Quasi-Monte Carlo techniques in Multipath radiosity

Francesc Castro, Mateu Sbert
Institut d'Informàtica i Aplicacions
Universitat de Girona
E-17071 Girona
Catalunya - Espanya
castro@ima.udg.es

October 2, 2002

Abstract

This paper reviews the Integral Geometry based Multipath algorithm, emphasizing in the use of quasi-Monte Carlo sequences instead of pseudo-random numbers. Several efficient quasi-Monte Carlo sequences are analyzed, together with their application to the Multipath method.

1 Introduction

One important issue in computer graphics is the obtaining of realistic images. Different simulation techniques can be used for this purpose. One of these techniques, the more common to deal with diffuse (or Lambertian) surfaces, is the radiosity technique [SP94]. This technique was borrowed from thermal engineering in the eighties [SH92].

Radiosity technique uses a discretization of the surfaces in the environment in polygons called *patches*. This drives us to the radiosity system of equations, a linear system of equations in which we have, for each patch

$$B_i = E_i + \rho_i \sum_j F_{ij} B_j \qquad (1)$$

where B_i is the outgoing radiosity of patch i (power per surface unit), E_i is the emittance of patch i (emitted power per surface unit) and ρ_i is the reflectance of patch i (fraction of the incoming power to patch i that is reflected). Note that B_i, E_i, ρ_i are supposed to be constant along patch i. Finally F_{ij} is the form factor from patch i to patch j. The form factor F_{ij} represents the fraction of power exiting patch i that lands on patch j, and it can be interpreted as the probability of a photon leaving patch i to reach patch j. Form factors are purely geometric quantities that, although unknown, are very easy to simulate using random lines. Thus Monte Carlo methods are frequently used to solve the radiosity system. The Multipath algorithm, introduced in [SPNP96], is a Monte Carlo radiosity algorithm that simulates the trajectory of light particles by means of an uniform density of lines. It belongs to a family of methods that use random global lines (or directions) to transport energy, called by different authors global Monte Carlo, global Radiosity or transillumination methods [SPP95, Neu95, SKFNC97]. The Multipath method will be reviewed in section 2.

On the other hand, the results obtained with Monte Carlo methods can be improved by using sequences of deterministic numbers instead of the classic pseudo-random values. These sequences, called quasi-Monte Carlo sequences, are known by their higher uniformity, which happens to be more important than randomness for the convergence rate of Monte Carlo methods. Quasi- Monte Carlo sequences are dealt with in section 3.

We study here quasi-Monte Carlo sequences applied to the Multipath algorithm. The results, that complete the work in [CMS98, CS99], are presented in section 4, and demonstrate the best behaviour of the quasi-Monte Carlo generation in front of classic pseudo-random generation. Finally the conclusions are presented in section 5.

2 The Multipath algorithm

2.1 Local lines and global lines

Classic Monte Carlo approaches to radiosity and global illumination [WEH89, Shi91, PM92, MP93, FP93] use to distribute the energy random lines that are local to the surfaces in the scene, that is, these lines are created by sampling a point on the surface and a random direction. We refer to these approaches as *local Monte Carlo techniques*.

In contrast, M.Sbert introduces in the context of radiosity [Sbe93, Sbe97] the use of an uniform density of lines in the sense of Integral Geometry [RPS51, San76]. This means that the density of lines must be homogeneous and isotropic. These lines are independent of the surfaces or patches in the scene, in contraposition to local lines, and they are referred to as *global lines* (see Fig. 1), and consistently the Monte Carlo approaches based on such lines are referred to as *global Monte Carlo*. It can be shown [Sbe97] that this uniform density submits on each patch the same density of lines used in a local approach in which importance sampling with pdf $\frac{cos\theta sin\theta}{\pi A}$ is done (A is the area of the patch and θ is the angle between the normal of the patch and the line).

This uniform density was obtained in [Sbe93] by sampling pairs of random points on the surface of a sphere that wrapped the whole scene. Other forms of generating this density are summarized in [CS99].

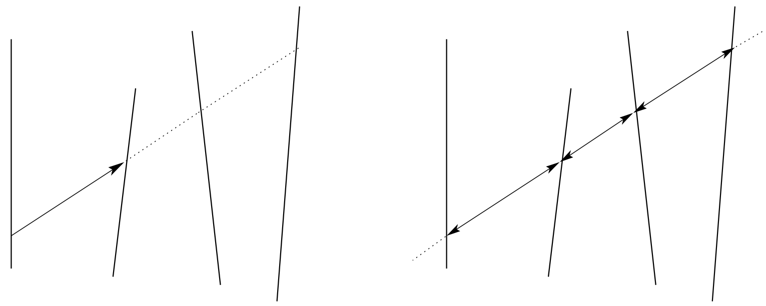


Figure 1: *(Left) Local lines: lines are cast from each patch, and power is transfered only to the nearest intersected patch. (Right) Global lines: lines are independent of the patches, and all intersections are used to transfer power.*

2.2 The Multipath algorithm

The above mentioned density of lines is also used in a simulation of the light particles paths: the Multipath method [SPNP96].

In the Multipath method every global line simulates the exchange of energy between several pairs of patches. In the random walk context we can say that every global line contributes to several paths (Fig. 2a). Now we review the Multipath algorithm:

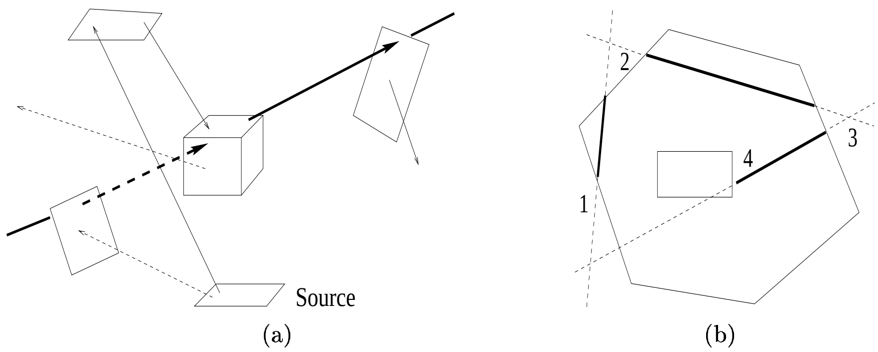


Figure 2: *(a) A global line (the thick one) simulates two paths, indicated with the continuous stroke and the dashed stroke (b) A path can contribute to the emission of power from several patches. In the 2D figure, path 1-2-3-4 simulates paths 1-2-3-4, 2-3-4 and 3-4. This is similar to the covering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

Global lines are intersected with the scene. For each line the intersections are sorted by distance in an intersection list. Each patch keeps two quantities. One records the power accumulated, the other one is the unshot power. For every pair of patches along the intersection list, they exchange their unshot power, decreased by the respective reflectances, and add to their accumulated power the unshot power of the other patch (also decreased by the reflectance). If a patch is a source, we keep also a third quantity, the emitted power per line exiting the source. Thus, if one of the patches of the pair is an emitter patch, we must add this emitted power per line to the unshot power. This "emitted per line" power is previously computed in the following way: given the number of 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 power by this number gives the predicted power per line. Note that each global line simulates different order reflections, and moreover both directions of the line are considered.

2.3 Strength and limits of the Multipath method

The strong points of the Multipath method are the following: First, all intersections of a line are used. Second, the power transfer is bidirectional. Third, each random line contributes to several paths (Fig. 2a). And finally, because the global density mimics on each patch the same local density, each path is used to transport different logical paths (see Fig. 2b), as with the covering path concept [Rub81].

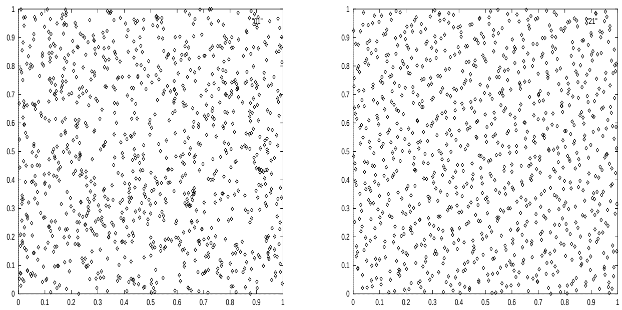
A limitation of the Multipath method is that in its first stages the distribution of power is only possible from light sources, and so most of the lines cast in these first stages (the lines that do not cross any light source) are wasted. To avoid this behaviour a preprocess, called *first shot*, is done [CMS98] [SKSMT00]. In this preprocess the primary power is cast from the source patches by generating local lines that exit from the surface of each emitter patch. 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.

3 Quasi-Monte Carlo generation

We call quasi-Monte Carlo sequences some sequences of deterministic (non-random) numbers that can improve the results obtained with pure Monte Carlo generation [Nie92]. There exist different number generators in the context of quasi-Monte Carlo, some of which will be reported in this paper.

The main advantage of quasi-Monte Carlo sequences of numbers in front of classic random numbers is their higher uniformity. This means that quasi-Monte Carlo numbers are spread in a more uniform way over the domain than classic random numbers. We can observe it in Fig. 3, where we can see 1000 2D points generated in both -Monte Carlo and quasi-Monte Carlo- ways.

The use of more uniform quasi-Monte Carlo numbers for integration produces convergence rates faster than the convergence rate $O(\frac{1}{\sqrt{N}})$ obtained when using classic Monte Carlo. Uniformity is measured by *discrepancy*. Discrepancy



(a) (b)

Figure 3: *(a) 1000 Monte Carlo generated 2D points. (b) 1000 quasi-Monte Carlo (Halton seq.) generated 2D points.*

[Nie92] is a measure of the deviation of a set of samples from the uniform distribution. In other words, the lower is the discrepancy of a set of samples, the higher is their uniformity. Quasi-Monte Carlo sequences are also known as *low discrepancy sequences.*

A quasi-Monte Carlo sequence is said to be *k-uniform* when it can be used in integrals of dimension k. If we want to integrate over a domain of dimension k, we can use one k uniform sequence, but also k independent 1-uniform sequences. Some of the low discrepance sequences that we will list next are only 1-uniform, so we have to consider k of these sequences to obtain a valid sequence of k-dimensional points.

Next we will describe the quasi-Monte Carlo sequences that will be used in chapter 4.

3.1 Radical inversion: Van der Corput sequence

The construction of some low discrepancy sequences is based on the radical inverse function [HW64]. Given a basis b, the radical inverse function Φ_b maps the natural numbers on the interval $I = [0, 1)$. This mapping is based on the representation of a natural number in basis b:

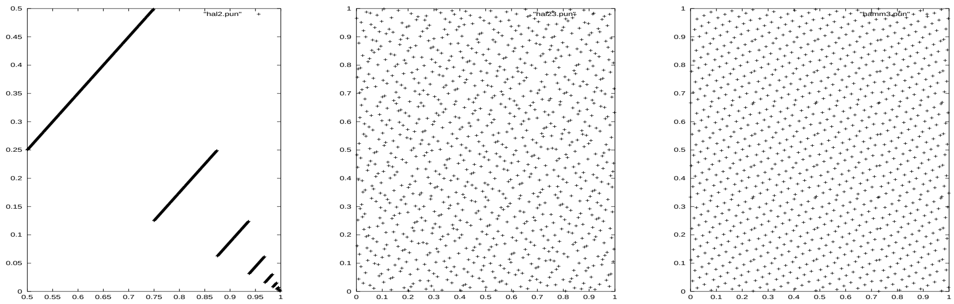
$$\Phi_b(i) = \sum_{j=0}^{\infty} a_j(i) b^{-j-1} \quad (2)$$

where $a_j(i)$ is the jth digit in the representation of i in basis b, that is:

$$i = \sum_{j=0}^{\infty} a_j(i) b^{j} \quad (3)$$

For instance, let us consider basis 2. Since the representation of the number 19 in basis 2 is 10011, the radical inverse of 19 in basis 2, $\Phi_2(19)$, is equal to the value represented by 0.11001 in the same basis (that is, 1/2+1/4+1/32). Note that it corresponds to mirror at the decimal point the representation of the number in basis b. The sequence of numbers obtained in this way is known as the Van der Corput sequence of basis b.

Van der Corput sequences are only 1-uniform, so they are not appropiate to integrate over domains of dimension greater than 1. In Fig. 4 (a) we can see the unusable distribution of the 2D points generated with this sequence.



(a) (b) (c)

Figure 4: *(a) 1000 2D points from a Van der Corput sequence (basis 2). (b) 1000 2D points from a Halton sequence (basis 2 and 3). (c) 1000 2D points from a Hammersley sequence (basis 3).*

3.2 The Halton sequence

The problem of 1-uniformity in Van der Corput sequences can be avoided by taking d (being d the dimension of the domain) independent Van der Corput sequences, as seen in Fig. 4 (b) for $d = 2$. This is called the Halton sequence, and it is a simple way to obtain low discrepancy sequences of dimension d. To generate d-tuples we use d Van der Corput sequences, using d different basis that are relative primes. For instance we can use as basis the first d prime numbers. Thus, the Halton sequence in d dimensions is built by

$$x_i = (\Phi_{p_1}(i), ..., \Phi_{p_d}(i)) \quad (4)$$

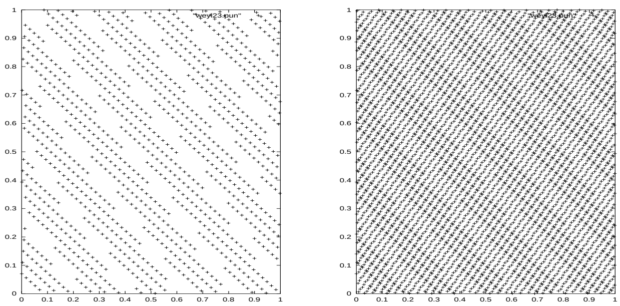
where $p_1, ..., p_d$ are the first d prime numbers.

3.3 The Hammersley sequence

The Hammersley sequence has a generation also based in radical inversion. The main difference with the Halton sequence is that the number of samples has to be fixed *a priori*. Let N be this number of samples, we have

$$x_i = (i/N, \Phi_{p_1}(i), ..., \Phi_{p_{d-1}}(i)), i \in (0, ..., N-1) \quad (5)$$

The Hammersley sequence appears to be a bit more uniform than the Halton sequence, as it can be appreciated in Fig. 4 (c). Note that, as in the Halton sequence, the Hammersley sequence is originated from d independent 1-uniform sequences (d being the dimension).



(a) (b)

Figure 5: *(a) 1000 2D points generated from 2 Weyl sequences (from $\sqrt{2}$ and $\sqrt{3}$) (b) 5000 2D points generated from 2 Weyl sequences (from $\sqrt{2}$ and $\sqrt{3}$). Note that the points tend to fill the empty spaces*

3.4 The Weyl sequence

The Weyl sequence is originated from the fractionary part of the multiples of the square root of prime numbers. That is

$$x_i = (frac(i \times \sqrt{p})) \tag{6}$$

where p is a prime number. The Weyl sequence is 1-uniform, so we need to use, as in the case of the Halton sequence, d different Weyl sequences, d being the dimension of the domain over which we want to integrate. Traditionally the square roots of the first d prime numbers are used:

$$x_i = (frac(i\sqrt{p_1}), ..., frac(i\sqrt{p_d})) \tag{7}$$

In Fig. 5 (a) and (b) we can observe the representation of 1000 and 5000 2D points generated from 2 Weyl sequences. Note that the points tend to fill the empty spaces.

3.5 Scrambled Halton sequence

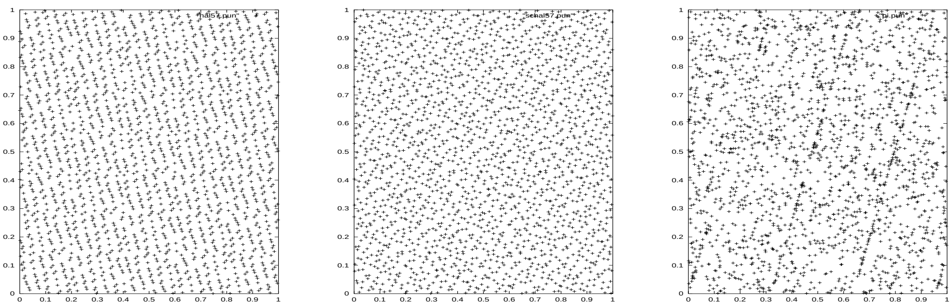
The radical inverse function in basis b has subsequences of b equidistant values spaced by $\frac{1}{b}$ that produce undesirable effects like alignement of points in lines. These efects are more noticeable when using high basis. These effects are reduced using scrambled sequences. Scrambled sequences consists of changing the order of quasi-Monte Carlo sequences. One of the algorithms to scramble the sequences is the Faure's permutation [Fau92]. This starts with the permutation (0,1) for basis $b = 2$, and it is extended to higher basis by distinguishing two cases:

- If b is even, we construct the permutation corresponding to basis b by multiplying by 2 the permutation corresponding to $b/2$ and then appending the same values incremented by 1.

- If the basis b is odd, we take the permutation correspondig to $b-1$, incrementing by 1 each value greater or equal than $\frac{b-1}{2}$ and inserting in the middle the value $\frac{b-1}{2}$.

In Fig. 6 we can see, on the left, 2000 2D points of a non-scrambled Halton sequence with basis 5 and 7, and on the right side 2000 points of a scrambled Halton sequence with the same basis. The last one seems to be more uniform. The effect of scrambling would be more noticeable when using higher basis.

A drawback of the scrambled Halton sequence is that it is not possible to generate each value from the previous one, as done in a standard Halton sequence. This produces an increase of cost in the generation of the random values.



(a) (b) (c)

Figure 6: *(a) 2000 2D points from a non-scrambled Halton sequence with basis 5 and 7. (b) 2000 2D points from a scrambled Halton sequence with the same basis (c) 1000 2D points from the sequence* $frac(\pi^i)$

3.6 A ∞-uniform sequence

Monte Carlo sequences are ∞-uniform. This means that they can be used to integrate in any dimension. A problem of many quasi-Monte Carlo sequences is that they lose the ∞-uniformity characteristic of the Monte Carlo generation. Sequences as Van der Corput or Weyl are only 1-uniform. As seen previously, this problem is solved by using a number d of independent 1-uniform sequences (d being the dimension).

There are low discrepancy sequences that are supposedly ∞-uniform. For instance the next one

$$x_i = frac(\pi^i) \tag{8}$$

is supposed to be ∞-uniform with probability 1 although nobody has proven it yet. This is also valid replacing π by any transcendent number. In Fig. 6 (c) we represent 1000 2D points obtained from this sequence. Although the sequence produces theoretically valid numbers, in practice the finite representation of π produces undesirable patterns when increasing the number of samples. Other sequences supposed to be ∞-uniform can be obtained, but they also present the same problem.

4 Using quasi-Monte Carlo sequences in the Multipath algorithm

4.1 Dimension of the low discrepancy sequences

The Multipath algorithm, like other Monte Carlo algorithms in global illumination, uses random lines to distribute the power. Each random line, regardless of being global or local, is generated from 4 values. This drives us to using low discrepancy sequences of dimension 4 instead of 2, as can be seen in [CMS98, CS99]. This allows to avoid the correlations between the values involved in the generation of a line. Note that the dimension of the quasi-Monte Carlo sequences is a critical point when using 1-dimensional sequences, as Halton, Hammersley, Weyl or Sobol sequences.

4.2 Results

Some experiments were presented in [CMS98] and [CS99], where we tested Halton and Sobol sequences. Mean Square Error (MSE) was reduced by 30-50 per cent, depending on the scene. Also, a smoother convergence was obtained. No special differences between the different quasi- Monte Carlo generators used were found, and it was concluded that the use of quasi-Monte Carlo sequences improved the performance of the Multipath method, getting an error that decreases faster than $\frac{1}{\sqrt{N}}$ characteristic of classic Monte Carlo.

Here we test the rest of the sequences presented in the previous section. We have first used a single scene -*SixCubes*- (see Fig. 10) composed by 6 cubes in a room. The graph in Fig. 7 shows that the best results are obtained using the Halton and Weyl sequences, where a reduction of the MSE by nearly 50 per cent is obtained. Scrambled Halton produces a MSE slightly lower, but the increase on cost due to the generation of the scrambled points makes unattractive the gain in the MSE, whereas the Hammersley sequence behaves clearly worse.

We have also tested a more complex scene -*Room*- (see Fig. 11) which contains a table and a desk with some small objects. The results in Fig. 11 indicate a reduction of the MSE by approximately 30 per cent using Weyl sequence. Halton and Scrambled Halton sequences produce a reduction of the MSE slightly smaller. Finally the Hammersley sequence does not produce any gain.

We have also tested the sequence $frac(\pi^i)$, that is supposed to be ∞-uniform. The results obtained with this sequence show no convergence in any of the tested scenes. This is due to the patterns presented by the values generated with this sequence (see previous section), that make it unusable if we need a large number of samples.

Summarizing all the results, we conclude that the scrambled Halton sequence produces a MSE slightly better than the Halton sequence, but its higher cost

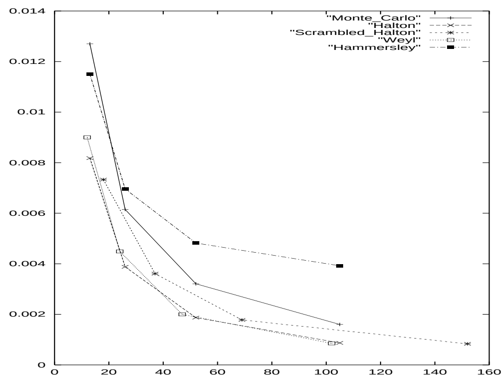


Figure 7: *Scene SixCubes. Horizontal axis: time (sec.). Vertical axis: mean square error*

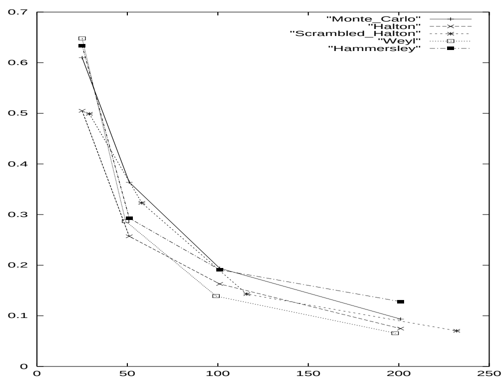


Figure 8: *Scene Room. Horizontal axis: time (sec.). Vertical axis: mean square error*

makes it unattractive. The Weyl sequence presents a performance very similar to the ordinary Halton sequence. The Hammersley sequence behaves clearly worse than Monte Carlo. The Sobol sequence, as shown in [CMS98], behaves similar to the Halton sequence. We can conclude that the use of low discrepancy sequences like Halton, Sobol and Weyl in the Multipath algorithm produces better results than classic Monte Carlo, with reductions of the MSE by 30-50 per cent, depending on the scene, with no increase in the cost.

In Figures 10 and 11 we compare pairs of images obtained with a similar cost, where the ones on the left have been generated using Monte Carlo values, and the ones on the right have used a quasi-Monte Carlo sequence -the Weyl sequence-. Note the reduction of the noise and thus the better quality of the quasi-Monte Carlo generated images.

4.3 Direct illumination and Quasi-Monte Carlo

Additional experiments have been done to establish the gain of quasi-Monte Carlo when considering only the direct illumination. We have used the scene *SixCubes*, and we have compared Monte Carlo random generation with Halton generation. The results are shown in Fig. 9. MSE has been reduced between 35 and 60 per cent. The important conclusion from these results is that the gain due to the use of quasi-Monte Carlo is clearly more noticeable in the distribution of direct illumination than in higher order reflections. Note that direct illumination is distributed in a preprocess, the first shot, in which local lines are used instead of the global lines used in the distribution of indirect illumination. However this difference in gain is not due to the origin -local or global- of the lines, but to the fact that quasi-Monte Carlo sequences are not ∞-uniform, and so there is a loss of the quasi-Monte Carlo gain when computing higher bounces. This problem was already discussed in [SKP99].

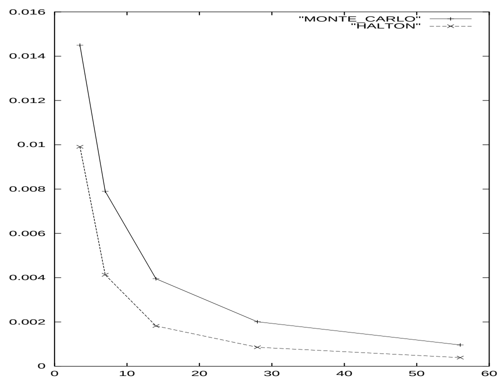
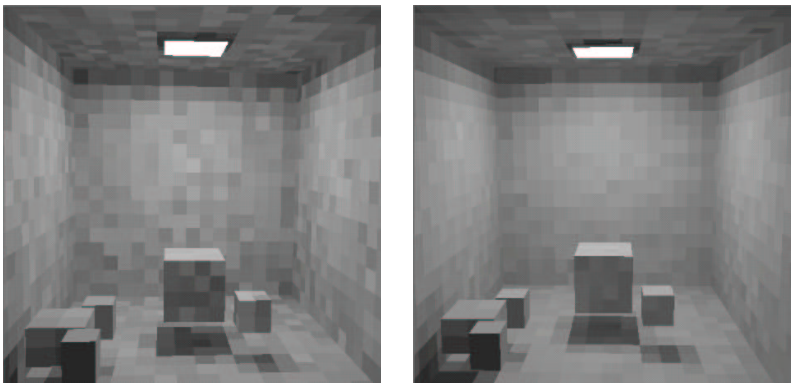


Figure 9: *Scene SixCubes. MSE in the simulation of direct illumination (Monte Carlo and Halton sequences)*



(a) (b)

Figure 10: *Scene SixCubes. (a) Monte Carlo generation. 320 K lines. Time: 6 sec. (b) Weyl sequence. 320 K lines. Time: 6 sec.*

(a) (b)

Figure 11: *Scene Room. (a) Monte Carlo generation. 3456 K lines. Time: 205 sec. (b) Weyl sequence. 3456 K lines. Time: 202 sec.*

5 Conclusions

We have reviewed several quasi-Monte Carlo sequences and applied them to the Multipath algorithm. The obtained results confirm the ones in [CMS98, CS99] with Halton and Sobol sequences in the sense of the usefulness of quasi-Monte Carlo sequences in the radiosity Multipath context.

The scrambled Halton sequence produces a MSE a bit lower than the ordinary Halton sequence, but as the generation of the scrambled values is more costly, the gain in this case is practically null. The Weyl sequence produces very similar results to the Halton sequence, and the cost of generating the values is more or less the same. In the case of Hammersley sequence, the performance has been clearly worse than Monte Carlo. Finally we have studied a ∞-uniform sequence, $frac(\pi^i)$, that produces no convergence because of the repetitions of the quasi-random values that make the sequence unusable if we need a large number of values.

On the other hand, we have studied quasi-Monte Carlo sequences applied to the distribution of the direct illumination, concluding that the gain is more noticeable than in the whole Multipath algorithm. This can be explained by the fact that the quasi-Monte Carlo sequences are not ∞-uniform.

In short, we conclude that it is possible to improve the performance of the Multipath algorithm by using quasi-Monte Carlo sequences of numbers instead of the classic pseudo-random values.

6 Acknowledgements

This project has been funded in part with grant number TIC 2001-2416-C03-01 from the Spanish Governement and grant number 2001SGR-00296 from the Catalan Governement.

References

[CMS98] F. Castro, R. Martinez, and M. Sbert. Quasi monte carlo and extended first shot improvements to the multipath method. *Proceedings SCCG'98, (Budmerice, Slovakia)*, 1998.

[CS99] F. Castro and M. Sbert. Aplication of quasi-monte carlo sampling to the multi-path method for radiosity. *Springer series Lecture Notes in Computational Science and Engineering, Springer*, 1999.

[Fau92] H. Faure. Good permutations for extreme discrepancy. *J.Number Theory, n.42*, pages 47–56, 1992.

[FP93] M. Feda and W. Purgathofer. Progressive ray refinement for monte carlo radiosity. *Proceedings of Eurographics Workshop on Rendering 1993*, pages 15–25, 1993.

[HW64] J. Halton and G. Weller. Radical-inverse quasi-random point sequence. *Comm. ACM 7 (1964), n. 12*, pages 261–269, 1964.

[MP93] S.P. Mudur and S.N. Pattanaik. Monte carlo methods for computer graphics. *State of the Art Report (STAR) series*, 1993.

[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*. Capital City Press, 1992.

[PM92] S.N. Pattanaik and S.P. Mudur. Computation of global illumination by monte carlo simulation of the particle model of light. *Proceedings of third Eurographics WorkShop on Rendering*, pages 71–83, 1992.

[RPS51] J. Rey-Pastor and L. Santalo. *Geometría Integral*. Espasa Calpe, 1951.

[Rub81] Reuven Y. Rubinstein. Simulation and the monte carlo method. 1981.

[San76] L. Santaló. *Integral Geometry and Geometric Probability*. Ed. Addison-Wesley, New York, 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.

[Sbe97] M. Sbert. The use of global random directions to compute radiosity. global monte carlo methods. *Ph.D. thesis. Universitat Politècnica de Catalunya, Barcelona*, 1997.

[SH92] Robert Siegel and John R. Howell. *Thermal Radiation Heat Transfer, 3rd Edition*. Hemisphere Publishing Corporation, New York, NY, 1992.

[Shi91] P. Shirley. Time complexity of monte carlo radiosity. *Proceedings of Eurographics 91*, pages 459–465, 1991.

[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.

[SKP99] L. Szirmay-Kalos and W. Purgathofer. Analysis of the quasi-monte carlo integration of the rendering equation. *Technical report TR-186-2-98-22*, 1999.

[SKSMT00] L. Szirmay-Kalos, M. Sbert, R. Martinez, and R. Tobler. Incoming first-shot for non-diffuse global illumination. *Proceedings of SCCG*, 2000.

[SP94] F.X. Sillion and C. Puech. *Radiosity and Global Illumination.* Morgan Kaufmann Pubilishers, Inc., 1994.

[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.

[WEH89] J. Wallace, K. Elmquist, and E. Haines. A ray tracing algorithm for progressive radiosity. *SIGGRAPH 89 Conference Proceedings, Annual Conference Series*, pages 315–324, 1989.