



Siniša Kolarić

**Towards direct spatial manipulation of virtual
3D objects using visual tracking and gesture
recognition of unmarked hands**

MSc Thesis

Thesis presented to the post-graduate program in Computer Science of the Department of Computer Science, PUC-Rio as partial fulfillment of the requirements for the degree of Master in Computer Science.

Adviser : Prof. Marcelo Gattass
Co-Adviser: Prof. Alberto Barbosa Raposo

Rio de Janeiro
March 2008



Siniša Kolarić

**Towards direct spatial manipulation of virtual
3D objects using visual tracking and gesture
recognition of unmarked hands**

Thesis presented to the post-graduate program in Computer Science of the Department of Computer Science, PUC-Rio as partial fulfillment of the requirements for the degree of Master in Computer Science. Approved by the following commission:

Prof. Marcelo Gattass

Adviser

Department of Computer Science — PUC-Rio

Prof. Alberto Barbosa Raposo

Co-Adviser

Department of Computer Science — PUC-Rio

Prof. Simone D. J. Barbosa

Department of Computer Science — PUC-Rio

Prof. Paulo Cezar P. de Carvalho

National Institute for Pure and Applied Mathematics (IMPA)

Prof. Waldemar Celes

Department of Computer Science — PUC-Rio

Prof. José Eugenio Leal

Head of the Science and Engineering Center — PUC-Rio

Rio de Janeiro, March 28, 2008

All rights reserved. Partial or full reproduction of this work without prior authorization by the university, the author or the adviser is prohibited.

Siniša Kolarić

Siniša Kolarić received his BSc degree in mathematics with a minor in computer science from the University of Zagreb, Croatia. He also concurrently studied theoretical physics for three years at the same university. Later on he worked in academia and industry for Croatian, USA and German organizations. Since 2006 he has been a graduate student at PUC-Rio and a researcher at Tecgraf/PUC-Rio. His scientific interests include computer-aided design, computational geometry and topology, solid modeling, 3D user interfaces and real-time interactive rendering.

Bibliographic data

Kolarić, Siniša

Towards direct spatial manipulation of virtual 3D objects using visual tracking and gesture recognition of unmarked hands / Siniša Kolarić; adviser: Marcelo Gattass; co-adviser: Alberto Barbosa Raposo. — 2008.

120 f.: il. (col.) ; 30 cm

Dissertação (Mestrado em Informática) — Pontifícia Universidade Católica do Rio de Janeiro, Rio de Janeiro, 2008. Inclui bibliografia.

1. Informática — Teses. 2. Manipulação direta espacial de objetos virtuais 3D. 3. Realidade aumentada. 4. Realidade mista. 5. Dispositivos de entrada 3D. 6. Técnicas de interação 3D. 7. Visão por computador. 8. Detecção de mãos. 9. Rastreamento de mãos. 10. Reconhecimento de gestos manuais. I. Gattass, Marcelo. II. Raposo, Alberto Barbosa. III. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. IV. Título.

CDD: 004

Katarina Kolarić née Bubek (1948–2007)

In memory of my mother.

Acknowledgments

I would foremostly like to thank my wife Ana Lúcia who has patiently allowed me to study at the expense of household chores, holidays, parent visits and many other little things. Although he is currently too little to understand, I would also like to thank our Raul for moments of pure joy (and sometimes much needed moments of distraction) that just a happy three-year old can provide.

Big thanks to my colleagues Pablo Carneiro Elias and Thiago de Almeida Bastos who initially helped me to make my ways around PUC-Rio and amortize the culture shock which I experienced when I arrived at the campus. They made my stay here much more easier than if I had tried to discover everything all by myself. Thanks especially for including me into your study group during the first year of my graduate studies at PUC — guys you've really been of much help. I hope that I have helped you at least half as much as you have helped me.

Next I would like to thank Prof. Marcelo Gattass who made me a member of the crew at Tecgraf and this way allowed me to have access to all the resources that just an institute can provide, as well as giving me an academic home. Further, thanks for great classes that actually motivated me to adopt a computer-vision approach to the problem described in this MSc thesis.

I would also like to thank Prof. Alberto Raposo, the coordinator of Virtual Reality group at Tecgraf, for creating a great environment to work, and doing everything possible to provide all the needed resources for this work to take place.

Thanks to all the members of the Virtual Reality group and to all the people Tecgraf is composed of — no wonder that Tecgraf has been doing so well lately — just a group of very talented people can achieve such an impressive string of successes.

Special thanks to Rosane and Cosme at the library of the Department of Computer Science, for their patience and expertise, and who sometimes had to endure my less than stellar record in devolving books on time.

Finally, thanks to PUC-Rio in general — I find this university to be a great place, in a great setting, and a real place of excellence. Being here is being part of something special.

Abstract

Kolarić, Siniša; Gattass, Marcelo; Raposo, Alberto Barbosa. **Towards direct spatial manipulation of virtual 3D objects using visual tracking and gesture recognition of unmarked hands**. Rio de Janeiro, 2008. 120p. MSc Thesis — Department of Computer Science, Pontifical Catholic University of Rio de Janeiro.

The need to perform spatial manipulations (like selection, translation, rotation, and scaling) of virtual 3D objects is common to many types of software applications, including computer-aided design (CAD), computer-aided modeling (CAM) and scientific and engineering visualization applications. In this work, a prototype application for manipulation of 3D virtual objects using free-hand 3D movements of bare (that is, unmarked, uninstrumented) hands, as well as using one-handed and two-handed manipulation gestures, is demonstrated. The user moves his hands in the work volume situated immediately above the desktop, and the system effectively integrates both hands (their centroids) into the virtual environment corresponding to this work volume. The hands are being detected and their posture recognized using the Viola-Jones detection method, and the hand posture recognition thus obtained is then used for switching between manipulation modes. Full 3D tracking of up to two hands is obtained by a combination of 2D "flocks-of-KLT-features" tracking and 3D reconstruction based on stereo triangulation.

Keywords

Direct manipulation of virtual 3D objects. Augmented reality. Mixed reality. 3D input devices. 3D interaction techniques. Computer vision. Hand detection. Hand tracking. Hand gesture recognition.

Resumo

Kolarić, Siniša; Gattass, Marcelo; Raposo, Alberto Barbosa. **Rumo à manipulação direta espacial de objetos virtuais 3D usando rastreamento baseado em visão e no reconhecimento de gestos de mãos sem marcadores.** Rio de Janeiro, 2008. 120p. Dissertação de mestrado — Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A necessidade de executar manipulações espaciais (como seleção, deslocamento, rotação, e escalamento) de objetos virtuais 3D é comum a muitos tipos de aplicações do software, inclusive aplicações de *computer-aided design* (CAD), *computer-aided modeling* (CAM) e aplicações de visualização científica e de engenharia. Neste trabalho é apresentado um protótipo de aplicação para manipulação de objetos virtuais 3D utilizando movimentos livres de mãos e sem o uso de marcadores, podendo-se fazer gestos com uma ou duas mãos. O usuário move as mãos no volume de trabalho situado imediatamente acima da mesa, e o sistema integra ambas as mãos (seus centróides) no ambiente virtual que corresponde a este volume de trabalho. As mãos são detectadas e seus gestos reconhecidos usando o método de detecção de Viola-Jones. Tal reconhecimento de gestos é assim usado para ligar e desligar modalidades da manipulação. O rastreamento 3D de até duas mãos é então obtido por uma combinação de rastreamento 2D chamado "*flocks-of-KLT-features*" e reconstrução 3D baseada em triangulação estéreo.

Palavras-chave

Manipulação direta espacial de objetos virtuais 3D. Realidade aumentada. Realidade mista. Dispositivos de entrada 3D. Técnicas de interação 3D. Visão por computador. Detecção de mãos. Rastreamento de mãos. Reconhecimento de gestos manuais.

Summary

1	Introduction	15
1.1	Historical context	16
1.2	The motivation	17
1.3	The scope covered by this dissertation	18
1.4	The structure of this MSc thesis	19
2	Human hand	20
2.1	Introduction	20
2.2	Human hand anatomy	20
2.3	Human hand modeling	22
3	Hand gestures for manipulation	26
3.1	One-handed and gestures in general	26
3.2	Two-handed gestures	28
3.3	Modeling of hand gestures	28
3.4	Hand gesture recognition	30
4	Interaction techniques for direct 3D manipulation	32
4.1	Selecting virtual 3D objects	32
4.2	Translating virtual 3D objects	33
4.3	Rotating virtual 3D objects	35
4.4	Scaling virtual 3D objects	35
5	Computer vision for hand recognition	37
5.1	Cameras	37
5.2	Digital images	38
5.3	Mono vision	39
5.4	Stereo vision	44
5.5	Color spaces	48
5.6	Human skin modeling	48
5.7	Image features	50
5.8	Hand detection	52
5.9	Hand segmentation	53
5.10	Hand pose estimation	54
5.11	Hand tracking	56
6	Prototype application	58
6.1	Requirements	58
6.2	Constraints, assumptions and restrictions	58
6.3	Hand postures defined	59
6.4	Manipulation operations implemented	61
6.5	Control flow	63
6.6	Hardware and software configuration	70
6.7	Tests and results	70

7	Conclusions and future work	81
7.1	Contributions	81
7.2	Future work	81
	Bibliography	83
A	Timeline of research in manipulation	91
A.1	Pre-1980s	91
A.2	1980s	94
A.3	1990s	95
A.4	2000-2008	106
B	Viola-Jones detection method	109
B.1	Haar-like features	109
B.2	Integral images	112
B.3	AdaBoost-based learning	113
B.4	Cascading strong classifiers	114
C	KLT features	116
D	Hartley-Sturm triangulation method	119

List of figures

2.1	A drawing of a human hand, with joints and bones emphasized	21
2.2	Muscles and tendons of the human hand	22
2.3	3-d.o.f. hand model	23
2.4	Rehg-Kanade 27 d.o.f. hand model	24
2.5	Wu-Huang 27 d.o.f. hand model	25
2.6	33 d.o.f. hand model by Nirei et al	25
3.1	Taxonomy of gestures. In this work mostly the manipulative gestures (see extreme left of the figure) will be considered	26
3.2	Gesture interpretation. The Recognition phase has the hand pose (Model Parameters), the database of all defined gestures (classes of trajectories) and a Grammar (serving to influence the gesture recognition depending on the current working context) as input parameters	30
4.1	Go-go technique extends the hand non-linearly for $\vec{r} > \vec{r}_r$	34
5.1	Pinhole camera, with pinhole (i.e. optical center) at \vec{C}	38
5.2	Pinhole camera with screen in the front of \vec{C}	38
5.3	A digital image consisting of 48×43 pixels	39
5.4	Coordinate systems in the world-camera-projection-raster image chain.	40
5.5	Going from 3D world to 3D camera coordinates ($CCS \longleftrightarrow WCS$)	41
5.6	Stereo rig. If the two cameras take a snapshot at the same instant, the two photos make a stereo pair of photos.	45
5.7	Triangulation. Knowing 3D positions of optical centers \vec{C}, \vec{C}' , focal lengths f, f' and 2D positions \vec{u}, \vec{u}' , we can determine 3D position \vec{X} using various triangulation methods (for example, <i>mid-point</i> and <i>polynomial</i> triangulation methods).	46
5.8	Mid-point triangulation method, which finds the point \vec{X} as the point that lies at the minimum distance to both rays: first ray from C through \vec{u} , and the second ray from C' through \vec{u}' .	47
5.9	Tracking an object by tracking its features	50
5.10	Taxonomy of hand pose estimation approaches	54
6.1	The user's workplace	59
6.2	Three hand postures utilized by the system: HAND_POSTURE_OPEN (left), HAND_POSTURE_POINTING (middle) and HAND_POSTURE_FIST (right)	60
6.3	OP_TRANSLATE operation, based on one HAND_POSTURE_FIST posture	62
6.4	The two-handed OP_ROTATE operation is based on two HAND_POSTURE_POINTING postures. An example of a CCW rotation shown	63
6.5	The two-handed OP_SCALE operation is based on two HAND_POSTURE_FIST postures	64

6.6	Detailed activity diagram for detection, tracking and posture recognition	69
6.7	3D plot of estimated hand positions, obtained by tracing a line, a circle and an “eight” in the workspace	71
6.8	A hit (left) and a hit and false hit (right). Posture HAND_POSTURE_OPEN	72
6.9	A hit and multiple false hits (left), and a miss (right). Posture HAND_POSTURE_OPEN	72
6.10	The application upon startup. No hand has been detected yet, therefore hands are not being tracked, thus no static gesture is being recognized, thus no manipulation operation is being performed	76
6.11	Application started to track hands, after both of them assumed posture HAND_POSTURE_OPEN. We can see that the application placed two flocks of KLT features on both hands	76
6.12	The right hand assumed posture HAND_POSTURE_POINTING, therefore the application started performing the operation OP_SELECT using the right hand	77
6.13	The right hand assumed posture HAND_POSTURE_FIST, therefore the application started performing the operation OP_TRANSLATE using the right hand	77
6.14	Both hands assumed posture HAND_POSTURE_OPEN, upon which the previous manipulation operation has been cancelled	78
6.15	The left hand assumed posture HAND_POSTURE_POINTING, therefore the application started performing the operation OP_SELECT using the left hand	78
6.16	The left hand assumed posture HAND_POSTURE_FIST, therefore the application started performing the operation OP_TRANSLATE using the left hand	79
6.17	Both hands assumed posture HAND_POSTURE_FIST, therefore the application started performing the operation OP_SCALE using both hands	79
6.18	Another example of OP_SCALE	80
6.19	Both hands assumed posture HAND_POSTURE_POINTING, therefore the application started performing the operation OP_ROTATE using both hands	80
A.1	Sutherland’s Sketchpad in use (Lincoln TX-2 console, lightpen)	92
A.2	Example of a drawing and calculation made in Sutherland’s Sketchpad: truss load	92
A.3	James H. Clark’s system: 3D-wand (left) and HMD armature (right)	93
A.4	James H. Clark’s system: 3D surface being edited (left) and its grid of control points (right)	93
A.5	“Put-That-There” system by Bolt: manipulating shapes on the wall-sized screen. The user currently points at the circular shape	95
A.6	3-Draw by Sachs <i>et al</i> — based on two 6-d.o.f. sensors and a conventional non-stereo display.	96
A.7	Krueger’s VIDEODESK. Splines are controlled by fingertip positions.	96

A.8	Widgets by Conner et al. Translating a knife along its x axis (a), rotating a knife along an axis (b), and scaling a knife along an axis (c)	97
A.9	Murakami's elastic cube for 3D deformation: schematic (left) and usage (right)	98
A.10	JDCAD: schematic (left) and cone selection technique (right)	99
A.11	Mine's local selection (left) and at-a-distance selection (right)	99
A.12	Deering's HoloSketch: head-tracked stereo glasses and 3D mouse/wand (left) and 3D fade-up menu (right)	100
A.13	CHIMP by Mine: Two-handed mode selection	101
A.14	Mine suggests proprioception as a way to address lack of haptic feedback	102
A.15	Responsive Workbench: stereo video projected on mirrors below the desk (left), and persons observing a 3D house model displayed in stereo (right)	102
A.16	Responsive Workbench: two-handed operation of zooming in	103
A.17	ErgoDesk by Forsberg et al	104
A.18	Some manipulation gestures by Nishino et al	104
A.19	"Surface drawing" by Schkolne et al: modeling a guitar in five steps	105
A.20	"Surface drawing" by Schkolne et al: hand motions create 3D shapes which "float" over the Responsive Workbench	105
A.21	Operation GRAB in Pratin's system	106
A.22	Operation SCALE implemented as opening/closing the fist in Pratin's system	107
A.23	The setup by Bettio et al. The user stands in front of a large stereo display, and manipulates the model using optically tracked hands.	108
B.1	Two types of rectangles used in the extended Viola-Jones method: 1) upright rectangle, and 2) rectangle inclined at 45° . We compute the sum of all gray-level intensities in rectangle r using function $\text{sum}(r)$.	110
B.2	Fourteen feature prototypes (templates) used in the extended Viola-Jones method	111
B.3	Example: computing a 6×2 -pixel "line feature" (see Figure B.2, feature (a) in the second row) whose top left corner is located at pixel $(5, 3)$	112
B.4	The value of pixel (x, y) of the integral image I_f is equal to the sum of all pixels left and up from (x, y) in image I	112
B.5	Cascade of strong classifiers using Haar-like features	115
C.1	Illustration of tracking based on KLT features. Window W is the current window, for example a rectangle of 10×10 pixels. J_W is the restriction of I on the current window W . I_W is the restriction of I on the previous window. What is being searched for, is the displacement vector \vec{d} , which enables us to position window W correctly in the current image.	117

List of tables

6.1	Training sets for HAND_POSTURE_OPEN	73
6.2	Detector performance for HAND_POSTURE_OPEN	73
6.3	Training sets for HAND_POSTURE_POINTING	74
6.4	Detector performance for HAND_POSTURE_POINTING	74
6.5	Training sets for HAND_POSTURE_FIST	75
6.6	Detector performance for HAND_POSTURE_FIST	75

People don't understand 3D. They experience it.

Ivan E. Sutherland, *American computer scientist*