Application of Quasi-Monte Carlo Sampling to the Multi Path Method for Radiosity

Francesc Castro and Mateu Sbert

Institut d'Informàtica i Aplicacions. Universitat de Girona, Lluis Santaló s/n, 17071 Girona, Spain
castro@ima.udg.es and mateu@ima.udg.es

Abstract. The multi path method is a Monte Carlo technique that solves the radiosity problem, i.e. the illumination in a scene with diffuse (also called lambertian) surfaces. This technique uses random global lines for the transport of energy, contrary to the classic Monte Carlo techniques, in which the lines used are local to the surface where they exited from. The multi path technique borrows results from Integral Geometry to predict the correct transfer of energy, and can be shown to be a random walk method, in which a geometric path corresponds to several logical paths.

We will study in this paper the application of quasi-Monte Carlo sequences to the random sampling of the global lines. Important improvements in the efficiency of the multi path method for certain sequences will be demonstrated. Alternative ways of generating global lines will also be studied in the context of quasi-Monte Carlo.

1 Introduction

The multi path method for radiosity is described in [SPNP96]. It is a member of a family of methods called by different authors global Monte Carlo, global Radiosity or transillumination methods [SPP95,SKFNC97,Neu95]. They use random global lines (or directions) to simulate the transport of energy, and in this way to compute the radiosity.

The global lines are independent of the surfaces or patches in the scene, in contraposition to local lines, used in the classic methods, which are dependent on the patches they are cast from. The global lines can take advantage of all the intersections with the scene. However a problem arises here. The random sequences used to generate the lines have a high discrepancy [Nie92]. This is inherent to all Monte Carlo methods. We have tried to avoid this problem by using quasi-Monte Carlo (qMC) low discrepancy sampling, that is, low discrepancy sequences (lds).

We will start with a brief review of the previous work. Then, in Sect. 3, we will consider different ways to produce random global lines. In Sect. 4, qMC sampling applied to multi path method will be discussed. Next we will present some results, and end with the conclusions and future work.

2 Previous Work

2.1 The Radiosity System of Equations

The *radiosity equation* [CW93] solves for the illumination in a diffuse environment. It can be written in the form

$$B(x) = E(x) + R(x) \int_{\mathcal{S}} B(x')V(x,x')\frac{\cos\theta\cos\theta'}{\pi r^2}d\sigma(x') \tag{1}$$

where

- $B(x)$ is the radiosity
- $E(x)$ is the emittance
- $R(x)$ is the reflectance
- $\mathcal{S}$ is the set of surfaces that form the environment
- x, x' are points on surfaces of the environment
- $d\sigma(x')$ is an area differential at point x'
- r is the distance between x and x'
- $V(x,x')$ is a visibility function equal to 1 if x and x' are mutually visible and 0 otherwise. The computation of this function is the main source of cost solving the equation
- θ, θ' are the angles the normals at x, x' form with the line joining them
- $V(x,x')\frac{\cos\theta\cos\theta'}{\pi r^2}$ is the differential form factor between x and x'.

To solve the radiosity equation we can use a finite element approach, discretising the scene in patches and considering radiosities, emissivities and reflectances constant over the patches. For every patch we have the following equation, obtained from the discretisation of equation (1) [CW93].

$$B_i = E_i + R_i \sum_j F_{ij} B_j \tag{2}$$

where

- B_i is the radiosity of patch i, the flux of energy per unit area that leaves patch i (W/m^2)
- E_i is the emittance of patch i, the flux of energy per unit area that emits patch i (W/m^2)
- R_i is the reflectance of patch i, the fraction of energy that is reflected by patch i
- F_{ij} is the form factor from patch i to patch j, that is, the fraction of energy that leaving patch i lands directly on patch j, defined [CW93] as the integral

$$F_{ij} = \frac{1}{A_i} \int_{A_i} \int_{A_j} \frac{\cos\theta \cos\theta'}{\pi r^2} V(x, x') d\sigma(x) d\sigma(x') \tag{3}$$

Considering the solid angle ω and the relation $d\omega = \frac{\cos\theta' d\sigma(x')}{r^2}$ we can rewrite the integral [CW93] as

$$F_{ij} = \frac{1}{\pi A_i} \int_{\Omega} \int_{A_i} V_{ij}(\omega, x) cos\theta d\sigma(x) d\omega \tag{4}$$

where we integrate on the hemisphere Ω and on the patch i (Fig. 1). $V_{ij}(\omega, x)$ is 1 if patch j is visible from point x in direction ω.

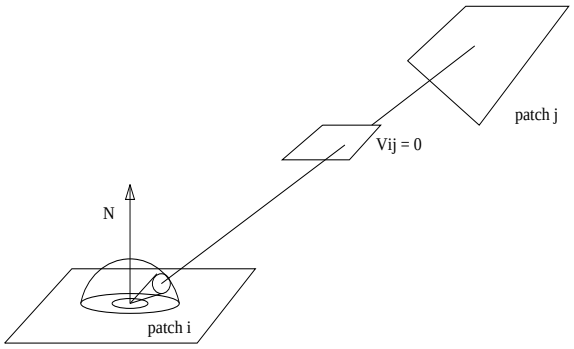


Fig. 1. *Monte Carlo integration of the form factors*

Thus we have a system of equations (2), with as many equations as patches in the scene. Our target will be to solve this system. The most expensive part is the computation of form factors. form factors only depend on the geometry of the scene.

Basically we have two ways to solve the radiosity system of equations:

- Computing explicitly the form factors, either the full matrix or a row at a time, to solve by numerical methods the system of equations.
- Simulating the exchange of energy between the patches, using i.e. a random walk. In this case the transition probabilities used will be the (unknown) form factors.

In the second approach, the form factors (4) are simulated using an importance sampling scheme with pdf $f(\omega, x) = \frac{cos\theta}{\pi A_i}$ [Sbe97b]. Thus, lines are cast from each patch, according to this density. This approach is called local (Fig. 2a).

An alternative scheme is to use a global density of lines in the sense of Integral Geometry, that is, homogeneous and isotropic [Sbe93,San76]. It can be shown [Sbe97b] that this global density submits on each patch the

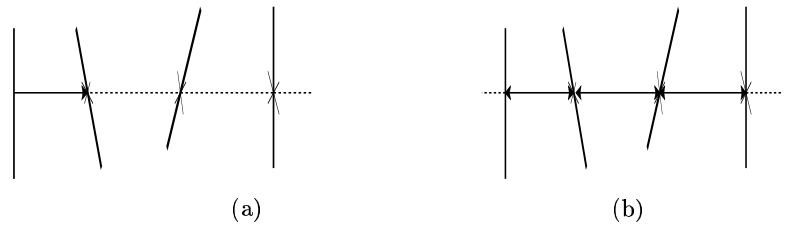


Fig. 2. *(a) Local approach: lines are casted from each patch, and energy is transfered only to nearest intersected patch (b) Global approach: lines are independent of the patches, and all intersections are used to transfer energy bidirectionally*

above considered importance sampling density. An important advantage of this approach in front of the local one is that each line segment can be used to transfer energy bidirectionally, (Fig. 2b). This is used by the multi path algorithm, explained in next section.

2.2 The Multi Path Algorithm

The multi path method exposed in [SPNP96] shows that it is possible to simulate a random walk using a global density of lines. This global density submits on each patch a local density according to the importance sampling seen in Sect. 2.1, using all line intersections to transfer energy bidirectionally.

We will briefly review the algorithm. Each global line will simulate the exchange of flux of energy between several pairs of patches. In this way, every global line contributes to several geometric paths (Fig. 3a).

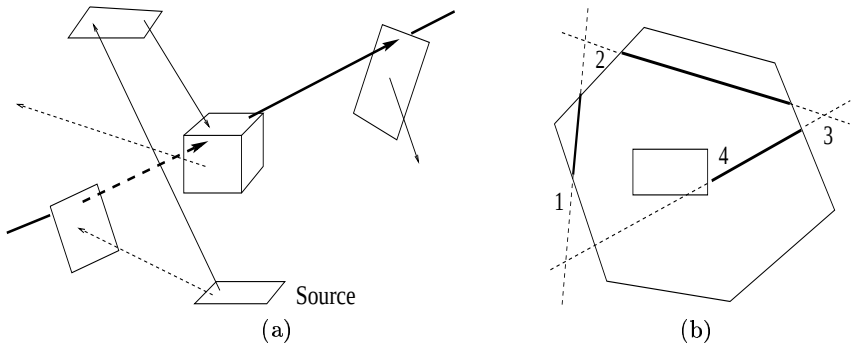


Fig. 3. *(a) multi path method. A global line (the thick one) simulates two geometric paths, indicated with the continuous stroke and the dashed stroke (b) multi path method. A geometric path can contribute to the emission of power from several patches. In the figure, geometric path 1-2-3-4 simulates paths 1-2-3-4, 2-3-4 and 3-4. This is similar to the covering paths [Rub81]*

Global lines are intersected with the scene. For each line the intersections are sorted by distance in an intersection list. Each patch (if not a source) has with it two quantities. One records the flux of energy accumulated, the other one is the unshot flux of energy. For every pair of patches along the intersection list, they exchange their unshot energies, decreased by the respective reflectances, and moreover they add to their accumulated energy the unshot energy of the other patch (also decreased by the reflectance). If a patch is a source, we keep also a third quantity, the emitted energy per line exiting the source. Thus, if one of the patches of the pair is an emitter patch, we must add this emitted energy per line to the unshot energy. This "emitted per line" energy is computed in the following way: given the number of the lines we are going to cast, we compute for any light source patch beforehand the forecast number of lines passing through it. This number is, for a planar patch, proportional to the area of the patch [San76]. The division of the total source energy by this number gives the predicted energy of one line.

There are three main advantages in the multi path method. The two first ones are the use of all intersections of a line and the *bidirectionality* of energy transfer. The other main advantage is that each path is used to transport different logical paths (see Fig. 3b), as with the covering paths [Rub81].

A drawback of the multi path method, due to its global nature, is that in first stages the distribution of power is only possible from light sources, and so most of the lines cast in these first stages (the lines that do not cross any light source) are wasted. To avoid this behaviour and gain in efficiency a preprocess, called *first shot*, is done [SPNP96,Sbe97a], in which the power is cast from the source patches in a local way. After that, the patches that have received some power will be the new sources instead of the original ones. Note that after this preprocess the power to be emitted is more widely distributed, decreasing the initial waste in global lines and thus becoming the multi path method much more efficient.

A comparison of the multi path method against the classic (local) methods can be found in [Sbe97b] and [SPNP96].

2.3 Random sampling of a Global Uniform Density of Lines

There are several methods to generate a uniform density of lines in a scene. The guideline is that the line density has to correspond to the form factor density, the integrand of (3) or (4) without the visibility function. Now we will comment some of them [Sbe97b,San76,Sol78]:

- **Pairs of random points on a bounding sphere.**
  In [San76] it is shown that a density of global lines intersecting a convex body is given by $\frac{\cos\theta\cos\theta'}{r^2}d\sigma d\sigma'$, where θ, θ' are the angles of the intersecting line with the normals in the intersecting points, $d\sigma, d\sigma'$ are the area differentials in the same points and r is the length of the chord. If the convex body is a sphere (see Fig. 4a), the density becomes simply

(save a constant factor) $d\sigma d\sigma'$. That is, taking pairs of uniform random points on the sphere surface we obtain a global uniform density of lines.

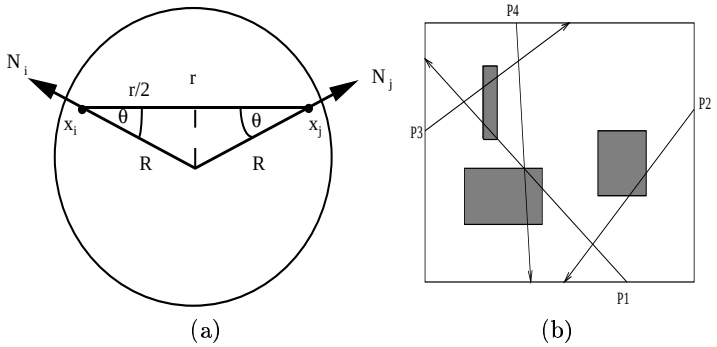


Fig. 4. *(a) Random sampling from pairs of points on a sphere (b) Random sampling using lines from the walls in a convex bounding box*

- **Lines from the walls of a convex bounding box**
  We can transform the density in a) into $\cos\theta d\omega$ (see 2.1). This new expression means that taking a uniform random point on the surface of the convex bounding box and a cosinus weighted uniformly distributed direction (Fig. 4b) we obtain the same global uniform density of lines as in a). Since this result is valid for any convex bounding box, it is useful to use the bounding box of the scene (if convex) to generate the lines.
  An advantage of casting the lines from the walls in front of using a bounding sphere is that no lines are wasted, because all the lines intersect the scene.
- **Maximum circle.**
  We sample a uniform random point on the surface of the bounding sphere (in fact, this is the same as sampling a uniform random direction). Then we take the circle orthogonal to this direction that contains the centre of the sphere (that is, a maximum circle). Finally, we sample a uniform random point on this circle. With this point and the direction, we have the global line (Fig. 5a). Note that this is equivalent to selecting a tangent plane (thus a point in the sphere) and a point in the projection of the sphere onto the plane.
- **Tangent planes: bundles of parallel lines**
  A last way to get random lines is using tangent planes. We must sample tangent planes to the sphere. To do this, we sample a uniform random point on the surface of the sphere, and then we construct the tangent plane at this point. For every plane we cast bundles of parallel lines orthogonal to the plane (Fig. 5b). Note that here the main point is the use, for each sampled direction, of bundles of parallel lines instead of a

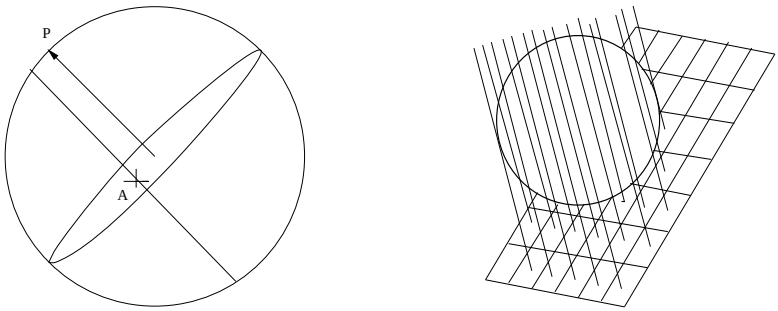


Fig. 5. *(a) Random sampling using maximum circle (b) Random sampling using tangent planes*

single line like in the previous case. To avoid lines always passing by the center of the scene, the plane is randomly jittered.
This technique offers several possibilities. An important point to consider is the balance between the number of directions and the number of lines per direction. On the other hand, the intersection of the lines with the scene can be accelerated in some ways, for instance by applying z-buffer techniques.

- **Non uniform densities.**
To end this section we present a couple of generation techniques that, although they seem to be uniform, they are not. The first one is using pairs of uniform random points from the surface of a *non spherical* convex bounding volume (Fig. 6a), thus it amounts to take a density proportional to $d\sigma d\sigma'$. But this density is incorrect in general because the factor $\frac{\cos\theta\cos\theta'}{r^2}$ in the uniform density only becomes constant for the sphere, as seen in the first point. Remember here that in the second point we took a point and a direction instead of two points.
The second one is using two parallel planes that surround the scene. Thus, if we sample one uniform random point on each plane, the lines that we get do not produce uniform distribution. The correct density should be proportional to $\frac{\cos\theta\cos\theta'}{r^2}d\sigma d\sigma'$, the one taken is proportional to $d\sigma d\sigma'$. The ratio is thus proportional to $\frac{\cos\theta\cos\theta'}{r^2}$. Now, from Fig. 6b, $\cos\theta = \cos\theta' = \frac{d}{r}$, and the ratio is proportional to $\frac{1}{r^4}$. This means that for twice the distance we cast 2^4 more lines than necessary. In other words, much more lines are cast in oblicuous directions that in orthogonal ones.

2.4 Quasi-Monte Carlo low discrepancy sampling

The main idea of using quasi-Monte Carlo low discrepancy sampling is to reduce the discrepancy of the sets of points. Intuitively, we can see the discrep-

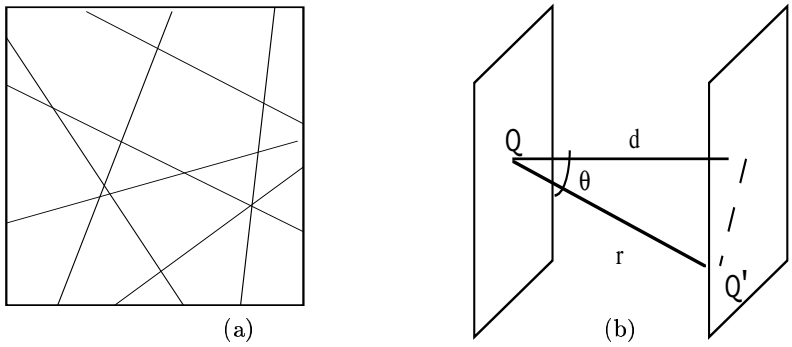


Fig. 6. *(a) Sampling lines from pairs of points on a bounding box we do not get a uniform distribution (b) Parallel planes*

ancy as a measure of how far is a set of points from the uniform distribution. A more formal and extensive definition of discrepancy can be seen in [Nie92]. If we look at Fig. 7 we can see on the left a set of 1000 2D points generated with a simple Monte Carlo generator, while on the right side the points have been generated using a quasi-Monte Carlo generator. We can easily see that the points on the right are closer to the uniform distribution that the ones on the left side. It seems that the points on the right have been generated "trying to fill empty spaces". In this case we usually talk of low discrepancy sequences (lds) of points.

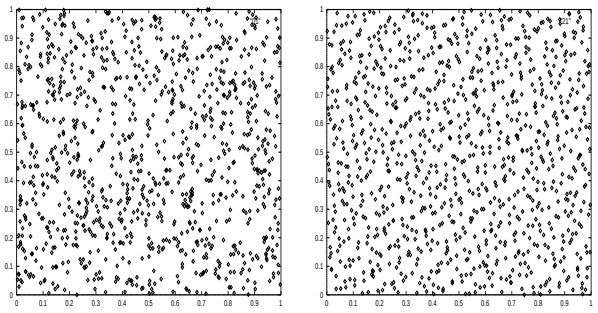


Fig. 7. *(a) Monte Carlo random sampling (1000 points) (b) Quasi-Monte Carlo low discrepancy sampling (1000 2-dimensional Halton points, using as basis 2 and 3)*

The qMC low discrepancy sampling produces sequences of k-dimensional points that in fact are obtained in a deterministic (that is, non-random) way. There exists a lot of different qMC sequences: Halton, Sobol, Niederreiter, Hammersley and others. Some of them are described with detail in [Nie92].

On the other hand, these sequences have been applied to the Monte Carlo methods for radiosity by several authors, obtaining in general satisfactory results. Now we review some of them:

[Kel96b] applies qMC techniques to the computation of the radiosities by using a quasi-Random Walk, obtaining a noticeable improvement with Halton sequences. This quasi-Random Walk uses an equal absorption probability for all patches, thus all patches are of the same length. In [NNB97] a new algorithm is presented that uses sets of rays (well distributed ray sets) constructed from Halton numbers and importance sampling. This algorithm is suitable for complex scenes, in which gains of half an order of magnitude are obtained. In [SKFNC97] Hammersley and scrambled Hammersley sequences were used in combination with the transillumination method to solve the radiosity problem.

On the other hand, we can see in [Kel96a] an application of lds to the computation of form factors, getting an speed-up of more than half an order of magnitude. Here, Halton and Hammersley sequences are applied, obtaining similar discrepancies with both. A different approach is presented in [Kel97]. The scene is not discretized in patches. This method is obtained directly from the radiance equation and applies jittered low discrepancy sampling.

Next we will study how the different qMC sequences applied to the global line generation can improve the multi path method.

3 Application of quasi-Monte Carlo Generators to the multi path Method

In the Sect. 2.3, we have studied different ways to simulate the global uniform density of lines that we need in the multi path method. The four presented ways are indeed equivalent, they all produce *the same density of lines.* But the interesting point is that they are equivalent *only if we use Monte Carlo random sampling.* If we apply quasi-Monte Carlo sequences, the results could be different. So we have applied some qMC sequences to the different techniques, and we have compared the results obtained with each one.

Basically we have applied Halton sequences, as defined in [Pre94]. Halton sequences use the representation of the numbers in basis b_i, where b_i is a prime number. In fact, they are sequences of k-tuples, where k is the dimension. Since each line needs 4 random values to be generated, $k = 4$. Thus, to get a sequence of 4-tuples, we make each component a Halton sequence with a different prime base. We have used the first 4 primes, namely 2, 3, 5 and 7.

Hammersley and Sobol sequences have also been used. In general, these sequences have not improved the results, but, as we will see later, in some cases have produced a very good performance.

4 Results

4.1 Mean Square Error

We have used mean square error (MSE) to establish a measure of the goodness of every estimation. The MSE is given by the next formula:

$$MSE = \frac{\sum_i A_i * (\hat{B}_i - B_i)^2}{\sum_i A_i} \tag{5}$$

where A_i is the area of patch i, $\hat{B}_i$ is the estimate value for its radiosity and B_i is the exact value. We obtain these reference values by running a multi path algorithm in which we cast tens of millions of lines. We have used as scene a cubical room with a table, four chairs around the table and a cube on it, and finally a light source near the ceiling. From now on, we will refer to this scene as "4CHAIRS".

We have compared a series of executions using Monte Carlo random sampling with five different qMC techniques: Four of them correspond to the application of Halton sequences, using 2,3,5 and 7 as basis, with the four techniques seen. In one case, sampling 2 points on the surface of the sphere, we have also obtained very good behaviour using Sobol sequences.

4.2 Asymptotical Behaviour of the Error

Table 1. *Asymptotical behaviour of the MSE. Scene 4CHAIRS*

METHOD	*SLOPE IN 4CHAIRS*
MONTE CARLO	-1.00
2 POINTS ON THE SPHERE (SOBOL)	-1.12
LOCAL FROM THE WALLS (HALTON)	-1.04
MAXIMUM CIRCLE (HALTON)	-1.02
TANGENT PLANES (HALTON)	-0.96

As is well known, in Monte Carlo the expected value of the mean square error decreases as $\frac{1}{N}$. This means that, if we take logarithms, the graph MSE vs. number of lines must be linear with slope -1. This is its asympthotical behaviour. Then, if we consider quasi-Monte Carlo low discrepancy sampling, we can expect a better behaviour, namely, a slope minor than -1 [SKFNC97]. We have studied the asymptotical behaviour of MSE in the scene 4CHAIRS, and the results are showed in Table 1 (they have been calculated by means of linear regression). There we have a result of -1 (as expected) in Monte Carlo, and, in general, values slightly lower than -1 in qMC. In Fig. 8 we have the log-log graph, in which we can observe the lower MSE and the slightly steeper slope (except in tangent planes technique) using qMC sequences. This means that there is not a clear gain in the asymptotical behaviour using qMC but a smoother convergence is obtained.

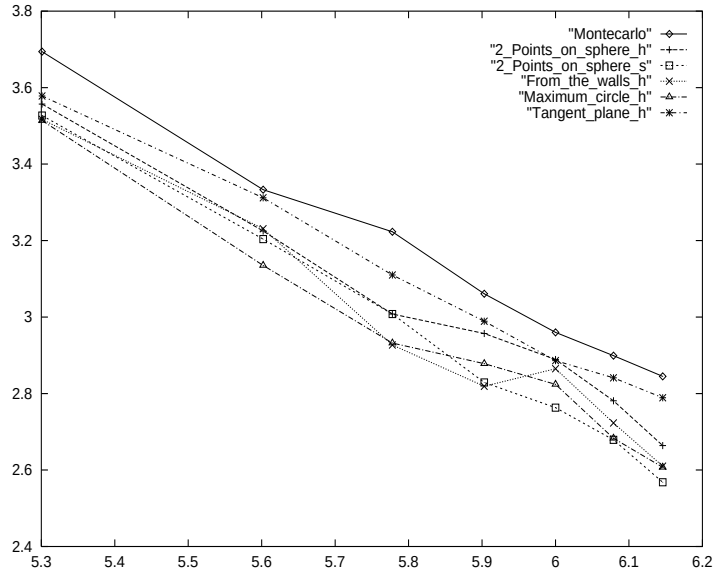


Fig. 8. *Scene 4CHAIRS. Log-log graph. MSE (in vertical axis) vs. number of lines*

It is important to remark that, in each technique, we have choosen the qMC sequence that has produced better results (between Halton and Sobol sequences).

We have also studied the asymptotical behaviour in a more complex scene, called OFFICE, that has about 36000 patches. In this case we compare Monte Carlo random sampling with several Halton low discrepancy sampling. To get more accurate results, several executions have been done and the average of the errors has been computed.

We can observe the log-log graph (Fig. 9) in which we have plotted log(execution time) versus log(MSE). The important point is that in the right side of the graph, the error obtained with the three qMC techniques (corresponding to the methods a), b) and c) seen in 2.3) is clearly lower than the error obtained using Monte Carlo sequences. As previously commented in the scene 4CHAIRS, the slope obtained using qMC is only slightly steeper, but a smoother convergence is observed.

If we compare our results with the ones obtained in [Kel96a], we find a lesser speed-up in ours. It can be due to the totally different nature of the method in [Kel96a]. It is also possible that there are some correlation problems produced by the fact that, in the multi path method, a global line transports several interreflection orders.

4.3 Images

We can see some of the obtained images in Fig. 10. Here we have generated the global lines from pairs of points on the surface of the sphere. It is quite

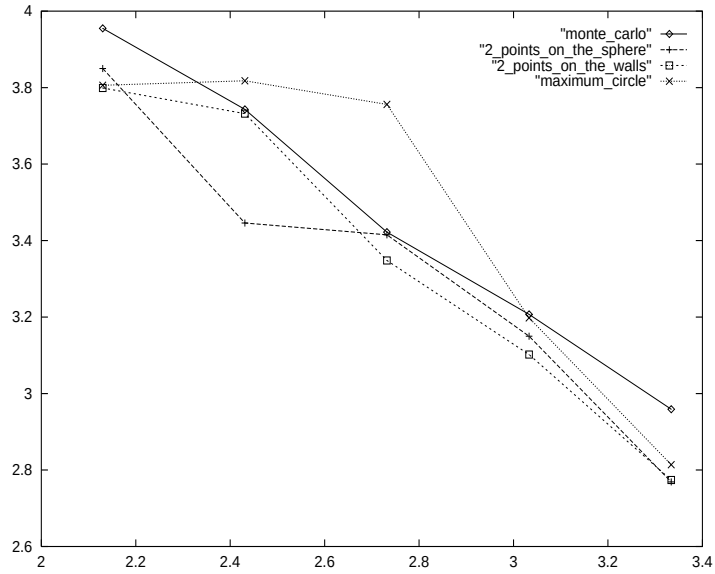


Fig. 9. *Scene OFFICE. Log-log graph. MSE (in vertical axis) vs. time (seconds) obtained in a SUN UltraSparc*

clear that, using the same number of rays, the obtained images are better using quasi-Monte Carlo. In particular, we can observe that the noise effect is much more reduced in quasi-Monte Carlo.

5 Conclusions

The multi path algorithm [SPNP96] is a Monte Carlo method to compute the radiosities of the patches in a scene. This algorithm uses uniform densities of lines, that deal with the transport of energy between the patches or polygons in the scene. In this paper we have reviewed some methods to produce this uniform density of lines. We have applied to them the quasi-Monte Carlo low discrepancy sampling. We have seen that qMC gives a lower error than Monte Carlo, that is, it needs less lines to get the same accuracy. In general, no important differences have been found between the different qMC sequences that we have tested. We have also studied the asymptotical behaviour, and concluded that the rate of convergence using qMC is not much faster than in MC, but in qMC a smoother convergence is obtained.

6 Acknowledgements

Many thanks to Ignacio Martín, Narcís Coll, Toni Sellarès and Roel Martínez for their useful help, and thanks also to Philippe Bekaert for his advice. The scene OFFICE has been designed by Greg Ward. This project has been

funded in part with a Spanish-Hungarian Proyecto de Cooperación conjunta from the Spanish Ministerio de Asuntos Exteriores and grant number TIC 98-0586-C03-02 from the Spanish Governement.

References

[CW93] M. Cohen and J. Wallace. *Radiosity and Realistic Image Synthesis.* Academic Press Professional, 1993.

[Kel96a] A. Keller. The Fast Calculation of Form Factors Using Low Discrepancy Sequences. *Proc. of SCCG96 Budmerice*, 1996.

[Kel96b] A. Keller. Quasi-Monte Carlo Radiosity. *Proceedings of Eurographics Workshop on Rendering*, pages 102–111, 1996.

[Kel97] A. Keller. Instant Radiosity. *Computer Graphics Proceedings, Siggraph'97*, pages 49–56, 1997.

[Neu95] L. Neumann. Monte Carlo Radiosity. *Computing, 55*, pages 23–42, 1995.

[Nie92] H. Niederreiter. Random Number Generation and Quasi-Monte Carlo Methods, nsf-cbms. 1992.

[NNB97] L. Neumann, A. Neumann, and P. Bekaert. Radiosity with Well Distributed Ray Sets. *Proceedings of Eurographics 97*, pages 261–269, 1997.

[Pre94] W.H. Press. Numerical Recipes in C. *Cambridge University Press*, 1994.

[Rub81] Reuven Y. Rubinstein. Simulation and the Monte Carlo Method. 1981.

[San76] L. Santaló. Integral Geometry and Geometric Probability. 1976.

[Sbe93] M. Sbert. An Integral Geometry Based Method for Fast Form-Factor Computation. *Computer Graphics Forum (proc. Eurographics'93)*, 12, N.3:409–420, 1993.

[Sbe97a] M. Sbert. Error and Complexity of Random Walk Monte Carlo Radiosity. *IEEE Transactions on Visualization and Computer Graphics, 3(1)*, 1997.

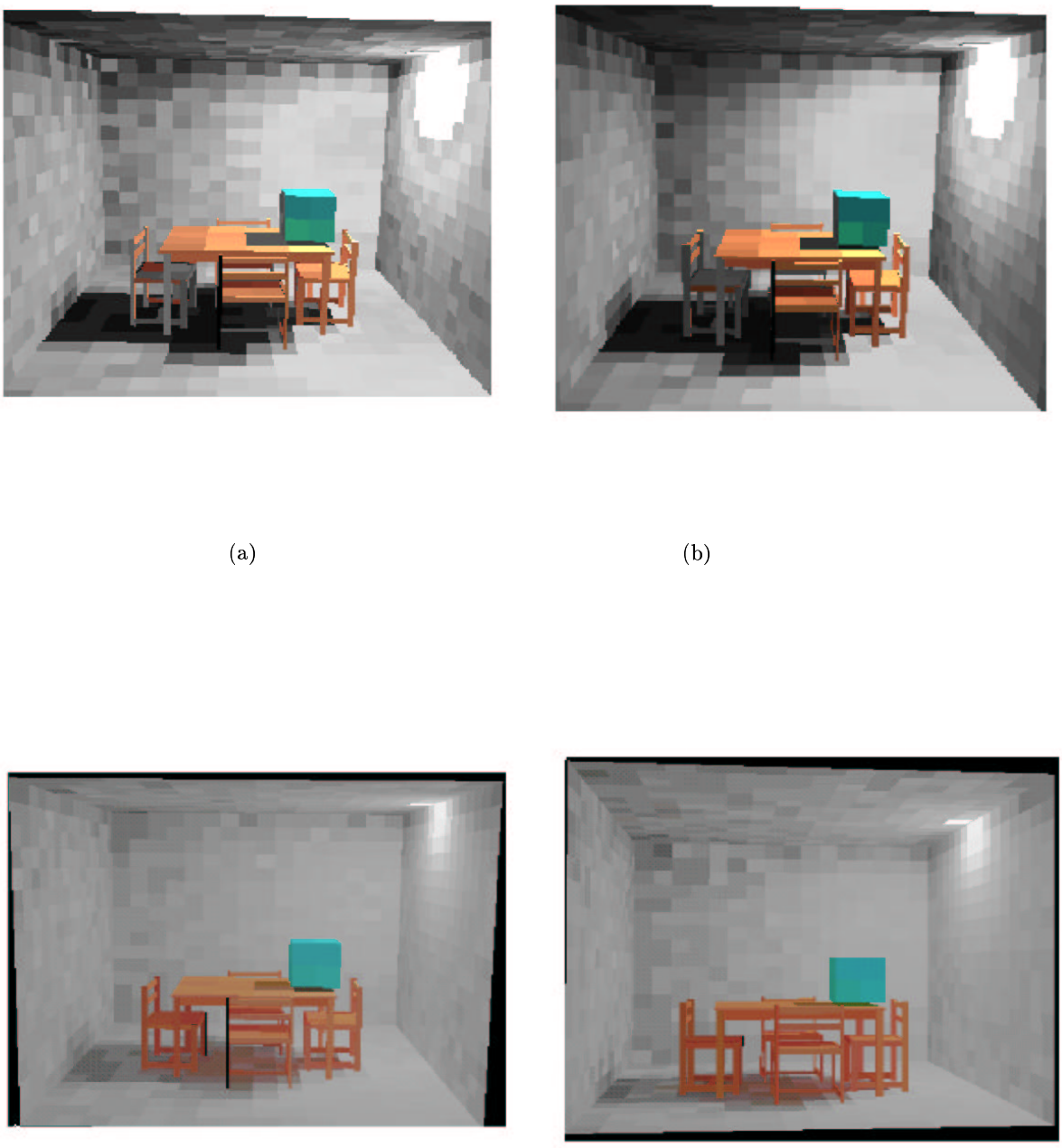
[Sbe97b] M. Sbert. The Use of Global Random Directions to Compute Radiosity. Global Monte Carlo Methods. *Ph.D. thesis. Universitat Politècnica de Catalunya, Barcelona*, 1997.

[SKFNC97] L. Szirmay-Kalos, T. Foris, L. Neumann, and B. Csebfalvi. An Analysis of Quasi-Monte Carlo Integration Applied to the Transillumination Radiosity Method. *Proceedings of Eurographics 97*, 1997.

[Sol78] H. Solomon. Geometric Probability, siaam-cbms 28. 1978.

[SPNP96] M. Sbert, X. Pueyo, L. Neumann, and W Purgathofer. Global Multi-Path Monte Carlo Algorithms for Radiosity. *The Visual Computer*, pages 47–61, 1996.

[SPP95] M. Sbert, F. Pérez, and X. Pueyo. Global Monte Carlo. A Progressive Solution. *Rendering Techniques '95. Springer Wien New York*, pages 231–239, 1995.



(a) (b)

(c) (d)

Fig. 10. *(a) Monte Carlo random sampling. Number of lines = 266000 (b) Halton low discrepancy sampling. Number of lines = 266000 (c) Monte Carlo random sampling. Number of lines = 532000 (d) Halton low discrepancy sampling. Number of lines = 532000*