Quasi Monte-Carlo and extended first shot improvements to the multi-path method

Francesc Castro, Roel Martínez, Mateu Sbert

Institut d'Informàtica i Aplicacions
Universitat de Girona
Lluis Santaló s/n, 17071 Girona, Spain.
phone +34 (972) 418439/8419
fax +34 (972) 418399
e-mail {castro|roel|mateu}@ima.udg.es

Abstract

We present in this paper two kind of improvements to the multi-path technique for radiosity. The first one is the use of quasi-Monte Carlo sequences to generate the lines used in the transport of energy. Different generators will be tried, and important gains in efficiency will be demonstrated. The second improvement is the overcoming of the bad behaviour of the algorithm when dealing with scenes where only a small portion of it is visible from the sources. This will be done by extending the so called first shot, to further smooth the undistributed radiosity all over the scene. In this way the transport of energy by the global lines will be optimal.

Keywords

Monte Carlo, Quasi Monte Carlo, Radiosity

1 Introduction

The multi-path method for radiosity was 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]) methods. They use random global lines (or directions) to transport energy. 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 casted from. The global lines can take advantage of all intersections with the scene, however some inefficiencies when using global lines arise from mainly two places. The first one is inherent to all Monte Carlo methods, which is the high discrepancy [Shi91] of random (or better seudo-random) sequences. This implies the use of a huge number of lines to obtain an acceptable image. The second inefficiency is inherent with global lines. The probability of a global line to intersect a patch is proportional to the area, thus, expanding the energy of a small area source with global lines is very inefficient. This inefficiency has been solved in part with a first shot preprocess, which expands the direct illumination and converts the receivers into secondary emitters. Neverthless, for certain scenes, such as ones with poor visibility from the sources to the scene (such as a source pointing to the ceiling), the inefficiency remains.

In this paper we will study the applicability of quasi Monte Carlo sequences to overcome the first inefficiency exposed. This is the step followed by all radiosity Monte Carlo methods, although in some of them it is very unclear the gain in efficiency [Kel96a]. And a solution to the second inefficiency will be given, based on extending the first shot to the secondary sources and beyond, so as to obtain an ideal situation for the energy transport with the global lines.

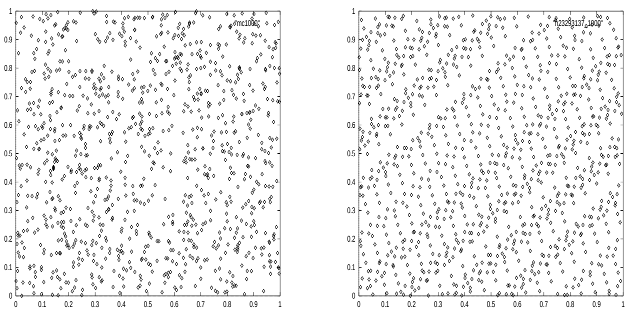
The organization of this paper is as follows. In next section we will remember the multi-path algorithm, and some applications of quasi-Monte Carlo sequences to radiosity. In section 3 the applicability of those sequences to the multi-path algorithm will be studied. In section 4 we will present the extended first shot improvement, and in section 5 our conclusions and future work.

2 Previous work

2.1 The multi-path algorithm

The multi-path method exposed in [SPNP96] shows that is possible to simulate a random walk using a global density of lines. We remember now the algorithm. We cast a predetermined number of random lines, taking pairs of random points on a sphere surrounding the scene. Each line will produce an intersection list, and we follow this list taking into account successive pairs of patches. Each patch (if not emitter) has with it two quantities. One records the energy accumulated, the other one is the unshot energy. For every pair of patches along the intersection list, the first patch of the pair will transmit its unsent energy to the second patch of the pair. So the unsent energy of the first patch is reset to zero, and the two quantities at the second patch, the accumulated and the unsent energy, are incremented. In the case of the source, we keep also a third quantity, the emitted energy per line exiting the source. We compute that energy in the following way: given the number of the lines we are going to cast, we compute for any light source beforehand the forecasted number of lines passign through it. This can be done with Integral Geometry methods ([San76]). The division of the total source energy by this number gives the predicted energy of one line. Then, if the first patch of a pair is a source patch, the energy transported to the second patch of the pair will also include this predicted energy portion.
For the multi-path method, as well as for all global Monte Carlo methods to work efficiently, we need first to "smooth" the scene ([SPNP96]), for example distributing the direct illumination in the usual way (sending rays from the source). We call this process the first shot. After that the new sources are the patches that have received direct illumination, and the old sources are no more emitters. In this way we avoid the problem of the many lines not transporting energy in the initial stage of the simulation. However the first shot remains inefficient for some kind of scenes, where just a small amount of patches are visible from the sources (see figure 6(a)).

2.2 Quasi Monte Carlo generation



(a) (b)

Figure 1: *(a) Monte Carlo generation. (b) Quasi Monte Carlo generation*

The concept of discrepancy appears in several papers, like [Shi91], [OA97] or [Kel96b]. We can define the discrepance as a measure for the deviation of a point set from uniform distribution. Look at figure 1 . On the left there is a set of 1000 2D points generated with a simple Monte Carlo generator. On the right the points have been generated with a quasi Monte Carlo generator. We can easily see that the discrepancy is lower in the right figure, because it has a more "uniform" distribution. Intuitively, we can talk about low discrepancy points when the new points "try" to fill the empty spaces.

We call quasi Monte Carlo numbers to some sequences (as Halton or Sobol sequences) of deterministic numbers (namely, non-random numbers) that can improve the results obtained with pure Monte Carlo generation. The improvement is caused by the fact that in quasi Monte Carlo generation, the generated points have a lower discrepancy (we can talk of low discrepancy sets of points).

Several authors have applied quasi Monte Carlo sequences in ray tracing or radiosity techniques. Keller ([Kel96a]), uses Halton and Hammersley sequences applied to the Form-Factors computation and to a random walk radiosity solution ([Kel96b]). However, it is unclear the gain in efficiency obtained in [Kel96b]. [OA97] uses them in Ray Tracing. In [SKFNC97] and [SKFP98] they are applied to the transillumination method, and in [SKCP98] to a random walk simulation. Unfortunately, no comparison is given with Monte Carlo sequences, although in [SKFNC97] it is shown from a theoretical basis the good applicability of quasi-Monte Carlo sequences to the transillumination method. Neumann et al. [NNB97] use well distributed ray sets generated with quasi Monte Carlo sequences to compute the radiosity, obtaining improvements of half an order of magnitude.

3 Application of quasi Monte Carlo generators to the multi-path method

If we reduce, by application of quasi Monte Carlo sequences, the discrepancy between the points, we get sets of points with a distribution more similar to a uniform distribution. So, it seems logical to expect better results in the estimation of the radiosity using this low discrepancy sequences.
In the work presented in this paper, we have used two different types of quasi Monte Carlo sequences:

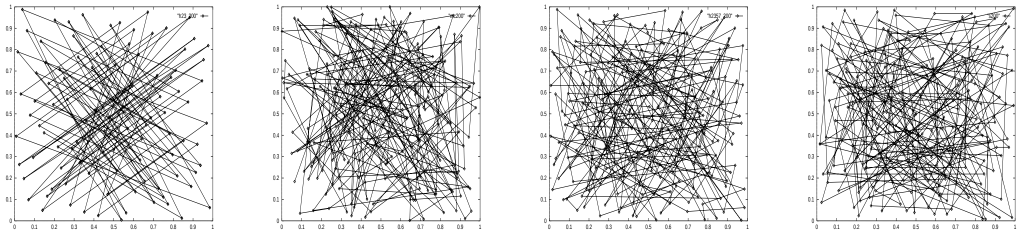
- Halton sequences.
- Sobol sequences.

Halton sequences, as defined in [Pre94], use the representation of the numbers in basis b_i, where b_i is a prime number. In fact, Halton sequences are sequences of k-tuples, where k is the dimension of the generation. Then, to get a sequence of k-tuples, we make each component a Halton sequence with a different prime base. Usually, the first k primes are used, but, as we will see later in this paper, other possibilities can be explored.

There exists a great variety of different low discrepance sequences that can be included in the generic name of Sobol sequences. In this paper we use an efficient variant proposed by Antonov and Saleev, that can be found in [Pre94]. We will not study in depth this variant.

We have applied these low discrepance sequences to the global Monte Carlo multi-path method, described in [SPNP96]. The next question is crucial: how many random (or quasi-random) numbers are needed to generate a random (or quasi-random) ray? In the multi-path algorithm, the answer to this question is four numbers. In the generation of global rays we obtain a ray from two points on the sphere, and we need two numbers for every point. In the first shot we need two numbers to obtain a random point on the surface of the patch, and two more numbers to get the direction.

A first approximation to the problem could be shown in two dimensional sequences, namely, sequences of 2-tuples. In this case, every item in the sequence corresponds to a point (or a direction in the first shot). The results obtained have not been good: the error in the computed radiosities has been enormous (so much with Halton numbers like with Sobol numbers), much worse that in



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

Figure 2: *(a) Halton generation of 2-dimension sequences. (b) Monte Carlo generation. (c) Halton generation of 4-dimension sequences. (d) Sobol generation of 4-dimension sequences*

Monte Carlo generation. The reason of that behavior is the strong correlation that exists between two consecutive points in the quasi Monte Carlo sequences of dimension 2. Looking figure 2 in (a) and comparing it with (b) we can understand the problem. In (a) we can see the order of points generated with a Halton sequence of dimension 2 (especifically using 2 and 3 as basis). This figure presents a important particularity: every point is quite "far" from the next one. Namely, there exist some important correlation. Since a ray is obtained from two points on the sphere, and every point corresponds to an item of the sequence, the correlation between these items can produce a density of rays very far from the prospective uniform density. Conversely, if we look at (b), that corresponds to a Monte Carlo generation, we can see that here the correlation is not so clear.

The key of the question is the following one: if the problem is the correlation between the two points that origine a ray, why do not we use a quasi Monte Carlo sequence with dimension 4 instead of 2?. In this way, every item in the sequence will correspond to a ray instead of a point, and so the correlation in the generation of a ray will disappear. If we look at figure 2 at (c), we can see a Halton generation of dimension 4, and now the strange effect observed in figure 2 at (a) is not so clear. We can affirm the same looking (d) , that corresponds to a Sobol generation in dimension 4. As we will see in the analysis of results, now the obtained radiosities are much more accurate.

3.1 Results

We compare here the results (radiosities) obtained using Monte Carlo with the results using the different quasi Monte Carlo sequences. We will compute the Mean Square Error (MSE) as known:

$$MSE = \frac{\sum_i A_i * (\hat{B}_i - B_i)^2}{\sum_i A_i} \tag{1}$$

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 executing a multi-path algorithm in which we cast a lot of lines.

Initially we have chosen a simple scene that consists of a cubic room with 5 cubes in its interior and one light source near the ceiling. We call this scene 5CUBES. Next, we will treat a more complex scene, called 4CHAIRS, that is a room with 4 chairs, a table and a cube on the table. In this scene, the light source is placed near the ceiling too.

Several executions have demonstrated that the time per execution does not depend on the method of generation of the quasi-random numbers. So, we have decided to do without the parameter time in both tables and graphs. We have only considered number of lines and MSE as the parameters to evaluate every method.

As is well known, in Monte Carlo the expected value of the Mean Square Error decreases as $\frac{1}{N}$. From here it is easy to prove the next formula:

METHOD	*SLOPE IN 5CUBES*	*SLOPE IN 4CHAIRS*
HALTON (2,7,3,5)	-1.098	-0.998
HALTON (11,13,17,19)	-0.844	-1.113
SOBOL	-1.102	-1.122

Table 1: *Asymptotical behaviour of the MSE*

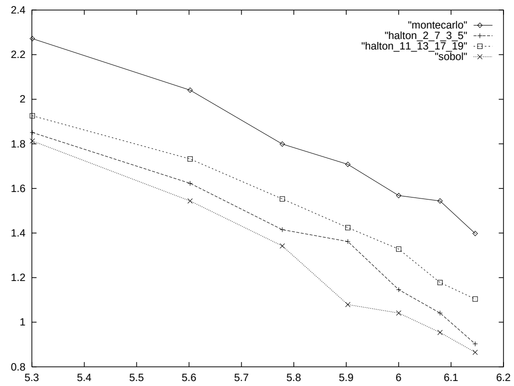


Figure 3: *Asymphtotical behaviour of the MSE in the scene 5CUBES*

$$log(MSE) \in O(-log(N)) \tag{2}$$

being N the number of casted rays. That means that, if we take logarithms, the graph MSE vs. number of line must be linear with slope -1. This is its asympthotical behaviour. Then, if we consider quasi Monte Carlo generation, we can expect a better behaviour, namely, a slope minor that -1. We have studied the asymptotical behaviour of MSE in both scenes 5CUBES and 4CHAIRS, and the results are showed in table 1. There we can see that the best asymthotical results are obtained, in both scenes, with Sobol generation. The error in Halton generation is very related with the choosed basis. So in scene 5CUBES we obtain optimal results using 2, 7, 3 and 5 as basis, whereas in scene 4CHAIRS the optimal error is found using 11, 13, 17 and 19. The relation between the tested scene and the optimal basis is, for the moment, empirical. In figures 3 and 4 we have the graphs of the asymptotical behaviour for both scenes. Note that, in every case, only the best Halton sequence is represented.

We can see some of the images obtained in figure 5. It is quite clear that, using the same number of rays, the obtained images are better in quasi Monte Carlo generation. Particularly, we can observe that the noise effect is more reduced in quasi Monte Carlo.

4 Extended First Shot

In the first shot the authors proposed to "smooth" the scene sending rays from the source, that is expanding direct illumination. This idea works fine but in some cases it is innefficient. For example, if a lot of patches of the scene are visible from the sources the direct illumination will work well because all these patches receive energy. But what happens if just a small amount of patches are visible from the sources? (see figure 6). After the first shot just the patches that are visible will have received energy, a small proportion of all patches. Then the global pass, the second step of the multi-path simulation, will work inneficiently because we waste a lot of lines not transporting energy at all.

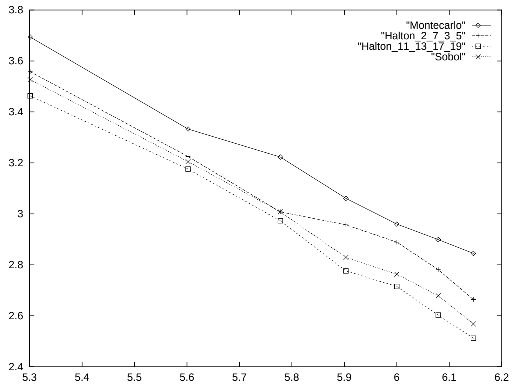
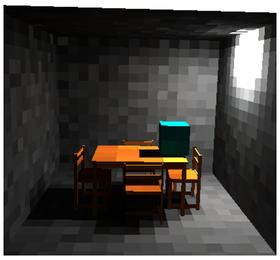


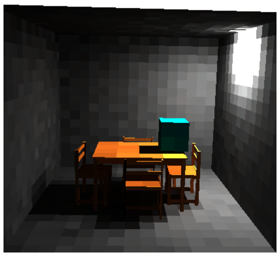
Figure 4: *Asymphtotical behaviour of the MSE in the scene 4CHAIRS*

(a)



(b)

(c)



(d)

Figure 5: *(a) Monte Carlo generation. Number of lines = 266000 (b) Halton (11,13,17,19) generation. Number of lines = 266000 (c) Monte Carlo generation. Number of lines = 532000 (d) Halton (11,13,17,19) generation. Number of lines = 532000*

4.1 Proposed algorithm

We propose here a new algorithm to solve this inefficiency. The new algorithm applies the first shot more than one time. In the first step of the new algorithm, we "smooth" the scene sending rays from the sources (we leave some energy undistributed so as not to waste the lines, used in the global step, hitting the original sources). All the rays transport a predetermined, by the user, amount of energy. After this first step, we use the patches that received energy from the sources like new sources and apply the first shot again (leaving again some energy undistributed, and just the patches with unsent energy greater than a predetermined threshold are used like new sources). This process is repeated until the unsent energy of all the patches is below the threshold or after a predetermined number of steps. We call this new algorithm *extended first shot*.

An important matter is the relationship between the number of local lines to use in the extended first shot and the number of global lines in the second step of the simulation. According to [Sbe97a] the optimal ratio is given by

$$\frac{N_l}{N_g} = \sqrt{\frac{n_{int}(1-R_{ave}^2)(1-R_{ave})}{(\frac{R_{s'}^2(1-R_{ave})}{f} + 2R_{s'}^3))}} \tag{3}$$

where, N_l, are the local lines cast, N_g the global lines, R_{ave} the average reflectivity of the scene, $R_{s'}$ the reflectivity of the "secondary" sources, n_{int} the average number of intersections a random line has with the scene and f the total area of the patches visible from the sources. We have used this distribution of lines for both first shot and extended first shot methods. For the second one, we suppose that after the first shot all (or almost all) the patches will have received energy. According to this supposition the value of the f will be taken equal to one.

It is important to note that this new algorithm is different from the ones in [FP93] and [Shi91]. Here we leave some energy undistributed in the sources, instead of sending the whole of it, thus the global lines, in the global step, are not wasted when hitting them. In [FP93] and [Shi91] all the unshot energy of all secondary sources is sent. We leave instead some undistributed energy (or radiosity), the same for all the patches and for all the steps in the extended first shot simulation. This undistributed energy (the threshold used in our algorithm) corresponds to the average radiosity of the scene, given by

$$\frac{R_{ave} \cdot \Phi_T}{A_T \cdot (1-R_{ave})} \tag{4}$$

where R_{ave} is the average reflectivity of the scene, Φ_T is the total power of the original sources and A_T is the total area of the scene.

The pseudocode of the extended first shot algorithm is as follows:

```
/* RADIOSITY array stores the radiosity of the patches  */
/* EMITTANCE array stores the emittance of the original sources */
RADIOSITY= EMITTANCE
/* UNSHOT_RADIOSITY array stores the unshot radiosity of the patches */
UNSHOT_RADIOSITY= EMITTANCE
/* The threshold is given by the average radiosity of the scene */
Threshold= (ReflectanceScene*TotalPowerOfOriginalSources)/
           (TotalArea*(1-ReflectanceScene))

MaximumNumberOfShots= NumberOfShotsDefinedByUser
for j=0 to MaximumNumberOfShots
  for i=0 to NumberOfPatches
    /* Only the patches that have unshot radiosity greater than the
       threshold are used like sources */
    if (UNSHOT_RADIOSITY[i] > Threshold)
```

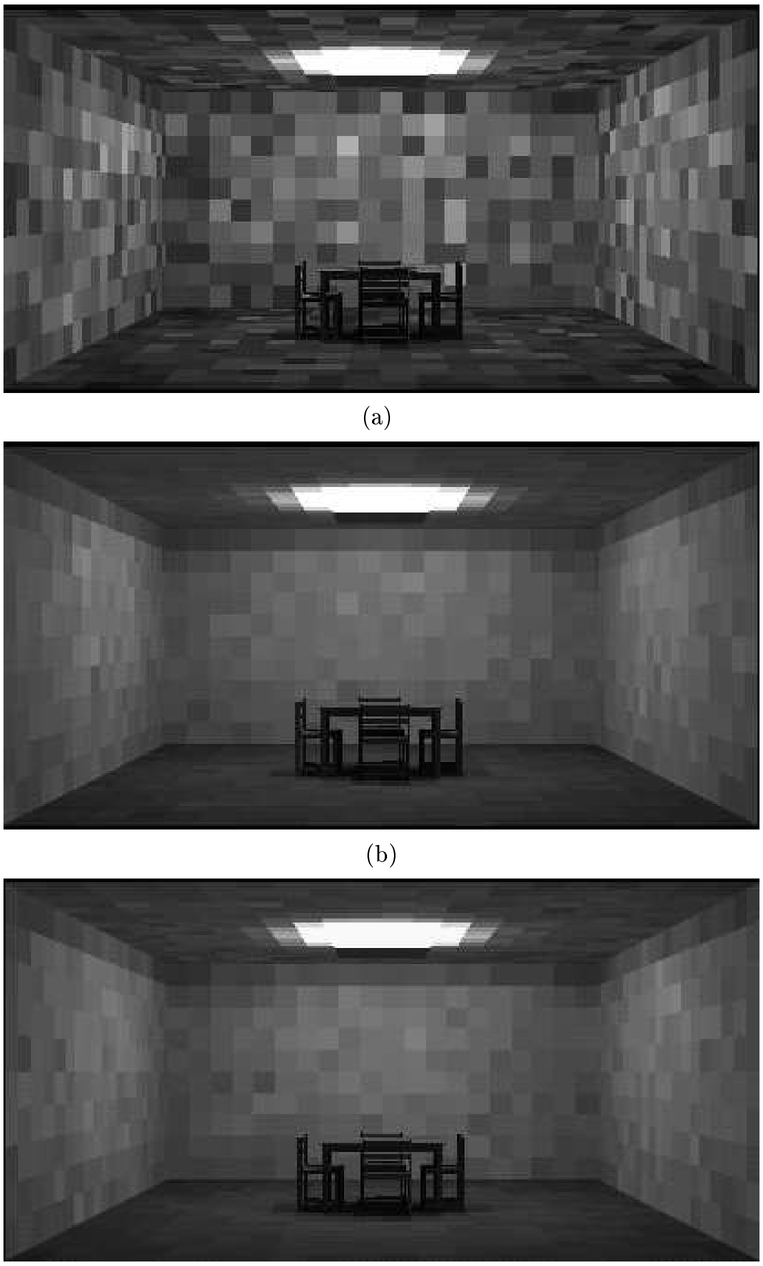
```
      /* The PowerByLine was defined by te user */
      /* AREA array stores the area of the patches*/
      /* Compute the number of lines send from the source */
      numberOfLines= ( AREA[i] * (UNSHOT_RADIOSITY[i] - Threshold))/PowerByLine

      for k=0 to numberOfLines
        /* find the first intersected patch by the line */
        hit= first intersected patch
        /* REFLECTANCE array stores reflectance of the patches */
        emittedRadiosity= PowerByLine/AREA[i]
        receivedRadiosity= (REFLECTANCE[hit] * PowerByLine)/AREA[hit]
        /* Actualize the RADIOSITY array */
        RADIOSITY[hit]=  RADIOSITY[hit] + receivedRadiosity
        /* Actualize the UNSHOT_ENERGY array */
        UNSHOT_RADIOSITY[hit]= UNSHOT_RADIOSITY[hit] + receivedRadiosity
        UNSHOT_RADIOSITY[i]= UNSHOT_RADIOSITY[i] - emittedRadiosity
      endfor
    endif
  endfor
endfor
```

4.2 Results

In this section we compare the performance of our extended first shot algorithm against the first shot and Φ_T local Monte Carlo algorithms [Sbe97a].

In figure 7 we compare the three methods with a reference solution (generated with a Φ_T local Monte Carlo method [Sbert97] with 16 million rays). All the presented results are the average of 6 executions. If figure 7 we show the results of the three algorithm for a scene where the illumination of the source is in the direction of the ceiling. All three images have the same amount of lines but the two multi-path methods have a different distribution of lines (see equation 1). For the first shot method, taking $R_{ave} = 0.577$, $R_{s'} = 0.7$, $f = 0.073$, $n_{int} = 2.1$. Thus $\frac{N_l}{N_g} = 0.41$, thus the number of global lines is more than twice the number of local lines. For the extended first shot method, taking $R_{ave}0.577$, $R_{s'} = 0.577$, $n_{int} = 2.1$ and $f = 1.0$ the result of $\frac{N_l}{N_g}$ distribution is 1.06. With $f = 1.0$ we suppose that after the extended first shot all the patches have energy (it doesn't matter the amount). In figure 7 we see that the extended first shot and the local Monte Carlo methods are better than the first shot method and in figure 8 we just compare (using the same values as in the previous figure) the extended first shot method against the local Monte Carlo method. Images for three executions each for a different method are presented in figure 6.

In figure 9 we compare the three methods for a scene equal to the one in figure 6 but where the illumination of the source is in direction of the floor. The multi-path methods have also here a different distribution of lines (see equation 3). For the first shot method, taking $R_{ave} = 0.577$, $R_{s'} = 0.65$, $f = 0.67$, $n_{int} = 2.1$. Thus $\frac{N_l}{N_g} = 0.85$, it means 20% more global lines than local lines. For the extended first shot method, taking $R_{ave} = 0.577$, $R_{s'} = 0.577$, $n_{int} = 2.1$ and $f = 1.0$ the result of $\frac{N_l}{N_g}$ distribution is 1.06, which is the same as in the first scene (that is, our heuristic for the extended first shot does not difference for the moment between the two cases). On this figure we see that the multi-path methods are better than the local Monte Carlo method (as already demonstrated in [Sbe97b] for the first shot case) and the extended first shot method does not has any noticeable improvement over the first shot method.



(a)

(b)

(c)

Figure 6: *Images generated with the first shot (a), extended first shot (b) and local Monte Carlo (c) methods with 1.5 million lines*

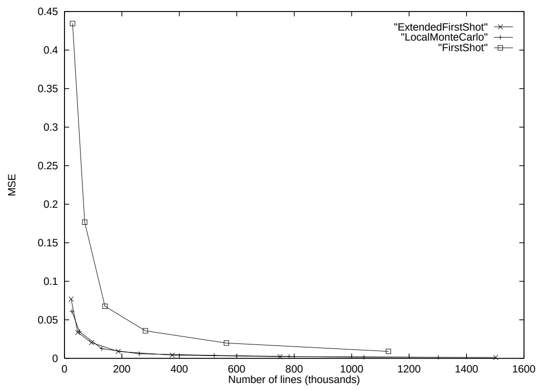


Figure 7: *MSE obtained by the first shot, extended first shot and and local Monte Carlo methods*

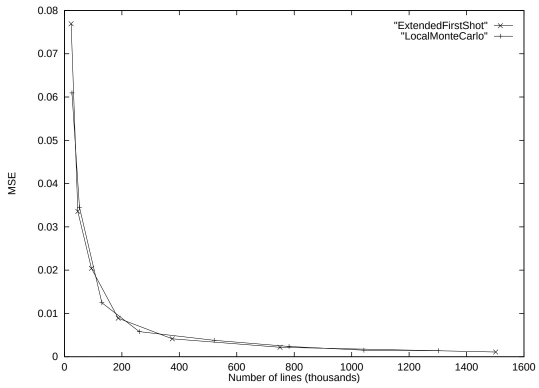


Figure 8: *MSE obtained by the extended first shot and local Monte Carlo methods (same values that the figure 7)*

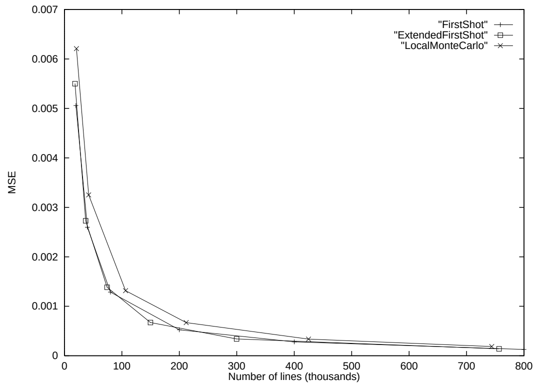


Figure 9: *MSE obtained by the first shot, extended first shot and and local Monte Carlo methods*

5 Conclusions and future work

We have presented here two kind of improvements for the multi-path technique for radiosity. The first one is the use of quasi-random sequences to generate the lines. We have shown that to avoid correlation on global lines we need to use four dimensional sequences. For almost all sequences used we obtained a better result than with pure Monte Carlo. However, some sequences showed not very good asymptotical behaviour, and in other cases, mainly the Halton ones, the best asymptotical behaviour was not maintained for the same sequence when switching between the two different scenes tested. This deserves further investigation. The second improvement is the overcoming of the bad behaviour of the multi-path algorithm when dealing with scenes where only a small portion of it is visible from the sources. This has been done by extending the so called first shot, to further smooth the undistributed radiosity all over the scene. The result obtained for this kind of scenes is of similar quality to the one obtained with a pure local random walk method (we used here for comparison the best one, as seen in [Sbe97a]). We added also a comparison with a well-behaved scene. For this scene we have shown that the extended first shot does not add any significative improvement. Further research will be directed to improve the heuristic used to obtain the ratio between local and global lines, and to investigate the relative efficiencies for more complex scenes.

6 Acknowledgements

Many thanks to Ignacio Martín for his collaboration, and also thanks to Philippe Beckaert for his advices. This project has been funded in part with a Spanish-Hungarian Joint Action from the Spanish Ministerio de Asuntos Exteriores and grant number TIC 95-0614-C03-03 from the Spanish Governement.

References

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

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

[NNB97] L. Neumann, A. Neumann, and P. Bekaert. Radiosity with well distributed ray sets. *Proceedings of Eurographics 97*, pages 261–269, 1997.

[OA97] R. Ohbuchi and M. Aono. Quasi-monte carlo rendering with adaptive sampling. *IBM Graphics*, 1997.

[Pre94] W.H. Press. Numerical recipes in c. *Cambridge University Press*, 1994.

[San76] L. Santaló. Integral geometry and geometric probability ed. addison-wesley, new york. 1976.

[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.

[Shi91] P. Shirley. Discrepancy as a quality measure for sample distributions. *Proceedings of Eurographics 91*, pages 183–194, 1991.

[SKCP98] L. Szirmay-Kalos, B. Csebfalvi, and W. Purgathofer. Importance driven quasi-random walk solution fo the rendering equation. *Proceedings of WSCG 98*, pages 378–385, 1998.

[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.

[SKFP98] L. Szirmay-Kalos, T. Foris, and W. Purgathofer. Quasi-monte carlo global light tracing with infinite number of rays. *Proceedings of WSCG 98*, pages 386–393, 1998.

[SPNP96] M. Sbert, X. Pueyo, L. Neumann, and W Purgathofer. Global multi-path monte carlo algorithms for radiosity. *The Visual Computer*, 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.