance $V(x)$, is the solution to the integral equation for importance at a point x [59]:

$$I(x) = V(x) + \int_S \rho(y) F(x,y) I(y) dA_y \tag{5.43}$$

Consider now the transition probability

$$P_{ij} = \frac{\int_{A_i} \int_{A_j} F(x,y) \rho(x) \rho(y) I(x) I(y) dA_x dA_y}{\int_{A_i} \rho(x) I(x) (I(x) - V(x)) dA_x} \tag{5.44}$$

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

$$w_i = \int_{A_i} \rho(x) I(x) (I(x) - V(x)) dx \tag{5.45}$$

and the reciprocity relation $w_i P_{ij} = w_j P_{ji}$ is fulfilled. As in the case of radiosity, if we divide patch i into i_1 and i_2, the hypotheses of the proposition 3 are

fulfilled. It can be seen that for $\rho(x)I(x) = k$, $\forall x$, which occurs when we take $V(x) = k(\frac{1}{\rho(x)} - 1)$, the importance case reverts to the visibility case.

Now, let us suppose that importances and reflectivities are constant along each patch. In this case, $w_i = A_i\rho_i I_i(I_i - V_i)$ and $P_{ij} = \frac{\rho_j F_{ji} I_j}{I_i - V_i}$. When $V_i = \delta_{ik}$, we have the null variance transition probabilities for a shooting random walk (section 2.1.10).

Since the hypotheses of the proposition 3 are fulfilled, this means that the increase in mutual information may be now calculated in the importance setting [2].

5.5 Summary

In this chapter, we have obtained measures for the entropy and complexity of a scene with diffuse illumination by using a different pair of discrete and continuous Markov chains. To study the complexity of a scene with illumination, we have applied an analog of the form factor matrix for the radiosity setting. The feasibility of this approach has been illustrated by a set of experiments.

Continuous scene radiosity mutual information expresses with maximum accuracy the information transfer in an illuminated scene. Similarly to the visibility setting, it represents the scene radiosity complexity and can be interpreted as the difficulty in obtaining an accurate meshing. We solved the mutual information integral by Monte Carlo integration, using an approximated solution for the radiosity.

To deal with the loss of information due to the discretisation, a general proposition gives us the increase in mutual information when a patch is subdivided into n subpatches. This increase has been calculated for visibility, radiosity, and importance. This proposition is fundamental for understanding the behaviour of the discrete mutual information with respect to the continuous mutual information.

[2] In this thesis, information theory measures for importance are not developed further.

Chapter 6

Refinement Criteria

As noted in the introduction, in the radiosity method, scene discretisation has to accurately represent illumination variations but it has to avoid unnecessary refinements that would increase the computational cost. To achieve this objective, in this chapter we introduce some mutual-information-based refinement criteria (oracles) for hierarchical radiosity which are based on the information-theory principles described in chapters 4 and 5.

This chapter is organized as follows. First, we analyze the feasibility of mutual information to be used as the basis for refinement criteria. Second, from the results obtained in section 5.4.2 we propose a first oracle based on an hypothetical increase in mutual information when a patch is refined. Third, we present a refinement criterion based on the loss of information transfer between two patches due to the discretisation. And finally, an oracle based on the information transfer loss of a patch is introduced for the visibility in flatland. Most of the contents of this chapter have been presented in [21, 22, 25].

6.1 Mutual Information Maximization

The oracles we will propose in this chapter have a common aim: *to maximize the discrete mutual information*. With this objective in mind, two approaches are analyzed:

- In the first approach, an oracle based on the *increase in mutual information when a patch is refined* (5.42) is theoretically explored. This approach presents some drawbacks which will be analyzed.
- The second approach is based on the *discretisation error* and avoids the disadvantages of the first one. It can be implemented on any hierarchical radiosity algorithm.

In this section, we show experimentally how the objective of maximizing the discrete mutual information can guide us towards an optimal discretisation. From (5.36), we analyze how mutual information varies for some common patch configurations and some simple scenes, and we provide some evidence that an optimal subdivision obtained by mutual information maximization corresponds satisfactorily to an optimal subdivision in terms of the mean square error.

The following results can be obtained from formula (5.36), form factor properties (section 2.1.3) and some closed form formulae for the unoccluded form factors [12, 82, 33]:

Partially occluded pair of patches

Consider the subdivision of patch i into two sub-patches (Figure 6.1a):

- **Proposition 4** *Of all the subdivisions of i with one sub-patch totally occluded to j, the maximum mutual information increase corresponds to the discontinuity mesh.*

  Proof. 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 (5.36) and assuming without loss of generality that $A_T = 1$,

$$\begin{aligned}(\Delta I)^v_{ij} &= 2(A_{i_a}F_{i_aj}\log\frac{F_{i_aj}}{A_j} + A_{i_b}F_{i_bj}\log\frac{F_{i_bj}}{A_j} - A_iF_{ij}\log\frac{F_{ij}}{A_j}) \\ &= 2((A_{i_c}F_{i_cj} + A_{i_d}F_{i_dj})\log\frac{(F_{ji_c} + F_{ji_d})}{(A_{i_c} + A_{i_d})} - A_iF_{ij}\log\frac{F_{ij}}{A_j}) \\ &= 2(A_{i_c}F_{i_cj}\log\frac{F_{ji_c}}{(A_{i_c} + A_{i_d})} - A_iF_{ij}\log\frac{F_{ij}}{A_j}) \qquad (6.1)\end{aligned}$$

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

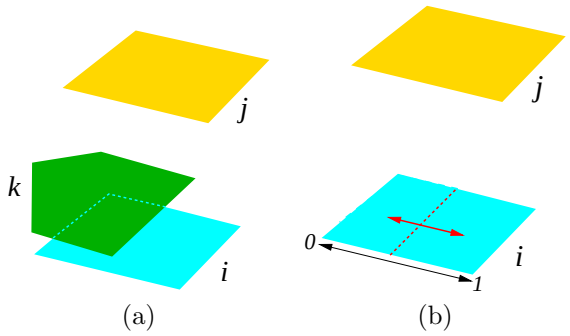


Figure 6.1: (a) Partially occluded patch pair. (b) Subdivision of a patch perpendicular to the radiosity gradient. The position of the cutting line is parametrized by the relative distance $0 \leq l \leq 1$ to one edge.

- **Proposition 5** *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.*

  Proof. 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.

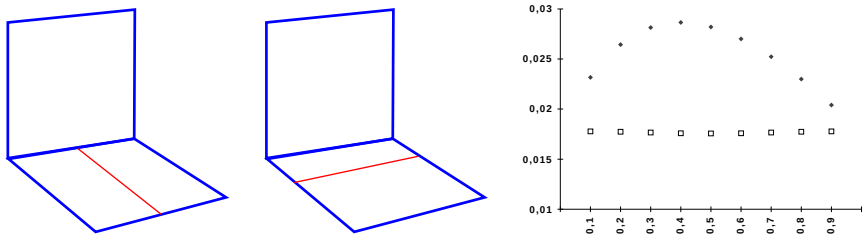
Then, from (5.36) and taking $A_T = 1$,

$$\begin{aligned}(\Delta I)^v_{ij} &= 2(A_{i_a}F_{i_aj}\log\frac{F_{i_aj}}{A_j} + A_{i_b}F_{i_bj}\log\frac{F_{i_bj}}{A_j} - A_iF_{ij}\log\frac{F_{ij}}{A_j}) \\ &= 2(A_{i_a}F_{i_aj}\log\frac{F_{i_aj}}{A_j} + A_{i_d}F_{i_dj}\log\frac{F_{ji_d}}{(A_{i_d}+A_{i_c})} - A_iF_{ij}\log\frac{F_{ij}}{A_j}) \\ &\leq 2(A_{i_a}F_{i_aj}\log\frac{F_{i_aj}}{A_j} + A_{i_d}F_{i_dj}\log\frac{F_{ji_d}}{A_{i_d}} - A_iF_{ij}\log\frac{F_{ij}}{A_j}) \quad (6.2)\end{aligned}$$

where the right-hand side of the inequality corresponds to the mutual information increase in the discontinuity meshing case, where we have assumed, by hypothesis, $F_{i_aj} = F_{i_dj}$. □

Two square patches with common edge

Consistent with observations in [9], orthogonal splitting (Figure 6.2a) leads to only a small gain in mutual information [1]. Nothing is gained by orthogonal splitting in the middle. When splitting along a line parallel with the common edge (Figure 6.2b) the maximum gain in mutual information results when splitting at a 40% relative distance from the edge (Figure 6.2c).



(a) (b) (c)

Figure 6.2: (c) Discrete mutual information, on the vertical axis, when dividing orthogonal (a,squares) and parallel (b,diamonds) to the common edge. Horizontal axis represents the displacement from one side edge (a) or the common edge (b).

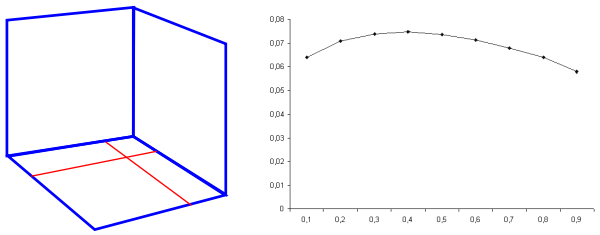
Three square patches with common edges

When subdividing a patch in a corner (Figure 6.3), the maximum mutual information gain [2] is obtained at a 39% relative distance 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 6.3b, squares).

Empty cube

In Figure 6.4, the resulting maximum mutual information subdivision is a bit displaced towards the edges with respect to the regular one. Figure 6.5 shows

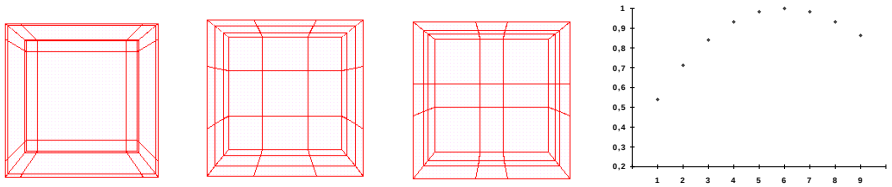
[1] In the mutual information computation we only consider the two patches involved.
[2] In the mutual information computation we only consider the three patches involved.



(a) (b)

Figure 6.3: (b) Discrete mutual information, on the vertical axis, when subdividing a patch in a corner (a). The horizontal axis represents the distance from the parallel division to one common edge.

an example with more subdivisions.



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

Figure 6.4: (d) Discrete mutual information, on vertical axis, for an empty cube. The horizontal axis represents the relative displacement of the nearest subdivision to a common edge, ranging from 0 to 10. (b) corresponds to the optimal case, with a value of almost 6, (a) with a value of 2 and (c) with a value of 8.

Mutual information maximization and mean square error

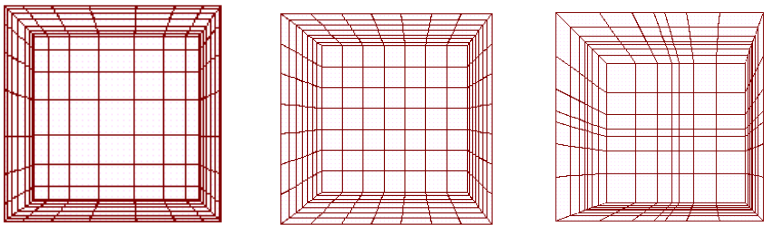
Consider two square patches, i and j, with the following characteristics: B_j is constant over j, $F(x,y)$ is approximately constant for $x \in S_i$ and $y \in S_j$ and the reflectivity is constant along each patch. Consider now that the radiosity on patch i varies along one axis parallel to one edge of i, $B(l) = l^n B$, where B is constant and l, which is between 0 and 1, parametrizes the patch (Figure 6.1(b)). The increase in mutual information when dividing patch i into sub-patches i_1 and i_2 is obtained from (5.42):

$$(\Delta I)^r_{ij} = 2\frac{B_i B_j}{\mathsf{A}_T}(A_{i_1} F_{i_1 j} \log F_{i_1 j} + A_{i_2} F_{i_2 j} \log F_{i_2 j} - A_i F_{ij} \log F_{ij})$$

The maximum increase in mutual information is achieved when the patch i is split perpendicular to the gradient (that 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 $l = 0.48, 0.61, 0.68, 0.74$.



(a) (b) (c)

Figure 6.5: An empty cube with (a) optimal ($I_S = 1.3569$), (b) regular ($I_S = 1.3331$) and (c) "bad" ($I_S = 1.2554$) subdivision.

Consider now the subdivision problem from the point of view of minimising the MSE error on patch i, while assuming constant values for the radiosities on the sub-patches (equal to the average of the continuous radiosity function $B(x)$). After some algebra, 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 MSE error.

6.2 An Oracle Based on the Increase in Mutual Information

In this section we present a theoretical analysis for an oracle based on the increase in mutual information when a patch is refined.

It has been seen (section 5.4.4) that, in the case of constant radiosity, the increase in mutual information is given by (5.42)

$$(\Delta I)_{ij}^r = 2\frac{B_i B_j}{\mathsf{A}_T}\left(\left(\sum_{k=1}^{m} A_{i_k} F_{i_k j} \log F_{i_k j}\right) - A_i F_{ij} \log F_{ij}\right)$$

Observe that this increase is strongly dependent on the subdivision type. To avoid this dependence, we take the maximum possible increase for a regular subdivision:

$$\begin{aligned} \max((\Delta I)_{ij}^r) &= \frac{2}{\mathsf{A}_T} B_i B_j \left(\left(\frac{A_i}{m} m F_{ij} \log(m F_{ij})\right) - A_i F_{ij} \log F_{ij}\right) \\ &= \frac{2}{\mathsf{A}_T} B_i B_j A_i F_{ij} \log m \propto B_i B_j A_i F_{ij} \end{aligned} \tag{6.3}$$

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 value can be rather different from the real gain obtained, when for instance

the form factors are fairly equal in the subdivisions, as they may be with two parallel patches at some distance and without occlusions (in this case there is no gain). Thus, the use of this quantity as an oracle for subdividing is not recommended. A better way is to use the full expression for ΔI, or at the least some information on form factor gradients along subdivisions should be taken into account (in general, the larger the gradient, the larger the increase in mutual information).

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

$$\sum_j B_i B_j A_i F_{ij} = A_i B_i \frac{B_i - E_i}{\rho_i} \tag{6.4}$$

to find the one with the largest generalized area A_i (section 5.2). Thus, lacking any other information, a heuristic method would be to subdivide the patch with the largest generalized area.

To sum up, one proposal for an oracle for hierarchical radiosity subdivision is as follows:

- A patch of the pair (i, j) will be a candidate for subdivision only when the quantity $B_i B_j A_i F_{ij} > \epsilon_1$. This discards subdivisions with a small potential increment of mutual information.
- If a pair (i, j) is considered, choose one of them: the one corresponding to

$$\max(A_i B_i \frac{B_i - E_i}{\rho_i}, A_j B_j \frac{B_j - E_j}{\rho_j}) \tag{6.5}$$

  that 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. The intention here is 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 have to point out here that the first step in the oracle is analogous to the power oracle $\rho_i A_i F_{ij} B_j$ (2.48), the second step can be seen as an extension to the heuristics of dividing the patch with the largest area, and the third step is analogous to the various gradient oracles used in hierarchical radiosity literature [9]. 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.

This oracle has two main drawbacks:

- It is based on a Markov chain whose stationary distribution and transition probabilities are taken from a converged scene. However, this oracle should be used on a hierarchical radiosity algorithm, where we do not have the final radiosities of patches.
- The hypothetical maximum mutual information gain can be quite different from the real increase obtained.

We will present in section 6.4 an oracle based on discretisation error that will overcome these disadvantages.

6.3 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. In this section we give the discretisation error for visibility and radiosity.

6.3.1 Mutual Information Matrix

In order to represent visibility and radiosity in the same formula, we adopt a general notation. So, $\mathbf{I}_S$ represents I_S or I_S, $\mathbf{F}_{ij}$ represents F_{ij} or F_{ij}, and $\mathbf{a}_i$ represents a_i or a_i in a consistent way. In the continuous case, $\mathbf{I}^c_S$ stands for I^c_S or I^c_S.

So, in general, discrete scene mutual information is given by

$$\mathbf{I}_S = \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \mathbf{a}_i \mathbf{F}_{ij} \log\Big(\frac{\mathbf{F}_{ij}}{\mathbf{a}_j}\Big) \tag{6.6}$$

From this formula, the term

$$\mathbf{I}_{ij} = \mathbf{a}_i \mathbf{F}_{ij} \log\Big(\frac{\mathbf{F}_{ij}}{\mathbf{a}_j}\Big) \tag{6.7}$$

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

$$\mathbf{I}_i = \sum_{j=1}^{n_p} \mathbf{a}_i \mathbf{F}_{ij} \log\Big(\frac{\mathbf{F}_{ij}}{\mathbf{a}_j}\Big) \tag{6.8}$$

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

$$\mathbf{I}_S = \sum_{i=1}^{n_p} \mathbf{I}_i = \sum_{i=1}^{n_p}\sum_{j=1}^{n_p} \mathbf{I}_{ij} \tag{6.9}$$

If we analyze the terms $\mathbf{I}_{ij}$, we observe that negative values appear when $\mathbf{F}_{ij} < \mathbf{a}_j$. This situation reflects a very low interaction between the two patches involved. On the other hand, using the concavity property of the logarithm function (2.67), it is easy to see that $\mathbf{I}_i \geq 0$ (substituting a_k, b_k, and n by $\mathbf{F}_{ij}$, $\mathbf{a}_j$, 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 for visibility and radiosity, we obtain:

- **For visibility**

  Continuous information transfer:

$$\begin{aligned} I_S^c &= \int_S \int_S \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \\ &= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \end{aligned} \quad (6.10)$$

  Continuous information transfer due to patch i:

$$I_i^c = \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \quad (6.11)$$

  Continuous information transfer between patches i and j:

$$I_{ij}^c = \int_{A_i} \int_{A_j} \frac{1}{A_T} F(x,y) \log(A_T F(x,y)) dA_x dA_y \quad (6.12)$$

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

- **For radiosity**

  Continuous information transfer:

$$\begin{aligned} \mathsf{I}_S^c &= \int_S \int_S \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c} \log \frac{\mathsf{A}_T^c F(x,y)}{B^{in}(x)B^{in}(y)} dA_x dA_y \\ &= \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c} \log \frac{\mathsf{A}_T^c F(x,y)}{B^{in}(x)B^{in}(y)} dA_x dA_y \end{aligned} \quad (6.13)$$

  Continuous information transfer due to patch i:

$$\mathsf{I}_i^c = \sum_{j=1}^{n_p} \int_{A_i} \int_{A_j} \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c} \log \frac{\mathsf{A}_T^c F(x,y)}{B^{in}(x)B^{in}(y)} dA_x dA_y \quad (6.14)$$

  Continuous information transfer between patches i and j:

$$\mathsf{I}_{ij}^c = \int_{A_i} \int_{A_j} \frac{F(x,y)B(x)B(y)}{\mathsf{A}_T^c} \log \frac{\mathsf{A}_T^c F(x,y)}{B^{in}(x)B^{in}(y)} dA_x dA_y \quad (6.15)$$

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

6.3.2 Discretisation Error Between Two Patches

As we have seen (sections 4.4.2 and 5.3), a general discretisation error for a scene can be given by

$$\delta = \mathbf{I}_S^c - \mathbf{I}_S \geq 0 \quad (6.16)$$

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 visibility and radiosity. For each one, three different Monte Carlo techniques can be used: patch-to-patch random lines, local lines and global lines.

- **For visibility**

  1. **Patch-to-patch random lines:** The computation of (6.12)

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

can be done with an area-to-area sampling (section 2.1.6), 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

$$I_{ij}^c \approx \frac{A_i A_j}{A_T}\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log\Big(F(x_k,y_k)A_T\Big) \quad (6.17)$$

where x_k and y_k are, respectively, the end-points on patches i and j of the k-th line.
From (6.7), I_{ij} can be expressed as

$$I_{ij} = \frac{A_i F_{ij}}{A_T}\log\Big(\frac{F_{ij}A_T}{A_j}\Big) = \frac{A_i A_j}{A_T}\frac{F_{ij}}{A_j}\log\Big(\frac{F_{ij}}{A_j}A_T\Big) \quad (6.18)$$

Now, taking $\frac{F_{ij}}{A_j} \approx \frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)$ (2.34), we obtain the *visibility discretisation error* between patches i and j:

$$\begin{aligned}
\delta_{ij}^v &= I_{ij}^c - I_{ij} \\
&\approx \frac{A_i A_j}{A_T}\Bigg(\frac{1}{N_{ij}}\Big(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log\Big(F(x_k,y_k)A_T\Big)\Big) \\
&- \frac{F_{ij}}{A_j}\log\Big(\frac{F_{ij}}{A_j}A_T\Big)\Bigg) \\
&= \frac{A_i A_j}{A_T}\Bigg(\frac{1}{N_{ij}}\Big(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log(F(x_k,y_k))\Big) \\
&- \frac{F_{ij}}{A_j}\log\Big(\frac{F_{ij}}{A_j}\Big)\Bigg) \\
&= \frac{A_i A_j}{A_T}\Bigg(\frac{1}{N_{ij}}\Big(\sum_{k=1}^{N_{ij}} F(x_k,y_k)\log(F(x_k,y_k))\Big) \\
&- \Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\Big)\log\Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k,y_k)\Big)\Bigg) \geq 0
\end{aligned} \quad (6.19)$$

where we have used the log-sum inequality (2.67). 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^v_{ij} is symmetric: $\delta^v_{ij} = \delta^v_{ji}$.

2. **Local or global lines:**
The computation of I^c_{ij} can also be done with uniformly distributed local or global lines (section 2.1.6). From (6.15) and (4.8), we obtain

$$I^c_{ij} \approx \frac{A_i F_{ij}}{A_T} \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} \log(A_T F(x_k, y_k)) \tag{6.20}$$

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

$$\delta^v_{ij} = I^c_{ij} - I_{ij} \approx \frac{A_i F_{ij}}{A_T} \Big(\frac{1}{N_{ij}} \Big(\sum_{k=1}^{N_{ij}} \log(F(x_k, y_k)) \Big) - \log \frac{F_{ij}}{A_j} \Big) \tag{6.21}$$

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

- **For radiosity**

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

$$\mathsf{I}^c_{ij} = \int_{A_i} \int_{A_j} \frac{F(x,y) B(x) B(y)}{\mathsf{A}^c_T} \log \frac{\mathsf{A}^c_T F(x,y)}{B^{in}(x) B^{in}(y)} dA_x dA_y$$

can also be done with an area-to-area sampling. Assuming constant approximations for the radiosity over the elements, we obtain for N_{ij} lines

$$\mathsf{I}^c_{ij} \approx \frac{A_i A_j B_i B_j}{\mathsf{A}_T} \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \log \Big(\frac{F(x_k, y_k) \mathsf{A}_T}{B^{in}_i B^{in}_j} \Big) \tag{6.22}$$

and from (6.7)

$$\mathsf{I}_{ij} = \frac{A_i A_j B_i B_j}{\mathsf{A}_T} \frac{F_{ij}}{A_j} \log \Big(\frac{F_{ij}}{A_j} \frac{\mathsf{A}_T}{B^{in}_i B^{in}_j} \Big) \tag{6.23}$$

Now, taking $\frac{F_{ij}}{A_j} \approx \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k)$ (2.34), we obtain the *radiosity discretisation error* between patches i and j:

$$\begin{aligned} \delta^r_{ij} &= \mathsf{I}^c_{ij} - \mathsf{I}_{ij} \\ &\approx \frac{A_i A_j B_i B_j}{\mathsf{A}_T} \Big(\frac{1}{N_{ij}} \Big(\sum_{k=1}^{N_{ij}} F(x_k, y_k) \log(F(x_k, y_k)) \Big) \\ &- \Big(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \Big) \log \Big(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \Big) \Big) \geq 0 \end{aligned} \tag{6.24}$$

Observe that

$$\delta^r_{ij} = \frac{B_i B_j A_T}{\mathsf{A}_T} \delta^v_{ij} \tag{6.25}$$

2. **Local or global lines:** The computation of I^c_{ij} can also be done with uniformly distributed local or global lines (sections 2.1.6). From (6.15) and (5.20), we obtain

$$\mathsf{I}^c_{ij} \approx \frac{A_i F_{ij} B_i B_j}{\mathsf{A}_T} \frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} \log\left(\frac{\mathsf{A}_T F(x_k, y_k)}{B_i^{in} B_j^{in}}\right) \tag{6.26}$$

Hence, we find

$$\delta^r_{ij} = \mathsf{I}^c_{ij} - \mathsf{I}_{ij} \approx \frac{A_i F_{ij} B_i B_j}{\mathsf{A}_T}\left(\frac{1}{N_{ij}}\Big(\sum_{k=1}^{N_{ij}} \log(F(x_k, y_k))\Big) - \log\frac{F_{ij}}{A_j}\right) \tag{6.27}$$

6.4 Kernel-Smoothness-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.

Unlike the first oracle proposed in section 6.2, this oracle is based on a *nearly* exact evaluation of the maximum potential gain in mutual information.

6.4.1 An Oracle Based on the Discretisation Error between Two Patches

A natural choice for a mutual-information-based oracle would be to use the radiosity discretisation error δ^r_{ij} (6.24):

$$\begin{aligned}
\delta^r_{ij} &= \mathsf{I}^c{}_{ij} - \mathsf{I}_{ij} \\
&\approx \frac{A_i A_j B_i B_j}{\mathsf{A}_T}\Big(\frac{1}{N_{ij}}\Big(\sum_{k=1}^{N_{ij}} F(x_k, y_k)\log(F(x_k, y_k))\Big) \\
&- \Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k, y_k)\Big)\log\Big(\frac{1}{N_{ij}}\sum_{k=1}^{N_{ij}} F(x_k, y_k)\Big)\Big) \geq 0
\end{aligned}$$

But this error expresses the information transfer loss in a converged scene whereas we are interested in an oracle which will be incorporated into a hierarchical radiosity algorithm. Note also that, at the beginning of the radiosity computation, most of the receiver B_i values are zero. However, a δ^r_{ij} oracle could be used with a nearly converged scene.

Thus, we take a different approach, similar to the classic smoothness-based oracles, which multiplies $\rho_i B_j$ (from the radiosity equation (2.6)) by an expression of the visibility gradient between the two patches involved. In our case, the visibility gradient is given by the discretisation error $\delta^v_{ij} = I^c_{ij} - I_{ij}$, which also represents the *variation of the radiosity kernel*.

Our oracle will be based on the following considerations:

- In the radiosity equation (2.6)

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

  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 2.1.12, such as

$$\rho_i (F^{max}_{ij} - F^{min}_{ij}) A_j B_j < \epsilon$$

  and

$$\rho_i max(F^{max}_{ij} - F^{av}_{ij}, F^{av}_{ij} - F^{min}_{ij}) A_j B_j < \epsilon$$

  , 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^v_{ij} between two patches) by $\rho_i B_j$. So, we find that the *mutual-information-based (MI-based) oracle* is given by

$$\rho_i \delta^v_{ij} B_j < \epsilon \tag{6.29}$$

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

$$\begin{aligned} &\frac{\rho_i A_i A_j B_j}{A_T} \Bigg(\frac{1}{N_{ij}} \Big(\sum_{k=1}^{N_{ij}} F(x_k, y_k) \log(F(x_k, y_k)) \Big) \\ - \quad &\Big(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \Big) \log \Big(\frac{1}{N_{ij}} \sum_{k=1}^{N_{ij}} F(x_k, y_k) \Big) \Bigg) < \epsilon \end{aligned} \tag{6.30}$$

Note that the differences between (6.24) and (6.30) are that A_T is substituted by A_T and B_i is substituted by ρ_i. 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^v_{ij} 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.

6.4.2 Empirical Results

To check the performance of the MI-based oracle, we have implemented a power-based oracle (2.48), a classic kernel-smoothness-based (KS-based) oracle (2.50) and our MI-based oracle (6.30) in the hierarchical Monte Carlo radiosity [5] method of the RenderPark[14] system (`www.renderpark.be`). It should be noted that our oracle can be used with any hierarchical radiosity method.

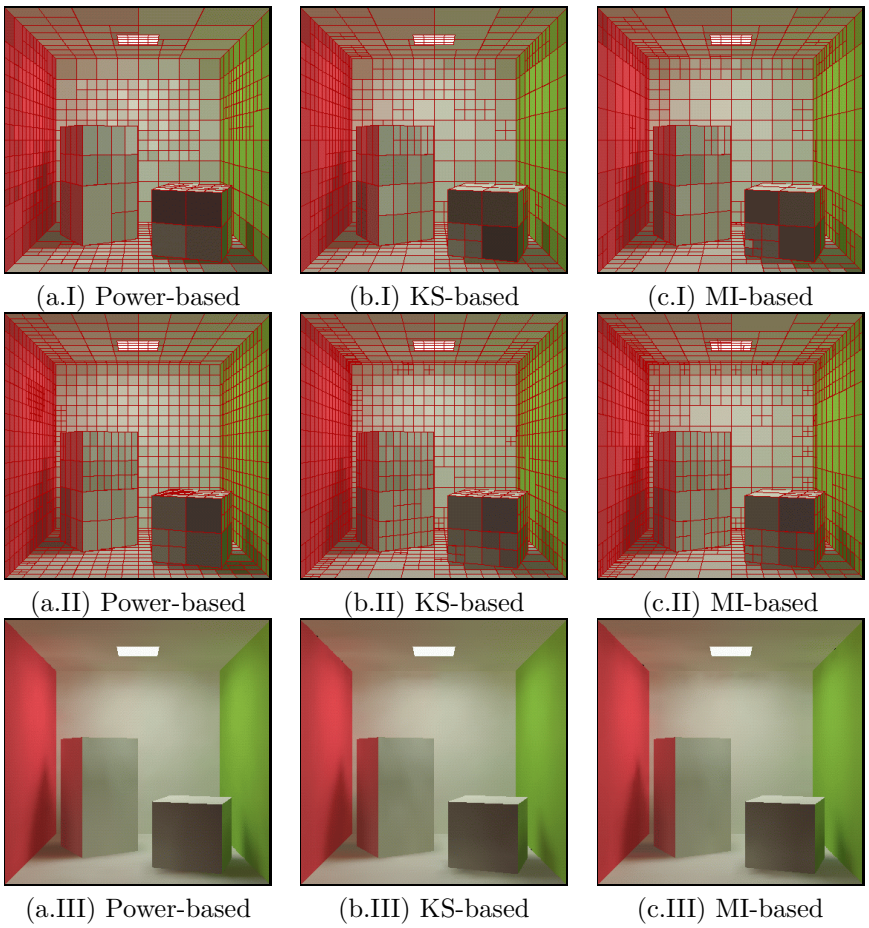
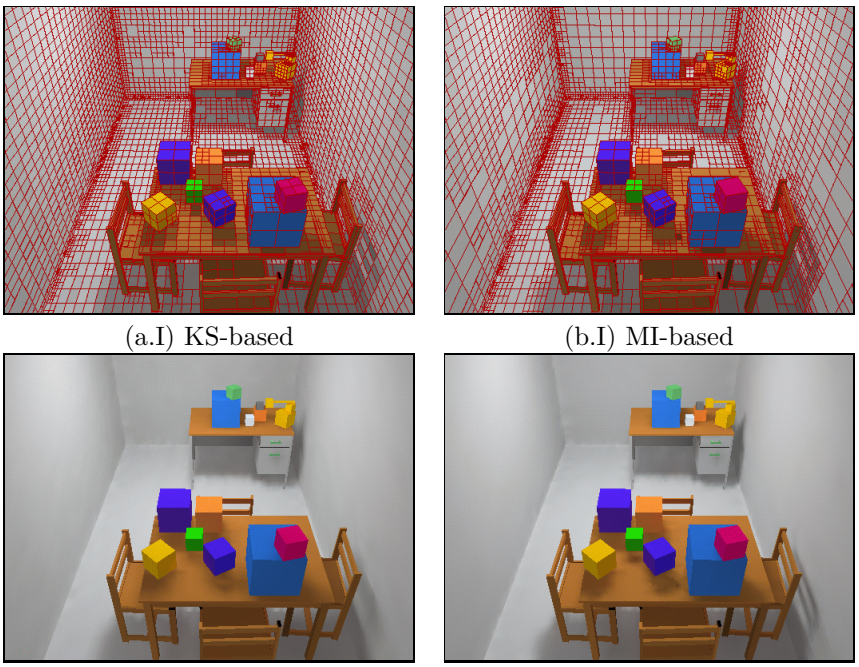


Figure 6.6: Power-based (a), KS-based (b) and MI-based (c) methods with the Cornell box scene. A coarse mesh is shown in (I) with 1051 (a.I), 1039 (b.I), and 1047 (c.I) patches, with 19472 rays for the radiosity computation. A fine mesh is shown in (II) with 1979 (a.II), 1955 (b.II), and 1995 (c.II) 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.

In our experiments, we use two scenes: the Cornell box (Figures 6.6 and 6.9) and the cube room (Figures 6.7, 6.8 and 6.10). Six different discretisations were generated for the Cornell box: three coarse (Figure 6.6I) and three finer ones (Figure 6.6II). These discretisations have been obtained from three meshing strategies based, respectively, on transported power (2.48) (Figures 6.6a.I and

6.6a.II), classic kernel smoothness (2.50) (Figures 6.6b.I and 6.6b.II), and mutual information (6.30) (Figures 6.6c.I and 6.6c.II). In a similar way, we compared our strategy with the KS-based strategy using two different views of the cube room scene (Figures 6.7 and 6.8). Both KS-based and MI-based oracles were evaluated for each discretisation decision with 10 element-to-element random lines (except in Figures 6.9 and 6.10, where only 4 rays were used). For the power-based oracle we used a cheap form factor estimate (see section 2.1.12).



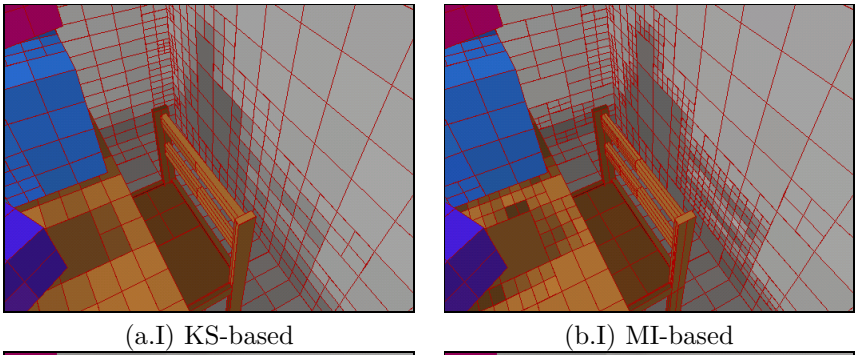
(a.I) KS-based (b.I) MI-based

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

Figure 6.7: 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.

In Figures 6.6I and 6.6II 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).

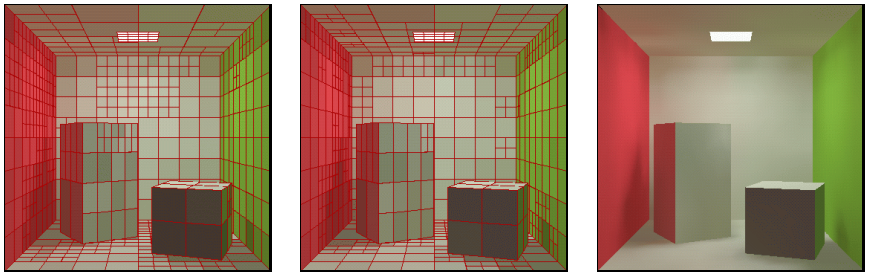
Figures 6.7 and 6.8 show the behaviour of the classic KS-based and MI-based oracle for the cube room scene. Observe the accurate representation of the shadow of the chair near the right wall (Figure 6.7b) and front wall (Figure 6.8b) obtained by the MI-based oracle. Observe also the much better discrimination in the mesh, seen for instance on the floor and walls, and how the shadows on the table are represented more accurately in Figure 6.8b.



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

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

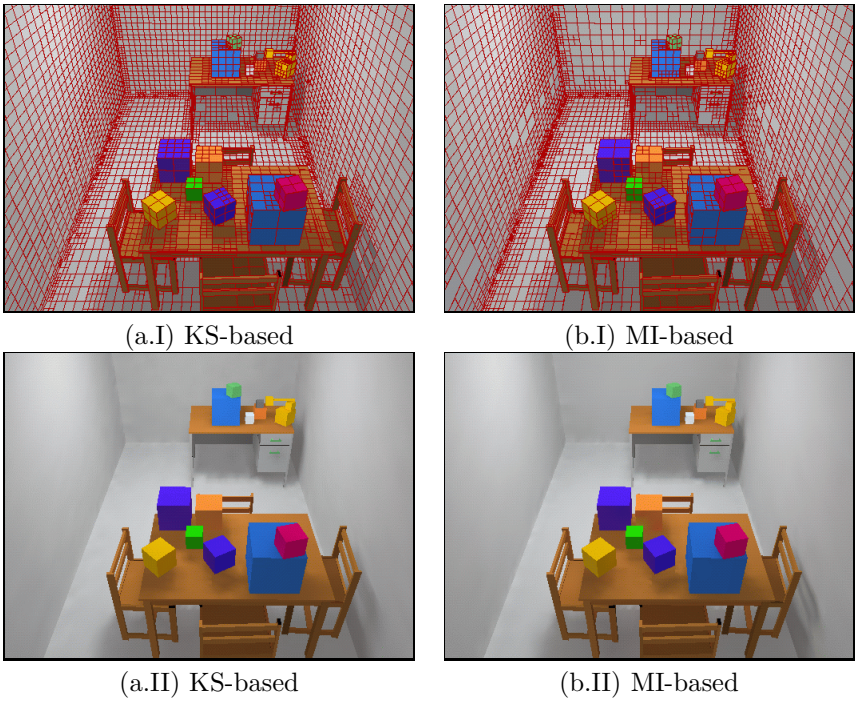
Figure 6.8: A different view of the scene shown in Figure 6.7.



(a) KS-based (b) MI-based (c) MI-based

Figure 6.9: 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).

In Figures 6.9 and 6.10, 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 6.9a (compare with Figures 6.6b.I) and the same happens in Figure 6.10a (compare with Figures 6.7a). On the other hand, the MI-based oracle maintains most of its good performance (compare Figure 6.9b with Figures 6.6c.II). See also the shadow of the chair near the right wall in Figures 6.10b and 6.7b.



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

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

Figure 6.10: 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.

In Figure 6.11 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.

6.5 Scene Discretisation in Flatland

In this section, the relationship between mutual information, entropy, and discretisation is studied in flatland. A discretisation algorithm based on the minimization of the difference $I_S^c - I_S$ will also be introduced.

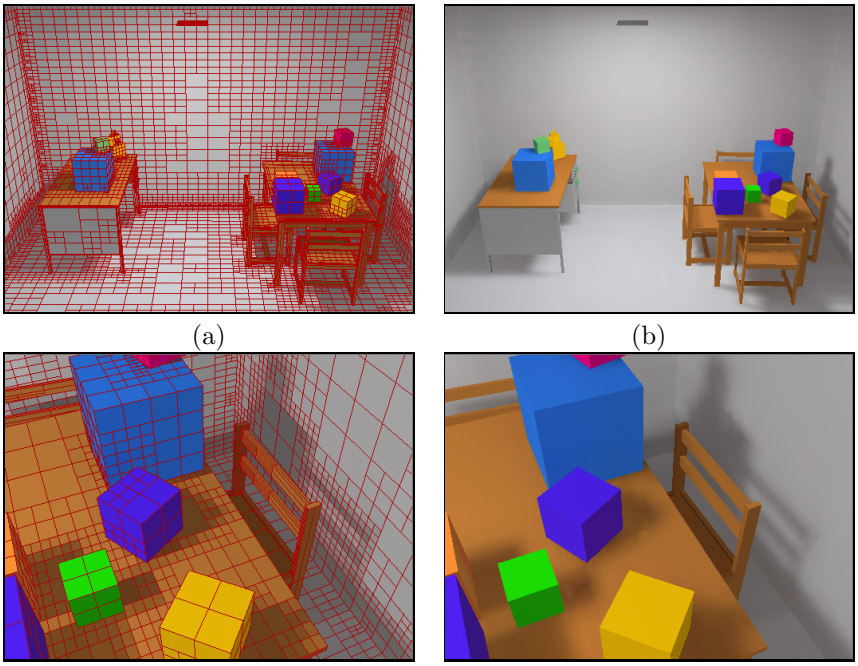


Figure 6.11: MI-based method with the scene shown in Figure 6.7. The number of patches is 18338. We cast 2684260 rays for radiosity computation and 10 rays for each oracle evaluation.

6.5.1 Mutual Information Matrix and Discretisation

In this section, a simple discretisation of an equilateral triangle is used to show the behaviour of I_S and H_S with respect to the discretisation.

Starting from a regular discretization of an equilateral triangle with six patches (Figure 6.12a), we subdivide symmetrically each patch into two parts and we analyse the behaviour of I_S and H_S with respect to the cutting point. As we can observe (Figure 6.12c) the maximum value for I_S is obtained when we cut at point P, at a distance of 15.6% from the vertex (Figure 6.12b), and H_S has the maximum value at point S. Obviously, H_P reaches the maximum (point T) with a regular discretisation.

If instead of two divisions we make three, the maximum for I_S is obtained with the discretisation shown in Figure 6.13a. Mutual information, entropy, computational error $E(MSE)$, and relative discretisation error (4.11) of the two discretisations (optimal and regular) of Figure 6.13 can be seen in Table 6.1. In the optimal case, the discretisation error is the lowest and so are the entropy and the computational error.

Similarly to section 6.3.1, we can consider that the terms

$$I_{ij} = l_i F_{ij} \log \frac{F_{ij}}{l_j} \tag{6.31}$$

correspond to a symmetric mutual information matrix ($I_{ij} = I_{ji}$), where each term represents the exchange (or transfer) of information between the patches

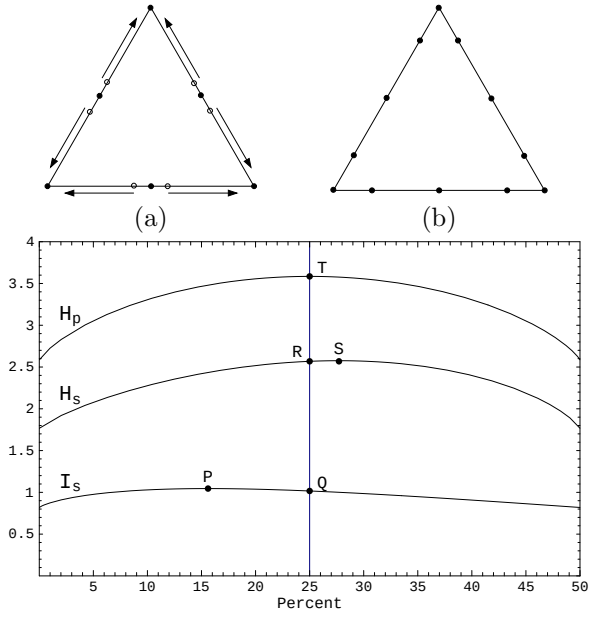


Figure 6.12: (a) Subdivision of an equilateral triangle. (b) Maximum I_S. (c) Points P and S indicate, respectively, the discretisation with maximum I_S and maximum H_S. The horizontal axis represents the distance from the cutting point to the vertex in percentage terms.

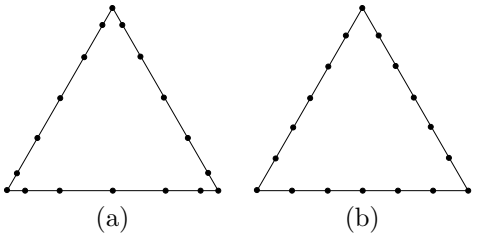


Figure 6.13: (a) Optimal and (b) regular discretisations.

scene	a	b
discretisation	optimal	regular
I_s^c	1.285	1.285
I_s	1.137	1.097
H_s	2.896	3.073
H_j	6.929	7.243
$E(MSE)$	14	15
$\overline{\delta}^v$ in %	11.52	14.63

Table 6.1: Values corresponding to the discretisations of Figure 6.13.

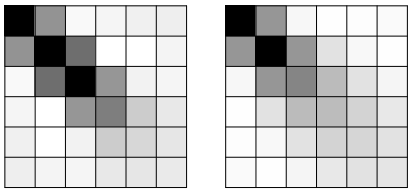
i and j. Also, the information transfer carried out by patch i is given by

$$I_i = l_i \sum_{j=1}^{n_p} F_{ij} \log \frac{F_{ij}}{l_j} \geq 0 \tag{6.32}$$

Thus, we have

$$I_S = \sum_{i=1}^{n_p} I_i = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} I_{ij} \geq 0 \tag{6.33}$$

Figures 6.14a-b (gray scaled maps where the darkest color corresponds to the highest values of I_{ij}) show the transfer of information between the patches on two sides of the triangles of Figure 6.13. Clearly, the maximum values are reached near the vertexes.



(a) (b)

Figure 6.14: (a,b) show the transfer of information between two sides of the triangles in Figures 6.13a and 6.13b respectively. The darkest colour indicates the highest value.

As we have seen in section 4.2, continuous mutual information I_S^c can be computed approximately by casting uniformly distributed global lines, and the average of the terms

$$\log(\frac{L_T cos\theta_x cos\theta_y}{2r_{xy}}) \tag{6.34}$$

represents the information transport in a scene. From this point of view, we can define, in analogous form, the terms

$$I_{ij}^c = I_{ji}^c \simeq \frac{1}{2N} \sum_{k=1}^{N_{ij}} \log(\frac{L_T \cos\theta_{x_k} \cos\theta_{y_k}}{2d(x_k, y_k)}) \quad \forall i \neq j \tag{6.35}$$

$$I_{ii}^c \approx \frac{1}{N} \sum_{k=1}^{N_{ii}} \log(\frac{L_T \cos\theta_{x_k} \cos\theta_{y_k}}{2d(x_k, y_k)}) \tag{6.36}$$

where N is the total number of pairs of points considered, i and j are two patches of a scene, and N_{ij} is the number of pairs of points which connect the patches i and j. Thus I_{ij}^c expresses the continuous information transfer between patches i and j.

We can also define the contribution of a patch i to the global complexity I_S^c as

$$I_i^c = \sum_{j=1}^{n_p} I_{ij}^c \tag{6.37}$$

This can be interpreted as the total information transferred by patch i.

Thus,

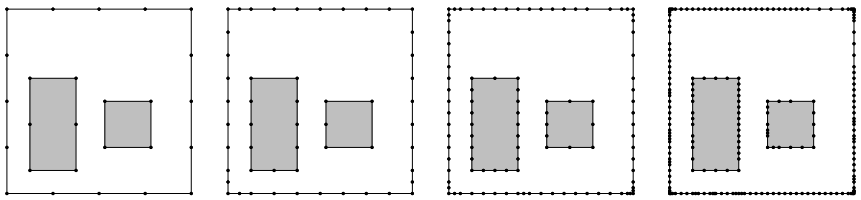
$$I_S^c = \sum_{i=1}^{n_p} I_i^c = \sum_{i=1}^{n_p} \sum_{j=1}^{n_p} I_{ij}^c \tag{6.38}$$

6.5.2 An Oracle Based on the Discretisation Error of a Patch

Finally, we introduce a refinement criterion based on the difference between I_i^c and I_i for each patch. This difference represents the discretisation error (DE) of patch i. Thus, as the objective is to reduce this as quickly as possible in order to obtain the maximum information transfer with the minimum number of patches, the algorithm proceeds by subdividing, into two equal parts, the patch that has the maximum loss, i.e., the maximum potential gain. This algorithm ends with a given discretisation accuracy or a given number of patches.

In the following experiments, this refinement criterion (DE-based oracle) is contrasted with another which divides, at each step, the patch with maximum information transfer I_i ("power-based" oracle).

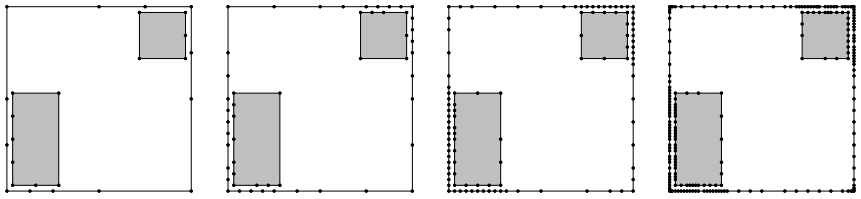
In Figures 6.15 (scene I) and 6.16 (scene II) we can see that the discretisation obtained with the DE-based oracle is finer at the corners and narrow spaces, and in general at the regions with a higher transfer of information. On the other hand, the discretisation obtained with the power-based oracle (Figures 6.17 and 6.18) is more uniform.



(I.a) (I.b) (I.c) (I.d)

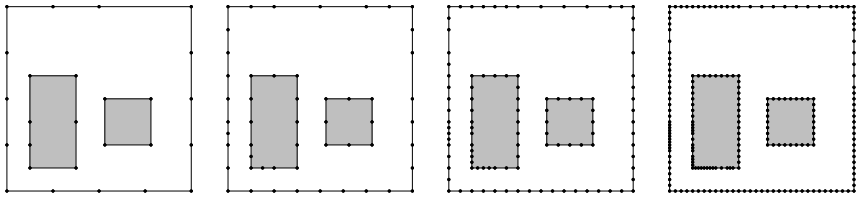
Figure 6.15: Four different discretisations for scene I obtained with the DE-based oracle: (a) 26, (b) 52, (c) 104 and (d) 208 patches respectively.

In Table 6.2, we compare three different discretisations of 208 patches (regular, power-based, and DE-based) for the scenes in Figures 6.15 and 6.17 (scene I) and 6.16 and 6.18 (scene II), and in Figure 6.19 we show the evolution of I_S and H_J with respect to the number of patches for the scene II. We can see that the best results are obtained with the DE-based oracle: the discretisation error is the lowest and so is the entropy. This fact is also important as entropy is directly related to computational error. In conclusion, the DE-based oracle shows good behaviour by reducing both errors.



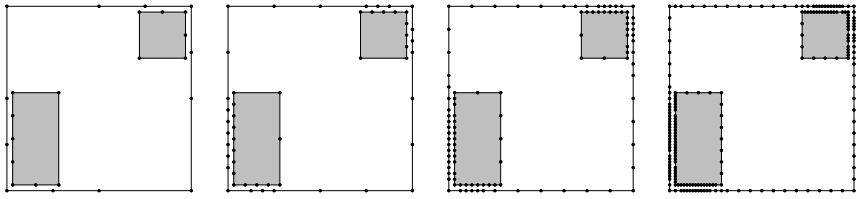
(II.a) (II.b) (II.c) (II.d)

Figure 6.16: Four different discretisations for scene II obtained with the DE-based oracle: (a) 26, (b) 52, (c) 104 and (d) 208 patches respectively.



(I.a) (I.b) (I.c) (I.d)

Figure 6.17: Four different discretisations for scene I obtained with the power-based oracle: (a) 26, (b) 52, (c) 104 and (d) 208 patches respectively.

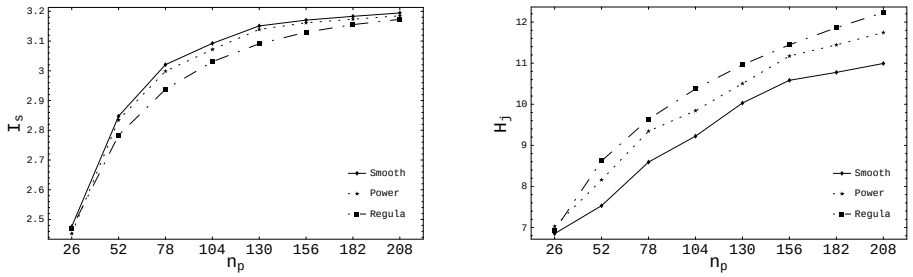


(II.a) (II.b) (II.c) (II.d)

Figure 6.18: Four different discretisations for scene II obtained with the power-based oracle: (a) 26, (b) 52, (c) 104 and (d) 208 patches respectively.

scene	*algorithm*	I_S^c	I_S	H_S	H_J	$E(MSE)$	$\overline{\delta}^o$ (%)
I	regular	2.519	2.492	5.208	12.908	68	1.07
	power-based		2.490	5.220	12.930	68	1.15
	DE-based		2.496	4.715	11.926	67	0.90
II	regular	3.245	3.173	4.530	12.233	63	2.21
	power-based		3.185	4.280	11.745	62	1.84
	DE-based		3.195	3.899	10.993	61	1.54

Table 6.2: Values obtained for three different discretisations (regular, power-based, DE-based), with 208 patches, corresponding to scenes I and II.



(a) (b)

Figure 6.19: Evolution of (a) I_S and (b) H_J of scene II with respect to the number of patches and three different discretizations: regular, power-based (power) and DE-based (smooth).

6.6 Summary

The assumption that, between different discretisations of the same scene, the most precise will be the one that has the minimum loss of information transfer is fundamental in dealing with the accuracy of the discretisation and in introducing some criteria of refinement applied to visibility and radiosity. In consequence, in this chapter, refinement has been driven by the criterion of maximizing the discrete mutual information.

Discrete and continuous mutual information can be seen as matrices where each element represents the exchange (or transfer) of information between two patches. The sum of a row of this matrix represents the information transferred by a patch.

Our experiments have illustrated the behaviour of mutual information when one or more patches are refined. First, we theoretically analyzed an oracle for hierarchical radiosity refinement based on the increase in mutual information when a patch is subdivided has been theoretically analyzed. The drawbacks of this method were overcome by an oracle based on the discretisation error.

We presented two different oracles based on the discretisation error:

- The first one is based on the discretisation error between two patches (for hierarchical radiosity)

  The discretisation error between two patches is used to decide the necessity of refinement. This oracle was compared with a kernel-smoothness-based oracle and a power-based oracle.

- The second one is based on the discretisation error on a patch (for visibility in flatland)

  As the objective was to reduce the discretisation error as quickly as possible in order to obtain the maximum information transfer with the minimum number of patches, the algorithm proceeds by subdividing (into two equal parts) the patch that has the maximum loss of information transfer, i.e., the maximum potential information gain. The results obtained with this oracle were contrasted with another one which is based simply on the value of the information transferred by each patch.

Chapter 7

Conclusions and Future Work

We present here the conclusions and the main contributions of this thesis, and also some directions for future research.

7.1 Conclusions

In this thesis, we have studied the *complexity of a scene* and new *criteria of refinement* in the visibility and radiosity settings from an information-theory point of view. Below, we review the principal concepts developed and results obtained in this dissertation:

- In chapter 3, a new way of looking at the visibility of a scene was developed from an information-theory point of view. Hence, a scene was interpreted as a communication channel.

  Beginning with this assumption, the most essential discrete information-theory measures, like entropy and mutual information, were defined in the visibility setting, and their properties analysed. In short, we show that scene entropy represents both randomness and unpredictability in a scene, and mutual information expresses both correlation and dependence between the different parts of a scene. From different experiments, we demonstrated that scene entropy and mutual information show a complementary behaviour.

  Entropy was also reinterpreted from the perspective of the local and global lines. We found that the uncertainty of a global line segment is at least twice the uncertainty of a local line. A close relationship between scene entropy and the variance of the form factor estimators was demonstrated: the bigger the entropy, the bigger the error in form factor computation. This means that, for a given error, we need to cast more lines for a scene with more entropy.

- In chapter 4, we analyzed the notion of scene complexity. After that, continuous scene mutual information, which expresses with maximum accuracy the information transfer in a scene, was proposed as the scene vis-

ibility complexity, and discrete scene mutual information as the visibility complexity of a discretised scene.

The continuous mutual information integral was solved by Monte Carlo integration, and the computation was done by casting uniformly distributed global or local lines. We showed that two fundamental properties are fulfilled in a scene:

- If any patch is divided into several subpatches, discrete mutual information increases or remains the same.
- Continuous scene visibility mutual information is the least upper bound to the discrete scene visibility mutual information.

We also saw that scene complexity, which measures the degree of correlation or dependence in a scene, also quantifies the difficulty in obtaining a precise discretisation, i.e., the higher the scene complexity the more difficult it is to obtain an accurate discretisation and probably more refinements will be necessary. The difference between continuous and discrete mutual information was interpreted as a global discretisation error and represents the loss of information transfer due to the discretisation.

Finally, the main reasons for the growth in complexity were described and a tentative scene classification was proposed in flatland.

- In chapter 5, our previous results were extended to the radiosity setting. To achieve this aim, we used an analog of the form factor matrix for the radiosity setting: the null variance probability transition matrix. Thus, both entropy and complexity of a scene with illumination, were defined by using a different pair of discrete and continuous Markov chains. A set of experiments demonstrated the feasibility of this approach.

  Similarly to the visibility case, continuous scene radiosity mutual information was interpreted as the difficulty in obtaining an accurate meshing. The mutual information integral was also solved by Monte Carlo integration, taking an approximate solution for the radiosity. A general proposition for visibility, radiosity, and importance enabled us to calculate the gain in mutual information resulting from refinement.

- In chapter 6, a mutual-information-based refinement criteria for hierarchical radiosity was introduced. As we saw in the previous chapters, from an information-theory point of view, the most accurate discretisation corresponds to the one with maximum discrete mutual information. Thus, our objective was to reduce the information transfer loss, i.e. the discretisation error.

  First, we demonstrated with different experiments that the objective of maximizing the discrete mutual information can lead us towards an optimal discretisation. After that, we analyzed, theoretically, a possible oracle based on the increase in mutual information when a patch is refined.

  Second, the discretisation error between two patches or elements was defined in the visibility and radiosity settings as the difference between continuous and discrete element-to-element mutual information. Then, we introduced a new smoothness-based refinement oracle for hierarchical radiosity based on this discretisation error. This oracle was then compared

with two classic refinement oracles based on transported power and kernel smoothness using a hierarchical Monte Carlo radiosity implementation. Experiments suggested that it is an improvement on previous oracles in that, at equal cost, it produces a better discretisation and also needs less visibility computations for a similar image quality.

Third, an oracle based on the information transfer loss of a patch has been introduced for the visibility in flatland. The algorithm proceeds by subdividing the patch that has the maximum loss of information transfer, i.e., the maximum potential information gain.

7.2 Main Contributions

The principal contributions of this thesis are described below. We also indicate the papers related to each contribution.

- The interpretation of a scene as an information channel, i.e., a system which contains and transfers information. With this assumption, the scene can be studied using information theory tools. [17, 21]
- Scene complexity definitions. Entropy and mutual information capture different but complementary aspects of the complexity of a scene. On the one hand, entropy measures the degree of randomness of a random walk in a scene and also the uncertainty in the form factor computation (using uniformly distributed random lines). On the other hand, mutual information quantifies the degree of structure or correlation of a scene, i.e., just how complicated the interrelations are between the different parts of a scene, and it is interpreted as the difficulty of obtaining a precise discretisation. In general, these measures express the difficulty in dealing with a scene. [23, 21]
- Mutual information, which also measures the information transfer in a scene, is used to obtain new refinement criteria. Our refinement oracles are based on the difference between continuous and discrete mutual information. This difference expresses the loss of information transfer due to the dicretisation and can be interpreted as the benefit to be gained by refining. Thus, among different discretisations of the same scene, the most accurate is the one with the highest discrete mutual information, i.e., the minimum loss of information transfer. [22, 25, 24]
- The work started in this thesis has paved the way for additional research and applications in areas such as pixel supersampling [65, 69, 68] and viewpoint selection [89, 91, 92].

7.3 Future Research

- The procedures presented in this thesis can be extended to the interior points of a scene and to the points in the environment. To achieve this objective, we can define, respectively, the entropy and mutual information fields and the entropy and mutual information densities. Point measures can be applied to

  - Studying the complexity of an interior region of a scene [66] and defining the animation complexity [67].

    The complexity of animation can be defined from the complexity at the interior points of a scene. Using a physical analogy, we can compute the work done in a particular animation.

  - Pixel adaptive supersampling [65, 69, 68].

    Two different information-theory contrast measures can be used in the recursive refinement. The first one is based on entropy, which measures the homogeneity of a pixel. And the second one is based on the difference between continuous and discrete mutual information field, which measures the distance to the ideal representation.

  - Defining a new refinement criterion for hierarchical radiosity.

    A new oracle based on the smoothness of the received radiosity can be introduced from the mutual information density at a point. The difference between continuous and discrete mutual information density will give us a measure of the discretisation error at a point.

- The concept of entropy can also be applied to the selection of good viewpoints with applications in areas such as image-based rendering, molecular visualization, and automatic exploration [90, 89, 91, 92].

  A method which automatically computes the goodness of a viewpoint in a scene has been introduced. This method obtains good views that can be used for image-based modeling, scene understanding, etc.

- Other aspects to be analyzed are listed below:

  - The definition of scene complexity can also be extended to non diffuse environments. A non-diffuse scene can be modeled by a 2-order Markov chain. Therefore, a statistical complexity measure like excess entropy can be used to quantify the complexity of a non-diffuse scene.

  - The behaviour of the long-range mutual information, i.e. the correlation after n rebounds of a ray, can be inspected in a scene. This can clearly show us what parts of a scene need more work.

  - Mutual information could also be used to simplify the mesh. Thus, a set of patches can be clustered if the loss of information transfer is sufficiently small.

  - A theoretical relationship between scene entropy and the variance of the form factor estimates will be investigated further.

7.4 Publications

- Publications that support the contents of this thesis:

  1. E. del Acebo, M. Feixas and M. Sbert. Form Factors and Information Theory. *Proceedings of the 3rd International Conference on Computer Graphics and Artificial Intelligence (3IA'98)*, Limoges, France, 1998.

2. M. Feixas, E. del Acebo and M. Sbert. Entropy of Scene Visibility. *Proceedings of Winter School on Computer Graphics and CAD Systems'99 (WSCG'99)*, Plzen, Czech Republic, 1999.
3. M. Feixas, E. del Acebo, Ph. Bekaert and M. Sbert. An Information Theory Framework for the Analysis of Scene Complexity. *Computer Graphics Forum (Proceedings of Eurographics'99)*, 18(3): 95–106, Milan, Italy, 1999. This paper obtained an honourable mention in EG'99 conference.
4. M. Feixas, E. del Acebo, Ph. Bekaert and M. Sbert. Information Theory Tools for Scene Discretisation. *Rendering Techniques'99 (Proceedings of the 10th Eurographics Workshop on Rendering)*, pages 95–106, Granada, Spain, 1999.
5. M. Feixas, J. Rigau and M. Sbert. Scene Visibility Complexity and Discretisation in Flatland. Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 2002. Technical Report, IIiA–02–12–RR.
6. M. Feixas, J.Rigau, Ph.Bekaert and M. Sbert. Information–Theoretic Oracle Based on Kernel Smoothness for Hierarchical Radiosity. *Short Presentations of Eurographics'02*, pages 325–333, Saarbrücken, Germany, 2002.

- Additional publications:

1. J. Rigau, M. Feixas and M. Sbert. Visibility Complexity of a Region in Flatland. *Short Presentations of Eurographics'00*, p. 159–163, Interlaken, Switzerland, 2000.
2. J. Rigau, M. Feixas and M. Sbert. Information Theory Point Measures in a Scene. Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 2000. Technical Report, IIiA–00–08–RR.
3. J. Rigau, M. Feixas and M. Sbert. Visibility Complexity of Animation in Flatland. *Proceedings of Winter School on Computer Graphics and CAD Systems'01*, Plzen, Czech Republic, 2001.
4. J. Rigau, M. Feixas and M. Sbert. View-Dependent Information Theory Measures for Pixel Sampling and Scene Discretisation in Flatland. *Proceedings of Spring Conference on Computer Graphics'01 (SCCG'01)*, p. 231–238, Budmerice, Slovak Republic, 2001.
5. J. Rigau, M. Feixas and M. Sbert. New Contrast Measures for Pixel Supersampling. *Advances in Modeling, Animation and Rendering (Proceedings of CGI'02)*, p. 439–451, Bradford, UK, 2002.
6. J. Rigau, M. Feixas and M. Sbert. Entropy-Based Adaptive Supersampling. *Poster Papers Proceedings of the 13th Eurographics Workshop on Rendering*, p. 63–70, Pisa, Italy, 2002.
7. P.P. Vázquez, M. Feixas, M. Sbert and W. Heidrich. Viewpoint Selection using Viewpoint Entropy. *Proceedings of Vision, Modeling, and Visualization 2001*, Stuttgart, Germany, 2001.

8. P.P. Vázquez, M. Feixas, M. Sbert and W. Heidrich. Image-Based Modeling Using Viewpoint Entropy. *Advances in Modeling, Animation and Rendering (Proceedings of CGI'02)*, p. 267–279, Bradford, UK, 2002.

9. P.P. Vázquez, M. Feixas, M. Sbert and A.Llobet. A New Tool for Obtaining Good Views for Molecules. *Proceedings of VisSym'02 (Eurographics-IEEE TCVG Symposium on Visualization)*, Barcelona, Spain, 2002.

10. F. Castro, M. Feixas and M. Sbert. Fuzzy Random Walk. *Advances in Modeling, Animation and Rendering (Proceedings of CGI'02)*, p. 389–396, Bradford, UK, 2002.

11. M. Sbert, M. Feixas, J. Rigau, F. Castro and P.P. Vázquez. Applications of Information Theory to Computer Graphics. *Proceedings of 5th International Conference on Computer Graphics and Artificial Intelligence (3IA'02)*, p. 21–36, Limoges, France, 2002.

Bibliography

[1] C. B. Anteneodo and A. R. Plastino. Some features of the López-Ruiz–Mancini–Calbet (lmc) statistical measure of complexity. *Physics Letters A*, 223:348–354, 1996.

[2] T. Arbel and F. P. Ferrie. Viewpoint selection by navigation through entropy maps. In *Proceedings of 7th International Conference on Computer Vision*, pages 248–254, September 1999. Held in Corfu, Greece.

[3] R. Badii and A. Politi. *Complexity. Hierarchical Structures and Scaling in Physics.* Cambridge University Press, 1997.

[4] P. Bekaert. *Hierarchical and Stochastic Algorithms for Radiosity.* PhD thesis, Department of Computer Science, Leuven, Belgium, December 1999.

[5] P. Bekaert, L. Neumann, A. Neumann, M. Sbert, and Y. D. Willems. Hierarchical Monte Carlo radiosity. In G. Drettakis and N. Max, editors, *Rendering Techniques'98 (Proceedings of the 9th Eurographics Workshop on Rendering)*, pages 259–268, New York (NY), USA, June 1998. Springer-Verlag Vienna-New York. Held in Vienna, Austria.

[6] P. Bekaert and Y. D. Willems. Error control for radiosity. In X. Pueyo and P. Schröder, editors, *Rendering Techniques'96 (Proceedings of the 7th Eurographics Workshop on Rendering)*, pages 153–164, New York (NY), USA, June 1996. Springer-Verlag Vienna-New York. Held in Vienna, Austria.

[7] C. H. Bennett. *Logical Depth and Physical Complexity*, pages 227–257. Oxford University Press, Oxford, UK, 1988.

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

[9] A. T. Campbell III. *Modeling Global Difusse Illumination for Image Synthesis.* PhD thesis, Department of Computer Science, Austin (TX), USA, 1991.

[10] F. Cazals and M. Sbert. Some integral geometry tools to estimate the complexity of 3d scenes. Research Report 3204, INRIA, July 1997.

[11] M. F. Cohen and D. P. Greenberg. The hemi-cube: A radiosity solution for complex environments. *Computer Graphics (Proceedings of SIGGRAPH'85)*, 19(3):31–40, July 1985. Held in San Francisco (CA), USA.

[12] M. F. Cohen and J. R. Wallace. *Radiosity and Realistic Image Synthesis*. Academic Press Professional, Boston (MA), USA, 1993.

[13] R. Coleman. *Stochastic Processes*. George Allen & Unwin Ltd., London, UK, 1974.

[14] Computer Graphics Research Group. *RenderPark: A Photorealistic Rendering Tool*. Katholieke Universiteit Leuven, Leuven, Belgium, November 2000.

[15] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley Series in Telecommunications, 1991.

[16] J. P. Crutchfield and N. Packard. Symbolic dynamics of noisy chaos. *Physica 7D*, pages 201–223, 1983.

[17] E. del Acebo, M. Feixas, and M. Sbert. Form factors and information theory. In *The 3rd International Conference on Computer Graphics and Artificial Intelligence (3IA'98)*, Limoges, France, 1998.

[18] N. Ebrahimi, E. Maasoumi, and E. S. Soofi. Ordering univariate distributions by entropy and variance. *Journal of Econometrics*, (90):317–336, 1999.

[19] M. Feda and W. Purgathofer. Progressive ray refinement for Monte Carlo radiosity. In M. F. Cohen, C. Puech, and F. X. Sillion, editors, *Eurographics Rendering Workshop 1993*, pages 15–26. Eurographics Association, June 1993. Held in Paris, France.

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

[21] M. Feixas, E. del Acebo, P. Bekaert, and M. Sbert. An information theory framework for the analysis of scene complexity. *Computer Graphics Forum (Proceedings of Eurographics'99)*, 18(3):95–106, September 1999.

[22] M. Feixas, E. del Acebo, P. Bekaert, and M. Sbert. Information theory tools for scene discretization. In D. Lischinski and G. W. Larson, editors, *Rendering Techniques'99 (Proceedings of the 10th Eurographics Workshop on Rendering)*, pages 95–106, New York (NY), USA, June 1999. Springer-Verlag Vienna-New York. Held in Granada, Spain.

[23] M. Feixas, E. del Acebo, and M. Sbert. Entropy of scene visibility. In *Proceedings of Winter School on Computer Graphics and CAD Systems'99*, pages 25–34, Plzen, Czech Republic, February 1999.

[24] M. Feixas, J. Rigau, P. Bekaert, and M. Sbert. Information-theoretic oracle based on kernel smoothness for hierarchical radiosity. In *Short Presentations (Eurographics'02)*, pages 325–333, September 2002. Held in Saarbrücken, Germany.

[25] M. Feixas, J. Rigau, and M. Sbert. Scene visibility complexity and discretization in flatland. Research Report IIiA–02–12–RR, Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 2002.

[26] M. Feixas and M. Sbert. 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.

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

[28] D. P. Feldman and J. P. Crutchfield. Discovering noncritical organization: Statistical mechanical, information theoretic and computational views of patterns in simple one-dimensional spin systems. Working Paper 98–04–026, Santa Fe Institute, Santa Fe (NM), USA, April 1998.

[29] D. P. Feldman and J. P. Crutchfield. Statistical measures of complexity: Why? *Physics Letters A*, 238(4/5):244–252, 1998.

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

[31] M. Gell-Mann and J. P. Crutchfield. Computation in physical and biological systems: Measures of complexity. `http://www.santafe.edu/sfi/research/focus/compPhysics`, 2001.

[32] M. Gell-Mann and S. Lloyd. Information measures, effective complexity, and total information. *Complexity*, 2(1):44–52, 1996.

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

[34] C. M. Goral, K. E. Torrance, D. P. Greenberg, and B. Battaile. Modelling the interaction of light between diffuse surfaces. *Computer Graphics (Proceedings of SIGGRAPH'84)*, 18(3):213–222, July 1984. Held in Minneapolis (MN), USA.

[35] S. J. Gortler, P. Schröder, M. F. Cohen, and P. Hanrahan. Wavelet radiosity. *Computer Graphics (Proceedings of SIGGRAPH'93)*, 27:221–230, August 1993. Held in Anaheim (CA), USA.

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

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

[38] J. M. Hammersley and D. C. Handscomb. *The Monte Carlo Method*. Cambridge University Press, 1964.

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

[40] P. S. Heckbert. Simulating global illumination using adaptive meshing. Technical Report UCB/CSD 91/636, Computer Science Division (EECS), University of California, Berkeley (CA), USA, 1991.

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

[42] J. E. Hopcroft and J. D. Ullmann. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, Reading (MA), USA, 1979.

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

[44] M. H. Kalos and P. A. Whitlock. *The Monte Carlo Method*. John Wiley & Sons, 1986.

[45] A. N. Kolmogorov. On the Shannon theory of information transmission in the case of continuous signals. *IRE Transactions on Information Theory*, 2:102–108, 1956.

[46] M. Li and P. Vitányi. *An Introduction to Kolmogorov Complexity and Its Applications*. Graduate Texts in Computer Science. Springer-Verlag Vienna-New York, New York (NY), USA, 1997.

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

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

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

[50] D. Lischinski, F. Tampieri, and D. P. Greenberg. Combining hierarchical radiosity and discontinuity meshing. *Computer Graphics (Proceedings of SIGGRAPH'93)*, 27:199–208, August 1993. Held in Anaheim (CA), USA.

[51] S. Lloyd and H. Pagels. Complexity as thermodymanic depth. *Annals of Physics*, 188:186–213, 1988.

[52] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, New York (NY), USA, 1995.

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

[54] L. Neumann, M. Feda, and W. Purgathofer. A new stochastic radiosity method for highly complex scenes. In S. Haas, S. Mueller, G. Sakas, and P. Shirley, editors, *Photorealistic Rendering Techniques (Proceedings of the 5th Eurographics Workshop on Rendering)*, pages 201–213, New York (NY), USA, June 1994. Springer-Verlag Vienna-New York. Held in Darmstad, Germany.

[55] L. Neumann, A. Neumann, and P. Bekaert. Radiosity with well distributed ray sets. *Computer Graphics Forum (Proceedings of Eurographics'97)*, 16(3):261–270, 1997. Held in Budapest, Hungary.

[56] T. Nishita and E. Nakame. Continuous tone representation of 3-D objects taking account of shadows and interreflection. *Computer Graphics (Proceedings of SIGGRAPH'85)*, 19(3):23–30, July 1985. Held in San Francisco (CA), USA.

[57] R. Orti. *Radiosité Dynamique 2D et Complexe de Visibilité*. PhD thesis, INRIA, Grenoble, France, 1997.

[58] A. Papoulis. *Probability, Random Variables, and Stochastic Processes*. McGraw-Hill, New York (NY), USA, 2 edition, 1984.

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

[60] S. N. Pattanaik and S. P. Mudur. Computation of global illumination by Monte Carlo simulation of the particle model of light. In *Proceedings of the 3th Eurographics Workshop on Rendering*, pages 71–83, May 1992. Held in Bristol, UK.

[61] S. N. Pattanaik and S. P. Mudur. The potential equation and importance in illumination computations. *Computer Graphics Forum*, 12(2):131–136, 1993.

[62] S. N. Pattanaik and S. P. Mudur. Adjoint equations and random walks for illumination computation. *ACM Transactions on Computer Graphics*, 14(1):77–102, January 1995.

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

[64] J. P. Pluim. *Mutual Information Based Registration of Medical Images*. PhD thesis, Image Sciences Institute, Utrecht, The Netherlands, 2001.

[65] J. Rigau, M. Feixas, P. Bekaert, and M. Sbert. View-dependent information theory measures for pixel sampling and scene discretization in flatland. In *Proceedings of Spring Conference on Computer Graphics'01*, pages 173–180, Los Alamitos (CA), USA, April 2001. IEEE Computer Society. Held in Budmerice, Slovak Republic.

[66] J. Rigau, M. Feixas, and M. Sbert. Visibility complexity of animation in flatland. Research Report IIiA–00–05–RR, Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 2000.

[67] J. Rigau, M. Feixas, and M. Sbert. Visibility complexity of animation in flatland. In *Proceedings of Winter School on Computer Graphics and CAD Systems'01*, volume 2, pages 352–359, Plzen, Czech Republic, February 2001.

[68] J. Rigau, M. Feixas, and M. Sbert. Entropy-based adaptive supersampling. In P. Debevec and S. Gibson, editors, *The 13th Eurographics Workshop on Rendering, Poster Papers Proceedings*, pages 63–70, Pisa, Italy, June 2002. National Research Council. Held in Pisa, Italy.

[69] J. Rigau, M. Feixas, and M. Sbert. *New Contrast Measures for Pixel Supersampling*, pages 439–451. Springer-Verlag London Limited, London, UK, July 2002. Proceedings of CGI'02, Bradford, UK.

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

[71] L. Santaló. *Integral Geometry and Geometric Probability*. Addison-Wesley, Reading (MA), USA, 1976.

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

[73] M. Sbert. *The Use of Global Random Directions to Compute Radiosity. Global Monte Carlo Methods*. PhD thesis, Departament de Llenguatges i Sistemes Informàtics, Barcelona, Spain, November 1996.

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

[75] M. Sbert, A. Brusi, R. F. 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, Girona, Spain, 1998.

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

[77] C. R. Shalizi and J. P. Crutchfield. Computational mechanics: Pattern and prediction, structure and simplicity. Working Paper 99–07–044, Santa Fe Insitute, Santa Fe (NM), USA, July 1999.

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

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

[80] P. Shirley. A ray tracing method for illumination calculation in diffuse–specular scenes. In *Proceedings of Graphics Interface'90*, pages 205–212, May 1990. Held in Halifax, Canada.

[81] R. Siegel and J. R. Howell. *Thermal Radiation Heat Transfer*. Hemisphere Publishing Corporation, New York (NY), USA, 3 edition, 1992.

[82] F. X. Sillion and C. Puech. *Radiosity and Global Illumination*. Morgan Kaufmann Publishers, San Francisco (CA), USA, 1994.

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

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

[85] C. Studholme. *Measures of 3D Medical Image Alignment*. PhD thesis, Computational Imaging Science Group, London, UK, August 1997.

[86] P. Szépfalusy and G. Györgyi. Entropy decay as a measure of stochasticity in chaotic systems. *Physical Review A*, 33(4):2852, 1986.

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

[88] J. C. van der Lubbe. *Information Theory*. Cambridge University Press, Cambridge, UK, 1997.

[89] P. P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. Viewpoint selection using viewpoint entropy. In T. Ertl, B. Girod, G. Greiner, H. Niemann, and H.-P. Seidel, editors, *Proceedings of Vision, Modeling, and Visualization 2001*, pages 273–280, Stuttgart, Germany, November 2001. Held in Stuttgart, Germany.

[90] P. P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. Viewpoint selection using viewpoint entropy. Research Report IIiA–01–04–RR, Institut d'Informàtica i Aplicacions, Universitat de Girona, Girona, Spain, 2001.

[91] P. P. Vázquez, M. Feixas, M. Sbert, and W. Heidrich. *Image-Based Modeling Using Viewpoint Entropy*, pages 267–279. Springer-Verlag London Limited, London, UK, July 2002. Proceedings of CGI'02, Bradford, UK.

[92] P. P. Vázquez, M. Feixas, M. Sbert, and A. Llobet. A new tool for obtaining good views for molecules. In D. Ebert, P. Brunet, and I. Navazo, editors, *Proceedings of VisSym'02 (Eurographics-IEEE TCVG Symposium on Visualization)*, pages 0–1, May 2002. Held in Barcelona, Spain.

[93] S. Verdú. Fifty years of shannon theory. *IEEE Transactions on Information Theory*, 44(6):2057–2078, October 1998.

[94] P. A. Viola. *Alignment by Maximization of Mutual Information*. PhD thesis, MIT Artificial Intelligence Laboratory (TR 1548), Massachusetts (MA), USA, 1995.

[95] W. M. Wells III, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1), 1996.

[96] W. H. Zurek. Algorithmic randomness and physical entropy. *Physical Review D*, 40(8):4731–4751, 1989.

[97] W. H. Zurek. *Complexity, Entropy and the Physics of Information.* Addison-Wesley, Reading (MA), USA, 1990.