A new approach of density estimation for global illumination

Fabien Lavignotte, Mathias Paulin
IRIT – Université Paul Sabatier
Toulouse, France

Abstract

This paper presents a new approach to generate view-independent global illumination solutions using kernel density estimation. Kernel density estimation allows smooth reconstruction of the radiance from hit points generated by shooting random walk or photon tracing. The advantage of this method is that an unbiased Monte-Carlo algorithm simulates light transport and that light reconstruction introduces error but this error is controllable and purely local. We present an approach that does not require storing the set of hit points generated by photon tracing contrary to previous implementation. A method to reduce error both from the light transport and the light reconstruction is also presented.

Keywords: *Global illumination, density estimation, Monte Carlo.*

1. INTRODUCTION

Kernel density estimation is a technique used in statistics and data analysis to construct a smooth estimate of a density function from observed data [5]. This method has been adapted to the global illumination problem. The observed data are the hit points of photons that trace random walks originating from the luminaries. The density of these hit points is proportional to the irradiance. So, a smooth estimate of the irradiance can be found using density estimation.

Density estimation was first used in global illumination to construct caustic maps in a multi-pass method [2][3]. Then, it has been used to construct a very coarse estimate of radiance known as photon map in distributed ray tracing [4]. Density estimation has also been used more directly to construct an estimation of the irradiance [5][7][10]. The two main difficulties of density estimation are highlighted in these latter papers: boundary bias and bandwidth selection. We will go into these problems more in depth in the next section.

Compared to other algorithms, density estimation has several advantages. The error on an estimate of a function can be separated into variance and bias. Pure Monte-Carlo techniques generate unbiased but very noisy estimates. Density estimation allows the introduction of some bias in order to smooth the estimate. This is known as the variance-bias trade-off of density estimation because reducing the bias increase the noise. So, we can choose between noise and bias.

Furthermore, the error in density estimation is purely local compared to finite element method. Indeed, error in finite element methods comes from the projection of the function to be estimated on a basis of functions. But, this error is propagated because of the interaction between all basis functions of the elements. This global error is very hard to control and measure. In density estimation, the error is purely local. The photon tracing phase that represents light transport is unbiased and is separated from the lighting reconstruction that introduces the error.

In this paper, we present a new approach to kernel density estimation. Previous implementations of density estimation for view-independent global illumination required to store the hit points to estimate the radiance at each vertices of the surfaces of the scene. We present a technique that avoids the separate estimation phase and that efficiently estimates the irradiance directly in the photon tracing phase. We also present how this new technique can easily eliminate boundary bias. Another consequence of our method is to reduce the variance of small surfaces. Indeed, photon tracing, as a shooting random walk method, generates variance inversely proportional to the area of the surfaces [1]. So, solutions with a small number of initial photons can be computed with our method without creating disturbing visual artefacts on small features of the scene.

2. DENSITY ESTIMATION FOR GLOBAL ILLUMINATION

2.1 Principles

The kernel density estimation constructs an estimate $\hat{f}$ of an unknown density function f from a set of observed data points $X_1,\ldots,X_n$:

$$\hat{f}(x) = \frac{1}{nh}\sum_{i=1}^{n} K\left(\frac{x - X_i}{h}\right) \quad (1)$$

with h the bandwidth or smoothing parameter and K a kernel function [6].

This method has been adapted to the global illumination problem [8]. The set of observed data points is built from photon tracing. Photon tracing simulates the light flow in the scene. First, a photon is generated from a light source according to its emitted radiance distribution. The photon generated is then traced in the environment until it reaches an object. The photon is ever absorbed or reflected according to the bi-directional reflectance distribution function or BRDF of the object. If reflected, the photon is then traced again from its hit point in a new direction according to the BRDF of the object.

So, the probability $p(x)$ that a photon hits an object at a given point x is:

$$p(x)=\frac{1}{\phi n}L_i(x) \quad (2)$$

with n the total number of photons hit, ϕ the amount of power carried by each photon, and $L_i(x)$ the incident irradiance at this point.

Combining (1) and (2), $L_i(x)$ can be estimated by:

$$\hat{L}_i(x)=\frac{\phi}{h}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h}\right)$$

So, we can reconstruct the irradiance at any point from the hit points $X_1,\ldots,X_n$ generated by the photon tracing.

2.2 Choice of the dimension

The incident irradiance $L_i(x)$ is a 3-dimensional function, so naively the kernel density estimation should be made in three dimensions. The problem is that the irradiance is only defined at the surface of the scene, so considering the problem in three dimensions is not the correct approach.

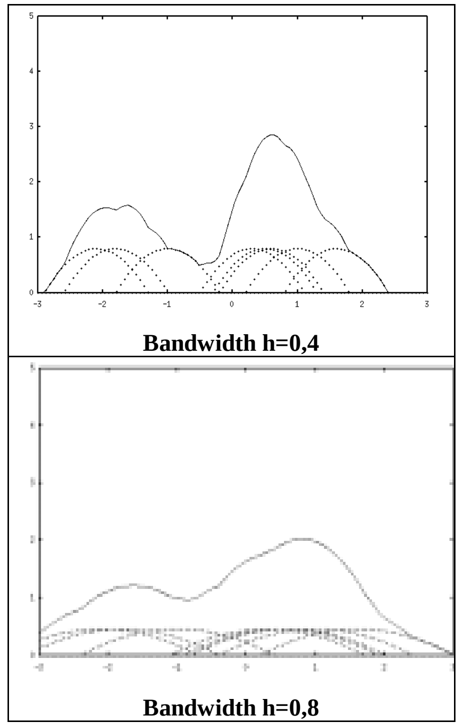
Instead of estimating the incident radiance $L_i(x)$ with $x\in R^3$, we can estimate the incident radiance $L_i^s(u)$ of each surface s in the scene with $u\in R^2$. Then, $L_i^s(u)$ is estimated with the subset $X_1^s,\ldots,X_{ns}^s$ of photons that hit the surface s. To use this subset in the estimation, we must find a transformation $\rho^s : R^3 \rightarrow R^2$, to transform the photons hit points in two-dimension. ρ^s must minimize metric distortion in order to obtain correct result. It is trivial for planar surfaces and much work has been done on this problem for general triangular meshes, but it's still an area of active research.

Jensen [4] overcomes this problem using a mixed solution. The hit points for the estimation are selected in three dimensions but the kernel has a two-dimension support. This solution used directly can lead to strong visual artefacts at boundaries surfaces but in his method density estimation is used only after a gathering step as a coarse approximation.

Volevich [10] used the same technique as Jensen, but he selects more carefully the hit points for the estimation using a heuristic.

2.3 Bandwidth selection

The accuracy of the estimation depends on the bandwidth and on the number of observed data points. If the bandwidth is too high with respect to the number of observed data points, the resulting estimate will be over-smooth, if the bandwidth is too low, the estimate will be too noisy. More formally, we can find in [9] that the integrated square bias is proportional to h^4, but the integrated variance is proportional to $(nh)^{-1}$.



Bandwidth h=0,4

Bandwidth h=0,8

Figure 1:Bandwidth selection

Choosing the optimal bandwidth is a difficult task. The figure above illustrates the importance of a good bandwidth selector. Hi-tech bandwidth selectors have been described in [8], but they are very expensive.

2.4 Local kernel density estimation

Walter [8] and Myszkowsky [5] use variable bandwidth, adaptively chosen for any point of estimation to decrease the error. This is known in statistics literature as local kernel density estimator:

$$\hat{f}(x)=\frac{1}{nh(x)}\sum_{i=1}^{n}K\left(\frac{x-X_i}{h(x)}\right)$$

with $h(x)$ a function that gives a different bandwidth for each point x.

To obtain an optimal bandwidth at point x, Walter looked for the value of h that minimize both the variance and the bias. The variance was estimated from the points data set and the bias was estimated with a difference of two estimators. This technique requires many additional kernel evaluations, twice the original number in Walter's view.

Another technique based on local kernel density estimator is to use the k-th nearest neighbour method, as reported by Jensen [4]. The bandwidth at $h(x)$ is chosen as the distance from x to its k-th nearest neighbour. This method is popular because there are a lot of fast algorithms for finding nearest neighbours in computer science. However, it was proven in [9], that the optimal $h(x)$ is $\left\{f(x)/f''(x)^2\right\}^{1/5}$, but in the case of the nearest neighbour method, $h(x)$ is essentially equivalent to $1/f(x)$, so it's not a very good choice in many cases.

2.5 Boundary bias

The kernel density estimator assumes that the domain of definition of the density function is unbounded. In the global illumination settings, we deal with bounded surfaces. If we do not pay attention of the boundary-induced bias, we have strong visual artefacts as illustrated below:

Figure2: Boundary bias in Cornell box

A simple technique to deal with boundary is to reflect the photons along it [9]. This reduces the bias only when the function is locally constant at the boundary.

Walter [8] resolved this problem using the local linear kernel density estimation method. It is complex to implement, but the runtime costs are said to be negligible in his implementation. In fact, local linear kernel density estimation only reduces the bias where the gradient of the function is locally linear.

3. DIRECT ESTIMATION

3.1 Principles

Our goal was to estimate the incident irradiance for a scene represented by a triangular mesh. Triangular mesh is the most common representation of objects, it consists of a set of vertices and a set of connected triangles. An estimation of the irradiance must be computed for each vertex of the triangular mesh. The triangular mesh needs to be tessellated enough to provide a correct representation of the irradiance because at display, irradiance at the vertices is interpolated linearly across the triangles.

The basic idea of our approach is to not store the hit points for the estimation. Instead, we directly evaluate the contribution of a photon after it hits a triangle.

The direct estimation of the irradiance has several obvious advantages:

- No needs for complex and costly structures for storing hit points data set
- No needs of cache mechanism for huge set of hit points data

The main inconvenient of direct estimation is that we cannot use the set of hit points as guidance for choosing the bandwidth contrary to previous approaches. We will present in the next section how this problem can be overcome.

First of all, we will present an efficient algorithm for direct kernel density estimation. The kernel we use for the estimation is the 2D Epanechnikov kernel that is defined by: $K(x)=\begin{cases} 2\pi^{-1}(1-|x|^2) & if\,|x|<1 \\ 0 & otherwise \end{cases}$. From the formula of the kernel density estimation presented in section 2.1 and the formula of the kernel, we can find that the contribution of the hit point x to the vertices v is $\frac{2\phi}{\pi h}(1-\left|\frac{v-x}{h}\right|^2)$ if $(v-x)<h$, 0 otherwise.

This means that only the vertices at a distance from the hit point less than the bandwidth h are affected by the hit point contribution. So, we need an algorithm to quickly find the neighbouring vertices. To do this, we use the edge connectivity of the triangular mesh as illustrated below:

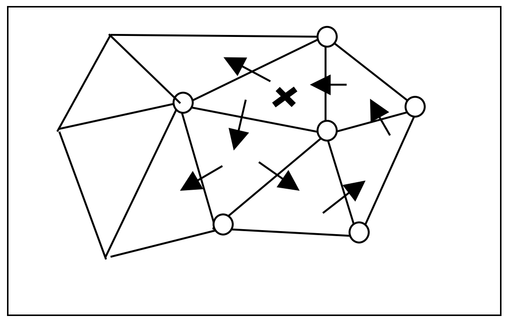


Figure 3: propagation of the contribution

The algorithm of the propagation is then:

addContribution (Point x):
    For each vertex v of the triangle Do
        d =(v –x) / h
        If |d| < 1 Then
        Add $2\phi(\pi h)^{-1}(1-|d|^2)$ *to* $\hat{L}(v)$
        For each edge e of the triangle Do
            addContribution(e,x)

addContribution(Edge e, Point x)
    Let v the vertex opposite to the edge e
    d =(v –x) / h
    If |d| < 1 Then
    Add $2\phi(\pi h)^{-1}(1-|d|^2)$ *to* $\hat{L}(v)$
        Let e' the edge that start from v
        addContribution(e',x)
        Let e" the edge that finish in v
        AddContribution(e",x)

To break cycle during propagation with the edge connectivity, we also need a flag for each vertex to know if the contribution of the hit point x has not already been added. For clarity, we have not shown the algorithm with this information.

3.2 Bandwidth selection

The problem of direct estimation is that we must choose a fixed and empirical bandwidth per surface at the start-up of the photon tracing phase. Fortunately, we present a method to select automatically a bandwidth.

First, a set of hit points is generated with a low number of initial photons. The mesh of the scene does not need to be tessellated at this point. Then, the bandwidth for each surface is chosen using least squares cross validation on hit points of this surface [6].

The method consists in searching the bandwidth h that minimize the function:

$$M(h) = \frac{1}{n^2 h}\sum_i \sum_j K^*\left(\frac{X_i - X_j}{h}\right) + \frac{2K(0)}{nh}$$

with $K^*(t) = K^{(2)}(t) - 2K(t)$.

A problem is that some surfaces may not have sufficient hit points to find a suitable bandwidth. So, bandwidth is computed only for surfaces with a sufficient number of hit points. The bandwidth of the other surfaces can then be chosen using an average of neighbouring surfaces.

Then, we adapt the bandwidth for a different number of initial photons. As shown in [9], the optimal bandwidth that minimizes the integrated mean squared error is proportional to $n^{-1/5}$. So, if the optimal bandwidth h_1 has been found for n_1 photons, optimal bandwidth h_2 for n_2 photons is then: $h_2 = \left(\frac{h_1^5 n_1}{n_2}\right)^{1/5}$.

The next phase of the density estimation algorithm is then performed without storing the hit points of the photons. Also, mesh is tessellated until the distance separating neighbouring vertices is less than a specified threshold that is equal to the bandwidth times a constant parameter usually between two and four.

4. REDUCING VISUAL ARTEFACTS

4.1 Extended triangles

A cause of strong visual artefacts with density estimation is boundary bias as explained in §2. Boundary bias is eliminated with a new method. We consider that the surfaces are not bounded for the estimation. That means that photon hit points used for the density estimation can be outside the triangle.

To obtain the hit points outside the triangles, the contour of the triangular mesh is extended by the bandwidth. Extended triangles are built from the extended contour and the basic contour using basic triangulation. Then, extended triangles are added to the photon tracing phase. The difference with normal triangle is that we do not stop the photon when it intersects an extended triangle, we only add its contribution as explained in ¤3.

With this approach the bias at the boundaries is naturally eliminated. The irradiance at the boundaries is not sub estimated because of the lack of hit points outside the border.

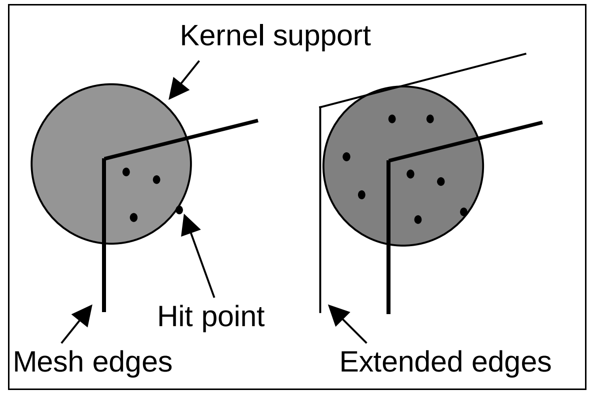


Figure 5: Boundary bias elimination

A direct consequence of this method is also to eliminate the higher variance on small surfaces because the kernel support of the estimation at the vertices of small surfaces has the same size than for large surfaces. That means also that we can use a small amount of photons to estimate the radiance without having disturbing noisy artefacts on small features of the scene.

The principal drawback of this method is that extended triangles add a lot of geometric complexity to photon tracing especially if the bandwidth is high. On the other hand, our method does really eliminate boundary bias contrary to previous approaches that make assumptions on the property of the function at the boundary.

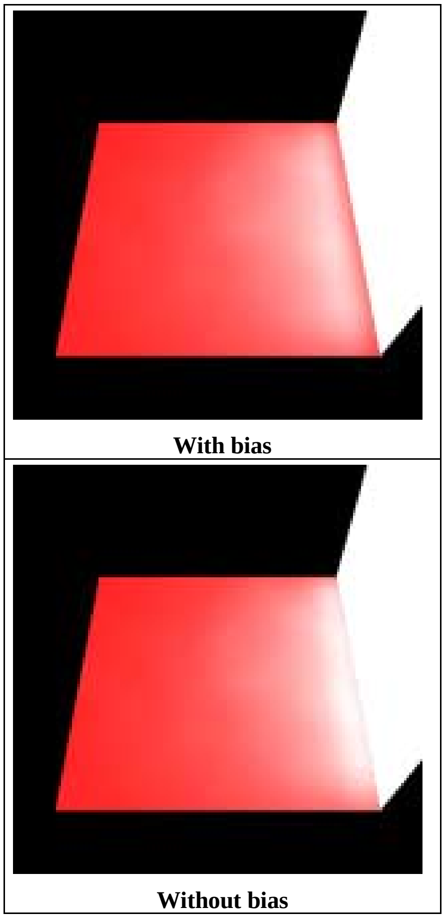
4.2 Bias from discontinuity

The method of extended triangles does not eliminate all visual artefacts. The problem appears at touching surfaces. A touching surface create boundary that causes also bias. We resolve this boundary bias as previously. We consider for the estimation that there is no boundary. That means that for the estimation at a vertex above the touching surface we consider there is no occlusion by the touching surface.

The photons are not stopped after a hit with a surface. But its contribution to the estimation of irradiance at a vertex after the first hit is only added if the vertex is above the first hit plane. So, a photon is stopped only after they are at a distance from the first hit plane greater than the bandwidth because then its contribution is null.

4.3 Bias from singularity

There is still a case of visual artefact with the current method. Indeed, when a photon is traced the probability to hit a point behind the plane of his origin is null. This is also a case of strong visual artefact as illustrated below:



With bias

Without bias

Figure 6: Singularity bias

This is in fact a case of boundary-induced bias where the boundary is the line of intersection between the origin plane of the photon and the kernel support of the vertex. To eliminate this boundary bias, we use a special boundary kernel. We choose the local constant kernel instead of the local linear kernel for efficiency reason because we must compute the boundary kernel for each photon-hit point near the boundary.

5. RESULTS

We have implemented an algorithm based on the approach presented in the previous section. Our implementation has several limitations for the moment in terms of realism:

- All surfaces are perfect diffuse reflectors
- Colour are represented by RGB
- Emitted radiance distribution is constant

A very interesting aspect of our algorithm is to need only one parameter to control the quality of the solution: the number of photons traced from the light sources. We use a tenth the number of photons to find the bandwidth for each surfaces using least squares cross validation. Other parameters control the efficiency of our algorithm. Indeed, efficiency depends of the size of the uniform grid used to accelerate photon tracing and the level of tessellation. The uniform grid has been greatly modified to take into account the various techniques of bias reduction.

In the table below, time are presented for different scenes with hundred of thousand and one million of photons. All the results have been obtained on an Athlon 900 MHz under Linux.

	100 000	1 000 000
Cornell Box	11,8s	76,92s
Room	27,86s	164,54s
Library	140,63s	725,34s

We show in this table how our algorithm scale with the complexity of the scene in terms of memory.

	# of initial triangles	# of tessellated triangles	Memory
Cornell Box	36	11808	11 MB
Room	338	15572	15 MB
Library	18246	26548	31 MB

Then, we compared our algorithm without the correction of bias to show the impact in terms of efficiency of our technique of bias reduction.

	100 000	1 000 000
Cornell Box	7,36s	68,23s
Room	10s	100,39s
Library	26,4s	220,21s

These results show clearly that our bias reduction algorithm has fewer penalties when the number of photons is high. Indeed, in this case, the bandwidth is smaller and many part of the algorithm depends on the size of the bandwidth. Especially, for the Library scene composed of many small triangles, the lack of efficiency of our bias reduction method is explained by the fact that a lot of triangles are extended by the bandwidth so it increase considerably the geometric complexity of the scene.

In the figure below, we compare visually random walk radiosity [1] with our method. Random walk radiosity is ten times faster on this scene than our technique. The artefacts on the photon tracing image are caused by the high variance on small triangles. Our method does not suffer of these artefacts.

Figure 7: Comparison with random walk radiosity

6. CONCLUSION

We have presented a new approach to density estimation for view independent global illumination. Our approach has several advantages compared to previous implementation. It has low memory requirement, and reduces visual artefacts. Also, a coarse estimate can be obtained very quickly with a small number of photons without creating disturbing visual artefact. Furthermore, an important advantage for usability of our method is that only one parameter controls the quality of the solution, the number of photons traced from the light sources.

Much work remains to be done. First, we want to improve the selection of the bandwidth. Least squares cross validation does not always give good result for very complex scene. The problem is that bandwidth selection is still an active area of research in statistics. Then, our method does not scale well in terms of speed with the complexity of the scene as shown in the results. The method must be modified to overcome this problem.

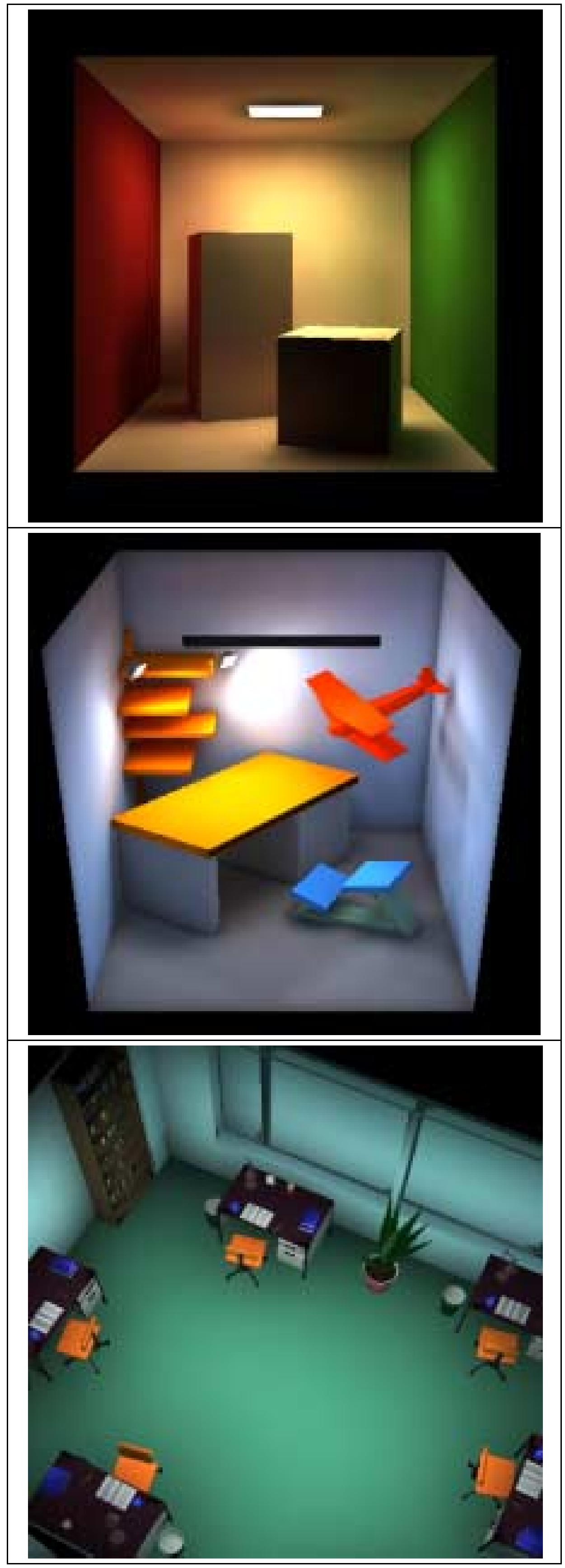


Figure 8: Cornell Box, Room and Library with our method

7. REFERENCES

[1] Philip Bekaert, Hierarchical and Stochastic Algorithms for Radiosity, *PhD Thesis, Katholieke Universiteit Leuven, December 1999.*

[2] Shenchang Eric Chen, Holly Rushmeier, Gavin Miller, and Douglas Turner. A progressive multi-pass method for global illumination. In P*roceedings of Siggraph'91*, July 1991.

[3] Paul S. Heckbert. Adaptative radiosity textures for bidrectionnal ray tracing. In *Proceedings of Siggraph'90,* August 1990.

[4] Henrik Wann Jensen. Global illumination using photon maps. In *Rendering Techniques '96*,1996.

[5] Karol Myszkowski. Lighting reconstruction using fast and adaptative density estimation techniques. In *Rendering techniques '97*, 1997.

[6] B.W. Silverman. Density Estimation for Statistics and Data Analysis. *Chapman and Hall, London and New York,* 1986.

[7] Peter Shirley, Bretton Wade, Philip M. Hubbard, Bruce Walter and Donald P.Greenberg. Global Illumination via density estimation. *In Proceeding of the Sixth Eurographics Workshop on Rendering*, June 1995.

[8] Bruce Walter. Density estimation techniques for global illumination. *PhD Thesis, Cornell University,* August 1998.

[9] M.P. Wand and M.C. Jones. Kernel smoothing. *Chapman and Hall, London and New York*, 1995.

[10]Vladimir Volevich, Karol Myszkowski - "Using the Visual Differences Predictor to Improve Performance of Progressive Global illumination Computation", ACM TOG Vol. 19 No. 1, April 2000, pp.122-161.

About the author

Fabien Lavignotte is a PhD student at IRIT.

Mathias Paulin is an associate professor at IRIT.
E-mail: **lavignot@irit.fr, paulin@irit.fr**