



**César Morais Palomo**

**Interactive image-based rendering for virtual  
view synthesis from depth images**

**DISSERTAÇÃO DE MESTRADO**

Dissertation presented to the Postgraduate Program in Informatics of the Departamento de Informática, PUC-Rio as partial fulfillment of the requirements for the degree of Mestre em Informática

Advisor: Prof. Marcelo Gattass

Rio de Janeiro  
July 2009



**César Moraes Palomo**

## **Interactive image-based rendering for virtual view synthesis from depth images**

Dissertation presented to the Postgraduate Program in Informatics, of the Departamento de Informática do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Mestre. Approved by the following commission:

**Prof. Marcelo Gattass**

Advisor

Departamento de Informática — PUC-Rio

**Prof. Alberto Barbosa Raposo**

Departamento de Informática — PUC-Rio

**Prof. Waldemar Celes Filho**

Departamento de Informática — PUC-Rio

**Prof. Paulo Cezar Pinto Carvalho**

Instituto de Matemática Pura e Aplicada — IMPA

**Prof. José Eugenio Leal**

Coordinator of the Centro Técnico Científico da PUC-Rio

Rio de Janeiro, July 06, 2009

All rights reserved.

### César Morais Palomo

César Morais Palomo graduated from the State University of Campinas with a BSc. in Computer Science in 2005. From 2002 to 2007 he worked as a contractor for several companies as a software engineer, designing and developing enterprise and critical systems for the industry and the Brazilian government. In 2007 he started the graduate program in Computer Science at PUC-Rio as a master candidate. In the same year, he joined the Computer Graphics Technology Group (Tecgraf), where he has been developing computer systems in the fields of computer graphics visualization, computer vision, virtual and augmented reality.

#### Bibliographic data

Palomo, C. M.

Interactive image-based rendering for virtual view synthesis from depth images / César Morais Palomo; advisor: Marcelo Gattass. — 2009.

52 f: il. ; 29,7 cm

Dissertação(Mestrado em Informática) - Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2009.

Inclui bibliografia.

1. Informática— Teses. 2. Renderização Baseada em Imagens. 3. Composição. 4. Programação em placas gráficas. 5. Mapa de profundidade. I. Gattass, M.. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

## Acknowledgments

To my advisor Professor Marcelo Gattass for the support, the invaluable attention and the incentive for the realization of this work.

To Ivana, whose grace and encouragement during difficult times were essential for the completion of this work.

To my mother and father, my brother and sister.

To my nephews Nicolas and Felipe for bringing me so much joy every moment I spend with them.

To my colleagues at TecGraf, who helped me whenever a doubt arised.

To the people of the Computer Science departament for their assistance.

To CNPq and PUC-Rio for the financial support.

## Abstract

Palomo, C. M.; Gattass, M.. **Interactive image-based rendering for virtual view synthesis from depth images**. Rio de Janeiro, 2009. 52p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Image-based modeling and rendering has been a very active research topic as a powerful alternative to traditional geometry-based techniques for image synthesis. In this area, computer vision algorithms are used to process and interpret real-world photos or videos in order to build a model of a scene, while computer graphics techniques use this model to create photorealistic images based on the captured photographs or videos.

The purpose of this work is to investigate rendering techniques capable of delivering visually accurate virtual views of a scene in real-time.

Even though this work is mainly focused on the rendering task, without the reconstruction of the depth map, it implicitly overcomes common errors in depth estimation, yielding virtual views with an acceptable level of realism. Tests with publicly available datasets are also presented to validate our framework and to illustrate some limitations in the IBR general approach.

## Keywords

Image-based rendering. Blending. GPU programming. Depth images.

## Resumo

Palomo, C. M.; Gattass, M.. **Renderização interativa baseada em imagens para síntese de vistas virtuais a partir de imagens com profundidade**. Rio de Janeiro, 2009. 52p. Dissertação de Mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

Modelagem e renderização baseadas em imagem tem sido uma área de pesquisa muito ativa nas últimas décadas, tendo recebido grande atenção como uma alternativa às técnicas tradicionais de síntese de imagens baseadas primariamente em geometria. Nesta área, algoritmos de visão computacional são usados para processar e interpretar fotos ou vídeos do mundo real a fim de construir um modelo representativo de uma cena, ao passo que técnicas de computação gráfica são usadas para tomar proveito desta representação e criar cenas foto-realistas.

O propósito deste trabalho é investigar técnicas de renderização capazes de gerar vistas virtuais de alta qualidade de uma cena, em tempo real. Para garantir a performance interativa do algoritmo, além de aplicar otimizações a métodos de renderização existentes, fazemos uso intenso da GPU para o processamento de geometria e das imagens para gerar as imagens finais.

Apesar do foco deste trabalho ser a renderização, sem reconstruir o mapa de profundidade a partir das fotos, ele implicitamente contorna possíveis problemas na estimativa da profundidade para que as cenas virtuais geradas apresentem um nível aceitável de realismo.

Testes com dados públicos são apresentados para validar o método proposto e para ilustrar deficiências dos métodos de renderização baseados em imagem em geral.

## Palavras-chave

Renderização Baseada em Imagens. Composição. Programação em placas gráficas. Mapa de profundidade.

# Contents

1	Introduction	11
2	Related Work	14
2.1	IBR real applications	14
2.2	Static scenes	16
2.3	Dynamic scenes	18
3	Rendering depth images	22
3.1	Image acquisition process	22
3.2	Depth image representation	24
3.3	3D warping	25
3.4	Artifacts inherent to 3D warping	26
4	Compositing	29
4.1	Virtual camera navigation	29
4.2	Blending algorithm	31
4.3	Limitations	32
5	IBR on the GPU	35
5.1	Conceptual algorithm	35
5.2	Creation of view-dependent geometry	36
5.3	Warping to novel viewpoint and occlusions identification	37
5.4	Compositing on the GPU	38
6	Results	42
6.1	Rendering quality	42
6.2	Limitations	45
6.3	Time-performance analysis	46
6.4	Summary	47
7	Conclusion and Future works	48
	Bibliography	49

## List of Figures

1.1	Skin and hair rendering are examples of challenges to purely geometry-based approaches.	11
1.2	Plenoptic function.	12
2.1	Bullet-time effect shows the necessity of counterbalancing the number of input cameras and quality of rendered images.	15
2.2	IBR goals: establish mapping between representation and image screen, and blend.	16
2.3	Layered Depth Images [30]. Input images (left) used to generate the layered representation of a scene (top right). It allows for reconstruction of views free from disocclusion problems (bottom).	17
2.4	View-dependent texture mapping [11]. Input images are projected onto reconstructed architectural model, and assembled to form a composite rendering. Top two pictures show images projected onto model, lower left shows results of blending those two renderings, and lower right shows final result of blending a total of 12 original images.	18
2.5	Kanade et al's Virtualized Reality geodesic dome [17].	19
2.6	Goldlücke results [13]. The regular triangular mesh causes inaccurate appearance at the vicinity of depth discontinuities.	20
2.7	Camera setup in Zitnick et al [38]. Eight cameras are used to capture 1024x768 images, synchronized with commissioned PtGrey concentrator units.	21
2.8	Rendering results for Zitnick et al [38]: (a) main layer M from one view rendered, with depth discontinuities erased; (b) boundary layer B rendered; (c) main layer M for other view rendered; (d) final blended result.	21
3.1	Imaging process in the pinhole camera model: image formation is a sequence of transformations between coordinate systems. We ignore here radial distortion.	22
3.2	Example of a depth image: color image + dense depth map. Darker pixels in depth map mean greater depth. Courtesy of Zitnick et al [38].	24
3.3	3D warping process. First, 3D mesh generated using depth map from reference camera $C_i$ is unprojected to global coordinate system. Then, mesh is projected into a virtual view using camera $C_{virtual}$ projection matrix.	25
3.4	3D warping result. Depth image (pair of images to the left) from a reference camera is projected into a virtual view (right image) using the described 3D warping process. In this case, virtual camera $C_{virtual}$ was placed slightly to the right of reference camera $C_i$ .	27



3.5	3D warping artifacts due to discontinuity in depth. The continuity assumption in depth map does not hold at objects' boundaries, which causes undesirable artifacts in the form of stretched triangles in warped view. When depth map (left) is warped to new viewpoint, the region to the right of the ballet dancer reveals occluded areas.	27
3.6	3D warping to a new viewpoint, with occluded regions drawn in black thank to our labeling schema.	28
4.1	Example of cameras setup: input cameras arranged along a 1D arc.	30
4.2	Navigation of virtual camera $C_{virtual}$ restricted to the lines linking center of adjacent pair of input cameras.	30
4.3	Virtual camera's parameters can be determined as a linear interpolation of adjacent pair of cameras.	30
4.4	Blending algorithm. Two reference views $C_i$ and $C_{i+1}$ , adjacent to virtual viewpoint $C_{virtual}$ , are used in composition stage.	31
4.5	Cases considered in the compositing process.	33
4.6	Frontier $r$ between regions $A$ and $B$ may become undesirably visible due to the stepped behavior of our compositing algorithm and there exist photometric (e.g. gain) differences in cameras used for capture.	33
4.7	Visible seams between occlusion and blended areas, to the right and to the left of woman's head.	34
5.1	Conceptual algorithm for novel view synthesis.	35
5.2	Vertex-buffer object used for improving performance during mesh generation.	37
5.3	Reference view rendering into a FBO, using vertex and fragment shaders.	38
5.4	Influence of angular distance in reference camera's contribution during compositing.	39
5.5	Penalty for pixels marked occluded based on angular distance.	40
6.1	Sample data from Ballet (left) and Breakdancers(right) sequences: color images on top, and corresponding depth images below [38].	43
6.2	Synthesized images for one frame of ballet sequence. Left and middle columns respectively correspond to cameras' 5 and 4 warping results, and right column is the final result. First row: occlusion areas not identified and rubber sheets appear. Second row: we apply the proposed labeling approach.	43
6.3	Synthesized images for one frame of breakdancers sequence. Left and middle columns respectively correspond to cameras' 3 and 2 warping results, and right column is the final result. First row: occlusion areas not identified and rubber sheets appear. Second row: we apply the proposed labeling approach.	43
6.4	Virtual camera positioned coincidently with camera's 5 position in ballet sequence, and cameras 5 and 6 used as reference cameras. Left column: original image. Middle column: synthetic image. Right column: differences between real and synthetic images. Negated for ease of printing: equal pixels in white.	44

- |      |   |    |
|------|---|----|
| 6.5  | Virtual camera positioned coincidently with camera's 5 position in ballet sequence, and cameras 5 and 6 used as reference cameras. Left column: original image. Middle column: synthetic image. Right column: differences between real and synthetic images. Negated for ease of printing: equal pixels in white. | 44 |
| 6.6  | Close-up of rendering result for frame 48 in ballet sequence. Left column: original photo from camera 7. Middle column: estimated depth map. Right column: visible seams below dancer caused by wrong depth estimates.  | 45 |
| 6.7  | Close-up of rendering result for frame 88 in breakdancers sequence. Left column: original photo from camera 7. Middle column: estimated depth map. Right column: visible seams close to dancer's foot caused by wrong depth estimates.  | 45 |
| 6.8  | Close-ups of rendering artifacts (shadowing) for frame 48 in ballet sequence and frame 88 in breakdancers sequence.   | 46 |
| 6.9  | Close-ups of rendering artifacts (due to matting absence) for frame 48 in ballet sequence and frame 88 in breakdancers sequence.  | 46 |
| 6.10 | Render time X input images' number of pixels.   | 47 |