



Edirlei Everson Soares de Lima

Video-Based Interactive Storytelling

TESE DE DOUTORADO

DEPARTAMENTO DE INFORMÁTICA
Programa de Pós-Graduação em Informática

Rio de Janeiro
August 2014



Edirlei Everson Soares de Lima

Video-Based Interactive Storytelling

TESE DE DOUTORADO

Thesis presented to the Programa de Pós-Graduação em
Informática of the Departamento de Informática, PUC-Rio as
partial fulfillment of the requirements for the degree of Doutor
em Informática

Advisor: Prof. Bruno Feijó

Rio de Janeiro
August 2014



Edirlei Everson Soares de Lima

Video-Based Interactive Storytelling

Thesis presented to the Programa de Pós-Graduação em Informática, of the Departamento de Informática do Centro Técnico Científico da PUC-Rio, as partial fulfillment of the requirements for the degree of Doutor.

Prof. Bruno Feijó

Advisor

Departamento de Informática - PUC-Rio

Prof. Simone Diniz Junqueira Barbosa

Departamento de Informática - PUC-Rio

Prof. Helio Côrtes Vieira Lopes

Departamento de Informática - PUC-Rio

Prof. Angelo Ernani Maia Ciarlini

UNIRIO

Prof. Sean Wolfgang Matsui Siqueira

UNIRIO

Prof. José Eugenio Leal

Coordinator of the Centro Técnico Científico da PUC-Rio

Rio de Janeiro, August 4th, 2014

All rights reserved. No part of this thesis may be reproduced in any form or by any means without prior written permission of the University, the author and the advisor.

Edirlei Everson Soares de Lima

Graduated in Computer Science at Universidade do Contestado (2008), and received his Master Degree in Computer Science from Universidade Federal de Santa Maria (2010). He joined the Doctorate program at PUC-Rio in 2010, researching on interactive storytelling, games, artificial intelligence and computer graphics. In 2011 and 2012, his research on video-based interactive storytelling received two honorable mentions from the International Telecommunication Union (ITU).

Bibliographic data

Lima, Edirlei Everson Soares de

Video-Based Interactive Storytelling / Edirlei Everson Soares de Lima ; Advisor: Bruno Feijó – 2014.

218 f. : il. (color.) ; 30 cm

Tese (doutorado) – Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Informática, 2014.

Inclui bibliografia

1. Informática – Teses. 2. Storytelling Interativo. 3. Dramatização Baseada em Vídeo. 4. Composição de Vídeo. 5. Cinematografia Virtual. I. Feijó, Bruno. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Informática. III. Título.

CDD: 004

Acknowledgments

First and foremost, I would like to express my deepest and sincerest gratitude to my advisor, Prof. Bruno Feijó, for his constant guidance, support and incentive during all these years of research. Besides being an outstanding advisor, he has been like a father and a very good friend to me. My appreciation for his continuous support in all the aspects of my research is immeasurable.

This thesis would also not have been possible without the expert guidance and support of many of my unofficial advisors. First, I would also like to thank Prof. Antonio Furtado for constantly guiding me with his extensive knowledge and enthusiasm. My grateful thanks also to Prof. Cesar Pozzer for his constant guidance and insightful discussions. Special thanks to Prof. Simone Barbosa for her precious advices on Human-Computer Interaction. And I would also like to thank Prof. Angelo Ciarlini for his support and collaboration on many of my research works.

I would also like to thank all the people that have been involved in the production of the prototype video-based interactive narratives. Special thanks to Marcelo Feijó for his excellent work in writing the scripts and directing the production of the interactive narratives; and to Bruno Riodi for his great work in editing and preparing the video material.

I would also like to thank CAPES (Coordination for the Improvement of Higher Education Personnel, linked to the Ministry of Education) and CNPq (National Council for Scientific and Technological Development, linked to the Ministry of Science, Technology, and Innovation) for the financial support for this research. Special thanks to the staff of the Department of Informatics (PUC-RIO) and to the ICAD/VisionLab for providing an excellent research environment.

Finally, I would also like to thank my parents for their constant support and encouragement.

Abstract

Lima, Edirlei Everson Soares de. **Video-Based Interactive Storytelling**. Rio de Janeiro, 2014. 218p. DSc Thesis - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

The generation of engaging visual representations for interactive storytelling represents a key challenge for the evolution and popularization of interactive narratives. Usually, interactive storytelling systems adopt computer graphics to represent the virtual story worlds, which facilitates the dynamic generation of visual content. Although animation is a powerful storytelling medium, live-action films still attract more attention from the general public. In addition, despite the recent progress in graphics rendering and the wide-scale acceptance of 3D animation in films, the visual quality of video is still far superior to that of real-time generated computer graphics. In the present thesis, we propose a new approach to create more engaging interactive narratives, denominated “Video-Based Interactive Storytelling”, where characters and virtual environments are replaced by real actors and settings, without losing the logical structure of the narrative. This work presents a general model for interactive storytelling systems that are based on video, including the authorial aspects of the production phases, and the technical aspects of the algorithms responsible for the real-time generation of interactive narratives using video compositing techniques.

Keywords

Interactive Storytelling; Video-Based Dramatization; Video Compositing; Virtual Cinematography.

Resumo

Lima, Edirlei Everson Soares de. **Storytelling Interativo Baseado em Vídeo**. Rio de Janeiro, 2014. 218p. Tese de Doutorado - Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro.

A geração de representações visuais envolventes para storytelling interativo é um dos desafios-chave para a evolução e popularização das narrativas interativas. Usualmente, sistemas de storytelling interativo utilizam computação gráfica para representar os mundos virtuais das histórias, o que facilita a geração dinâmica de conteúdos visuais. Embora animação tridimensional seja um poderoso meio para contar histórias, filmes com atores reais continuam atraindo mais atenção do público em geral. Além disso, apesar dos recentes progressos em renderização gráfica e da ampla aceitação de animação 3D em filmes, a qualidade visual do vídeo continua sendo muito superior aos gráficos gerados computacionalmente em tempo real. Na presente tese propomos uma nova abordagem para criar narrativas interativas mais envolventes, denominada “Storytelling Interativo Baseado em Vídeo”, onde os personagens e ambientes virtuais são substituídos por atores e cenários reais, sem perder a estrutura lógica da narrativa. Este trabalho apresenta um modelo geral para sistemas de storytelling interativo baseados em vídeo, incluindo os aspectos autorais das fases de produção e os aspectos técnicos dos algoritmos responsáveis pela geração em tempo real de narrativas interativas usando técnicas de composição de vídeo.

Palavras-chave

Storytelling Interativo; Dramatização Baseada em Vídeo; Composição de Vídeo; Cinematografia Virtual.

Contents

1 Introduction	16
1.1. Objectives	19
1.2. Contributions	19
1.3. Thesis Structure	20
2 Interactive Storytelling	22
2.1. Story Dramatization	27
2.1.1. Text-Based Dramatization	27
2.1.2. 2D/3D Dramatization	30
2.1.3. Video-Based Dramatization	35
2.1.4. Other Forms of Dramatization	37
2.2. Logtell	39
2.2.1. Story Generation	41
2.2.2. User Interaction	43
2.2.3. Dramatization	44
2.3. Conclusion	46
3 Cinematography	48
3.1. Shot	49
3.2. Camera Movements	51
3.3. Continuity	52
3.4. Filming Methods	54
3.5. Editing	54
3.6. Matting and Compositing	56
3.7. Light and Color	59
3.8. Music	60
3.9. Film Crew	61
3.10. Conclusion	62
4 Video-Based Interactive Storytelling	63
4.1. Related Work	64

4.2. System Requirements	69
4.3. Operating Cycle and System Modules	70
4.3.1. Story Generation	73
4.3.2. User Interaction	75
4.3.3. Story Dramatization	76
4.4. Architecture	77
4.4.1. Story Generator	77
4.4.2. User Interaction	79
4.4.3. Story Dramatization	80
4.5. Conclusion	82
5 Interactive Film Production	83
5.1. Pre-production	83
5.1.1. Story Definition	84
5.1.1.1. Static Schema	84
5.1.1.2. Dynamic Schema	88
5.1.1.3. Behavioral Schema	93
5.1.1.4. Detailed Schema	94
5.1.2. Shooting Script Generation	96
5.2. Production	99
5.2.1. Set Construction and Camera Setup	100
5.2.2. Shooting Actions	105
5.2.3. Shooting Locations	107
5.2.4. Shooting Static Scenes	109
5.3. Post-Production	110
5.3.1. Editing Actions	110
5.3.2. Editing Locations	114
5.3.3. Actor Definition	117
5.3.4. Location Definition	119
5.3.5. Static Scenes Definition	121
5.3.6. Narrative Resource Pack	121
5.4. Conclusion	122
6 Video-Based Dramatization System	124

6.1. Methods and Libraries	124
6.1.1. Image and Video Processing	124
6.1.2. Artificial Neural Networks	125
6.1.3. Emotions and Relations Network	126
6.2. Cinematography Agents	128
6.2.1. Film Directing	129
6.2.1.1. Scriptwriter	130
6.2.1.2. Director	132
6.2.1.3. Actors	134
6.2.2. Film Compositing	137
6.2.2.1. Placing Actors and Establishing the Line of Action	139
6.2.2.2. Camera Placement and Definition	141
6.2.2.3. Video Editing	147
6.2.2.4. Color and Lighting Effects	155
6.2.2.5. Frame Compositing	157
6.2.3. Film Scoring	161
6.2.3.1. The Music Director Agent	162
6.3. Conclusion	163
7 User Interaction	164
7.1. Natural Language Processing	165
7.2. Interaction Mechanisms	171
7.2.1. Social Interaction	172
7.2.1.1. Interaction by Comments	174
7.2.1.2. Interaction by Preferences	174
7.2.1.3. Interaction by Poll	175
7.2.2. Mobile Interaction	176
7.3. Conclusion	177
8 Application and Evaluation	179
8.1. Technical Evaluation	181
8.1.1. Video Editing	181
8.1.1.1. Shot Selection	181
8.1.1.2. Transition Selection	183

8.1.2. Photography and Music	186
8.1.3. Frame Compositing	187
8.1.4. Natural Language Interface	188
8.2. Visual Evaluation	190
9 Conclusion	192
9.1. Concluding Remarks	192
9.2. Contributions	194
9.3. Publications and Awards	197
9.4. Limitations and Directions for Future Research	198
References	201

List of Figures

Figure 2.1: Story graph of The Mystery of Chimney Rock.	24
Figure 2.2: Example of story generated by Tale-Spin.	27
Figure 2.3: Example of story generated by Universe.	29
Figure 2.4: Example of story generated by Ministrel.	30
Figure 2.5: Graphical interactive storytelling systems.	33
Figure 2.6: Other graphical interactive storytelling systems.	34
Figure 2.7: Video-based interactive storytelling systems.	36
Figure 2.8: Interactive storytelling systems that explore other forms of story dramatization.	38
Figure 2.9: Overview of the story generation process.	40
Figure 2.10: Logtell architecture.	41
Figure 2.11: Example of automaton representing the possibilities for the dramatization of a kidnap event.	42
Figure 2.12: User interface for continuous interaction.	44
Figure 2.13: Scenes from the 3D dramatization module of the Logtell system.	45
Figure 3.1: The structure of a film.	49
Figure 3.2: Shot types.	51
Figure 3.3: The compositing process using the chroma key technique.	58
Figure 3.4: Example of scene created using the chroma key and matte painting techniques.	58
Figure 3.5: Example of scene created using the chroma key techniques.	59
Figure 4.1: System modules.	71
Figure 4.2: Overview of the story generation process.	72
Figure 4.3: Activity diagram of the proposed system.	74
Figure 4.4: The new architecture of the story generator server of Logtell.	78
Figure 4.5: Multimodal interaction architecture.	79
Figure 4.6: The architecture of the user interaction server.	80
Figure 4.7: The architecture of the video-based story dramatization system.	82

Figure 5.1: Entity-Relationship diagram of the static scheme.	87
Figure 5.2: Example of generic operator that has three different specializations.	91
Figure 5.3: Example of composite operator that has four sub-operators.	92
Figure 5.4: Example of automaton describing the dramatization of the event follow(CH_1 , CH_2).	95
Figure 5.5: Structure of the GEXF file that describes the nondeterministic automata of the narrative.	96
Figure 5.6: Example of a two column shooting script.	97
Figure 5.7: Segment of a shooting script automatically generated by the system.	98
Figure 5.8: Full circle filming setup.	101
Figure 5.9: Semicircle filming setup.	102
Figure 5.10: One-quarter filming setup.	103
Figure 5.11: Single camera film setup.	104
Figure 5.12: In-place walking action being performed over a treadmill.	106
Figure 5.13: Camera placement for filming locations.	108
Figure 5.14: Location with 2 layers (L_1 and L_2).	109
Figure 5.15: The master scene structure.	110
Figure 5.16: Example of alpha mask extracted from the green screen video.	111
Figure 5.17: Adobe After Effects user interface.	112
Figure 5.18: The loop detector tool.	113
Figure 5.19: Example of results of the action editing phase.	114
Figure 5.20: Location with a foreground layer defined by an image and its respective alpha mask.	115
Figure 5.21: Location with 5 waypoints (W_1 , W_2 , W_3 , W_4 and W_5). The front line F_1 and the far line F_2 delimit the region for waypoints.	116
Figure 5.22: The interactive tool for waypoint placement.	117
Figure 5.23: Structure of the XML file that describes the actors of the narrative.	118
Figure 5.24: Structure of the XML file that describes the locations of the narrative.	120

Figure 5.25: Structure of the XML file that describes the static scenes of the narrative.	121
Figure 6.1: Emotions and Relations Network.	126
Figure 6.2: An overview of the video compositing process.	129
Figure 6.3: Flowchart of the directing process.	130
Figure 6.3: Template of a behavior class inherited from the class BehaviorBase.	135
Figure 6.4: Scene of a character walking from W_2 to W_1 .	136
Figure 6.5: Flowchart of the compositing process.	138
Figure 6.6: Example of scene using the proposed method to calculate the size of the actor.	140
Figure 6.7: Examples of line of action.	142
Figure 6.8: Triangle Systems.	143
Figure 6.9: Full Triangle System.	143
Figure 6.10: Examples of shots that can be simulated using the video material available in a scene of a dialog between two characters.	144
Figure 6.11: Window system.	146
Figure 6.12: Neural network system.	149
Figure 6.13: Structure of the camera selection neural network for a scene of a dialog between two characters.	150
Figure 6.14: Similarity Scale (the values of α and β are experimental).	153
Figure 6.15: Example of a transition computation.	155
Figure 6.16: Parallel video compositing architecture.	158
Figure 6.17: Pseudocode of the compositing algorithm.	159
Figure 7.1: Flowchart of the global and local interaction processes executed by the Suggestion Manager module.	165
Figure 7.2: Phrase structure tree (a) and the typed dependencies (b) of " <i>The wolf should eat the grandmother!</i> ".	166
Figure 7.3: Example of anaphora problem in the sentence " <i>John saves the grandmother and marries her</i> ".	168
Figure 7.4: Example of negation in the sentence " <i>The wolf should not eat Anne!</i> ".	168
Figure 7.5: Example of omitted subject in the sentence " <i>Kill the wolf!</i> ".	169

Figure 7.6: The process of extracting valid first-order logic sentences. S_x is the input text phrase and P_{xn} is the output list of predicates.	170
Figure 7.7: Example of an introduction message.	172
Figure 7.8: Activity diagram of the social interaction module.	173
Figure 7.9: Example of user comment expressing a suggestion on Facebook.	174
Figure 7.10: Example of user “liking” a system generated suggestion on Facebook.	175
Figure 7.11: Example of poll with story suggestions generated by the system on Facebook.	176
Figure 7.12: Mobile user interface. Image (a) shows the main screen of the mobile application; and image (b) shows the interface during a local interaction.	177
Figure 8.1: Scenes from “The Game of Love”.	179
Figure 8.2: Scenes from “ <i>Modern Little Red Riding Hood</i> ”.	180
Figure 8.3: Recognition rate of the shot selection method with training sets ranging from 10 to 50 samples.	182
Figure 8.4: Example of a transition computation between two shots (C_{79} , C_{80}) of The Lord of the Rings: The Return of the King.	184
Figure 8.5: Recognition rate of the visual effects and music selection method with training sets ranging from 10 to 50 samples.	187
Figure 8.6: Performance results of the parallel composing architecture with the number of actors in the frame ranging from 1 to 4 and with the number of compositing threads ranging from 1 to 8.	188

List of Tables

Table 2.1: List of the main interactive storytelling systems and their respective story generation models and dramatization methods.	46
Table 6.1: List Emotional profiles used by the Director of Photography agent.	157
Table 6.2: List Emotional profiles used by the Music Director agent.	163
Table 8.1: Recognition rate of the shot selection method with training sets ranging from 10 to 50 samples.	182
Table 8.2: Comparison between the original transitions in the Lord of the Rings: The Return of the King with the transitions selected by our method.	184
Table 8.3: Comparison between the original transitions in Psycho with the transitions selected by our method.	185
Table 8.4: Performance results of the transition selection method with different video resolutions.	186
Table 8.5: Recognition rate of the visual effects and music selection method with training sets ranging from 10 to 50 samples.	186
Table 8.6: Description of the selected basic actions used in the visual evaluation test.	190
Table 8.7: Visual comparison between the selected frames of the scenes composed by the human subjects and the corresponding frames automatically generated by the proposed video-based dramatization system for the three basic actions.	191
Table 8.8: Comparison between the times spent by the human professionals and the system to compose the scenes representing the three basic actions.	191