
Advanced Real-Time Manipulation of Video Streams

Jan Herling

Advanced Real-Time Manipulation of Video Streams

Jan Herling
Erfurt, Germany

PhD Thesis, Ilmenau University of Technology, Germany, 2013

ISBN 978-3-658-05809-8

ISBN 978-3-658-05810-4 (eBook)

DOI 10.1007/978-3-658-05810-4

The Deutsche Nationalbibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data are available in the Internet at <http://dnb.d-nb.de>.

Library of Congress Control Number: 2014938061

Springer Vieweg

© Springer Fachmedien Wiesbaden 2014

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law. The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use. While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer Vieweg is a brand of Springer DE.

Springer DE is part of Springer Science+Business Media.

www.springer-vieweg.de

*Never trust a live video transmission -
even if you've manipulated it yourself.*

Abstract

Diminished Reality is a new fascinating technology that removes real-world content from live video streams. This sensational live video manipulation actually removes real objects and generates a coherent video stream in real-time. Viewers cannot detect modified content. Existing approaches are restricted to moving objects and static or almost static cameras and do not allow real-time manipulation of video content. This work presents a new and innovative approach for real-time object removal with arbitrary camera movements.

Two major challenges are presented. A high quality image inpainting method, applicable within a few milliseconds to each frame of the generated video stream is required in addition to a frame-to-frame coherence without any knowledge about future or previous frames. To determine areas to be removed, even from heterogeneous backgrounds, our approach uses a new and powerful real-time capable selection strategy based on fingerprints. Our image inpainting approach itself was inspired by previous layered and randomized approaches. Applying a new and unique initialization strategy as well as a new cost function to minimize coherence deviations based on a combination of spatial and appearance costs, the approach provides high quality results in real-time. Our approach for frame-to-frame coherence applies a homography to remove objects from mostly planar backgrounds, and is applicable even for rotational camera movement around the object to be removed.

Applied to well-known test images, our approach guarantees similar or even better quality compared to that of other state-of-the-art inpainting approaches. In addition, it performs approximately two magnitudes faster. An initial user test revealed that video manipulations based on the approach are barely detectable even if viewers are generally aware of the possibility of changed content. Based on these results, this work opens up a world of new opportunities for interactive and real-time video manipulation especially in the fields of TV and movie production as well as in advertising.

Kurzfassung

Diminished Reality ist eine neue faszinierende Technologie, die es ermöglicht, reale Inhalte aus Live-Kamerabildern zu entfernen. Die Live-Videomanipulation entfernt reale Objekte und erzeugt in Echtzeit einen kohärenten Video-Stream. Dabei können die Betrachter keine Manipulation feststellen. Derzeit existierende Ansätze sind auf sich bewegende Objekte und statische oder weitestgehend statische Kameras beschränkt und erlauben keine Manipulation des Videoinhalts in Echtzeit. In dieser Arbeit wird ein innovativer Ansatz vorgestellt, der das Entfernen von Objekten bei beliebigen Kamerabewegungen ermöglicht.

Dabei gilt es vorrangig, zwei Herausforderungen zu lösen. Zum einen wird eine Bild-Inpainting-Methode benötigt, die innerhalb weniger Millisekunden auf jedes Kamerabild angewandt werden kann und qualitativ hochwertige Ergebnisse liefert. Zum anderen muss ohne die Verwendung vergangener oder zukünftiger Kamerabilder eine Bild-zu-Bild-Kohärenz erzeugt werden. In dieser Arbeit wird eine neuartige und leistungsfähige Selektionsstrategie vorgestellt, die auf sogenannten Fingerabdrücken basiert, damit unerwünschte Inhalte selbst auf heterogenen Hintergründen bestimmt werden können. Der Inpainting-Ansatz wurde ausgehend von bereits bekannten randomisierten Ansätzen und Verfahren mit mehreren Bildebenen entwickelt. Qualitativ hochwertige Ergebnisse können in Echtzeit durch die Anwendung einer einzigartigen Initialisierungsstrategie und einer neuartigen Kostenfunktion, die die Kohärenzabweichung miniert, erzeugt werden. Die Kostenfunktion kombiniert räumliche und erscheinungsbasierende Kosten. Der auf Homographie basierte Ansatz erzeugt eine Bild-zu-Bild-Kohärenz beim Entfernen von Objekten, vor überwiegend ebenen Hintergründen und unterstützt Rotationsbewegungen der Kamera um das zu entfernende Objekt.

Obwohl der Ansatz etwa zwei Zehnerpotenzen schneller ist als aktuelle, vergleichbare Inpainting-Ansätze, zeigt sich, dass er für bekannte Testbilder durchgängig qualitativ gleichwertige oder bessere Ergebnisse erzeugt. Ein erster Test mit Nutzern zeigt, dass Videomanipulationen, die mit diesem Ansatz durchgeführt werden, für Testpersonen kaum zu erkennen sind, sogar wenn diese sich der Manipulationsmöglichkeiten bewusst sind. Basierend auf diesen Ergebnissen eröffnet der in der Arbeit vorgestellte Ansatz eine

Fülle neuer Möglichkeiten für interaktive und echtzeitfähige Manipulationen, die vor allem in TV- und Filmproduktionen sowie im Bereich der Werbung eingesetzt werden können.

Acknowledgements

First and foremost, I would like to thank my doctoral advisor Wolfgang Broll. Wolfgang's guidance and support have made a tremendous impact on the outcome of this work. After serving as my mentor during my training at the Fraunhofer Institute for Applied Information Technology in Sankt Augustin, Wolfgang offered me the chance to pursue my doctorate under his supervision. He continually provided critical feedback and suggestions and helped me to stay on track when it seemed that progress in my research had come to a halt. He always found the right words to motivate me, and was available for questions and support round the clock for a period of several years. The discussions we had were invaluable. I would like to express my deepest gratitude for so many years of cooperation and the many shared laughs and happy moments.

I would also like to express my appreciation to Beat Brüderlin for reviewing this thesis and providing detailed feedback and advice regarding my work. The uncomplicated and direct cooperation and communication allowed me to significantly improve my thesis. His enthusiasm for my research gave me self-confidence and inspired my work.

Further, I would like to extend my thanks to my external reviewer Mark Billingham. I am very thankful for his detailed feedback, which made it possible for me to intensify the quality and impact of my work. In 2009, Mark gave me the opportunity to participate in an exchange program at the HIT Lab New Zealand. During this time, I had the chance to meet engineers researching similar topics. I was able to benefit from the wide variety of their experience and research projects conducted by Mark and his employees.

Advice and constructive recommendations given by Sarah Brüntje were invaluable regarding the mathematical background of this thesis. I also would like to gratefully acknowledge the support of Sandra Pöschel and the helpful suggestions she provided regarding the analysis and evaluation of my work. A big thanks goes to Lisa Czok. Lisa is a native speaker and she helped me in the last phase of my doctoral thesis. She proofread and edited the entire thesis regarding grammar and spelling issues in an incredibly short time so that it could be released within the specified time frame.

Jan Herling

This dissertation would not have been possible without the encouragement and commitment of my parents Elke and Johannes. They allowed me to experience an optimal education and provided unconditional support. They always believed in me and my work.

Thank you so much.

Contents

1	Introduction	1
1.1	Objective	4
1.2	Outline	8
2	Related Work	9
2.1	Static Image Processing	9
2.1.1	Texture Synthesis	10
2.1.2	Image Inpainting	15
2.1.3	Image Composition	25
2.1.4	Image Manipulation	27
2.1.5	Discussion	31
2.2	Video Inpainting	32
2.2.1	Almost Stationary Camera Motion	32
2.2.2	Dynamic Camera Motion	38
2.2.3	Discussion	40
2.3	Mediated Reality	42
2.3.1	Mixed Reality	42
2.3.2	Diminished Reality	42
2.3.3	Discussion	45
3	Concept	47
3.1	Real-Time Image Inpainting	48
3.2	Real-Time Video Inpainting	51
4	Image Inpainting	53
4.1	Mapping Function	53
4.2	Cost Function	55
4.2.1	Spatial Cost	55
4.2.2	Appearance Cost	58
4.3	Iterative Refinement	60
4.4	Initialization	64
4.4.1	Randomized Erosion Filter	65

4.4.2	Contour Initialization	66
4.4.3	Patch Initialization	69
4.4.4	Discussion	79
4.5	Implicit Constraints	82
4.6	Explicit Constraints	84
4.6.1	Area Constraints	85
4.6.2	Structural Constraints	88
4.7	Analysis	93
4.7.1	Convergence	93
4.7.2	Complexity	97
4.8	Implementation Issues	99
4.9	Results	107
4.10	Limitations	118
4.10.1	Pixel-based Inpainting of Homogenous Content	118
4.10.2	Perspective Image Inpainting	120
4.11	Discussion	121
5	Video Inpainting	123
5.1	Object Selection	124
5.2	Object Tracking	130
5.2.1	Heterogeneous Environments	131
5.2.2	Intermediate Environments	132
5.2.3	Homogenous Environments	132
5.2.4	Contour Refinement	133
5.2.5	Discussion	133
5.3	Mapping Propagation	134
5.4	Inpainting Pipeline	135
5.5	Compensation of Ambient Lighting Changes	136
5.6	Extended Appearance Cost	138
5.7	Results	140
5.7.1	Performance Issues	140
5.7.2	User Study	142
5.7.3	Visual Results	157
5.8	Limitations	161
5.8.1	Object Selection and Tracking	161
5.8.2	Video Inpainting	161
5.9	Discussion	164

6	Conclusion	165
6.1	Summary	165
6.2	Future Work	170
6.2.1	Image Inpainting	170
6.2.2	Video Inpainting	172
6.2.3	Fields of Application	173
7	Spatial Cost Convergence	175
7.1	Spatial Cost for Local Mappings	175
7.2	Spatial Cost for Neighbors	179
7.3	Spatial Cost for Non-Neighbors	187
8	Appearance Cost Convergence	189
8.1	Appearance Cost for Local Mappings	189
8.2	Appearance Cost for Neighbors	193
8.3	Appearance Cost for Non-Neighbors	198
A	Appendix	201
A.1	Patents	201
A.2	Additional Patch Initialization Comparisons	201
A.3	Additional Initialization Comparisons	204
A.4	Additional Image Inpainting Results	208
A.5	Additional Video Inpainting Results	216
A.6	Additional Study Results	227
A.6.1	User Ratings	227
A.6.2	Evaluation Video	232
A.7	Performance Measurements	232
	Bibliography	235

List of Tables

2.1	Summary of the introduced image inpainting approaches . . .	26
3.1	Overview of the derived image inpainting approach	51
4.1	Performance values of patch initialization	77
4.2	Performances of individual initialization approaches	81
4.3	Performance comparison for the <i>Elephant</i> image	109
4.4	Performance comparison for the <i>Bungee</i> image	110
4.5	Performance comparison for the <i>Outlook</i> image	111
4.6	Performance comparison for the <i>Blobs</i> image	113
4.7	Performance comparison for the <i>Baby</i> image	115
4.8	Performance of our image inpainting approach	117
5.1	Performance of fingerprint selection	140
5.2	Performance of the video inpainting	141
5.3	The five values the test subjects could select as answer	144
5.4	Distribution of the test subjects	145
5.5	T-test analyzing ratings for individual backgrounds	155
A.1	Accumulated ratings of the test subjects	231

List of Figures

1.1	Pictures of the <i>Basilica St Mary of Health</i> in Venice	2
1.2	Live video manipulation	5
1.3	Inpainting result example for the <i>Train</i> image	7
2.1	Image processing overview	10
2.2	The two steps of texture synthesis by Efros and Leung	12
2.3	Texture synthesis result by Efros and Leung	12
2.4	Texture synthesis by Wei and Levoy	13
2.5	Texture synthesis result by Xu et al.	14
2.6	Texture synthesis by Efros and Freeman	14
2.7	Image Inpainting vs. Image Completion	16
2.8	Image restoration by Bertalmio et al.	17
2.9	Image inpainting result by Drori et al.	18
2.10	Image inpainting scheme by Criminisi et al.	19
2.11	Image inpainting results by Criminisi et al.	21
2.12	Image inpainting scheme by Sun et al.	22
2.13	Gradient-based image inpainting by Shen et al.	23
2.14	Information propagation of PatchMatch by Barnes et al.	28
2.15	Seam carving image resizing by Avidan et al.	29
2.16	Update step of the completion approach of Wexler et al.	37
2.17	Scheme of the active contour tracking approach	40
2.18	Mixed Reality continuum as defined by Milgram	42
2.19	Diminished Reality result of our previous approach	44
4.1	The two cost constraints of the mapping function f	55
4.2	Spatial cost for neighboring mappings	57
4.3	Two symmetric neighborhoods	58
4.4	Scheme of the pyramid refinement	61
4.5	Iterative layer refinement	62
4.6	Comparison of individual spatial weightings	62
4.7	Scheme of the multithreading inpainting realization	64
4.8	Comparison of the standard and randomized erosion filter	66

4.9	Contour mapping initialization scheme	68
4.10	Contour mapping initialization for a real image	68
4.11	Patch similarity determination for patch initialization	71
4.12	Inpainting priority of the patch mapping initialization	73
4.13	Determination of the direction of the inpainting border	74
4.14	Area of interest for randomized searches	75
4.15	Patch initialization of the <i>Pyramid</i> image of Xu et al.	77
4.16	Comparison of individual patch initialization modes	78
4.17	Initialization comparison for the <i>Elephant</i> image	79
4.18	Initialization comparison for the <i>Sign</i> image	80
4.19	Initialization comparison for the <i>Bungee</i> image	81
4.20	Inpainting with individual data formats	83
4.21	Inpainting with additional texture information	84
4.22	Inverse importance map of an area constraint	86
4.23	Inpainting result for the <i>Ruin</i> image	87
4.24	Constraint weighting graph	89
4.25	Line constraint costs	89
4.26	Distance determination for finite lines	90
4.27	Constraint image inpainting example	92
4.28	Appearance neighborhood for a given point $p \in T$	95
4.29	Determination of the appearance cost difference	96
4.30	Performance comparison for a mask with constant size	98
4.31	Performance for a growing mask with constant image size	99
4.32	Convergence of the number of optimized mappings	101
4.33	Convergence of the ratio of optimized mappings	102
4.34	Convergence with weak spatial weighting	102
4.35	Convergence with strong spatial weighting	103
4.36	Comparison of spatial weightings on a fine pyramid layer	104
4.37	Comparison of spatial weightings on a coarse pyramid layer	104
4.38	Cost ratio for individual pyramid layers	105
4.39	Mapping changes for individual inpainting images	106
4.40	Cost ratio for the exact inpainting approach	106
4.41	Visual comparison between exact and fast inpainting	107
4.42	Result comparison for the <i>Elephant</i> image	109
4.43	Result comparison for the <i>Bungee</i> image	110
4.44	Result comparison for the <i>Outlook</i> image	111
4.45	Result comparison for the <i>Blobs</i> image	112
4.46	Result comparison for the <i>Blobs</i> image of Kwok et al.	113
4.47	Result comparison for the <i>Sign</i> image	114
4.48	Result comparison for the <i>Baby</i> image	115

4.49	Inpainting result for the <i>Wall</i> image	116
4.50	Image inpainting with a soft gradient background	119
4.51	Image inpainting with perspective distortion	120
4.52	Inpainting of an individually transformed artificial image . . .	121
5.1	Real-time selection results using fingerprint segmentation . .	126
5.2	Object selection and tracking scheme	128
5.3	Contour tracking scheme	131
5.4	Mapping forwarding by application of a homography	134
5.5	Lighting compensation of the reference model	137
5.6	Comparison of a default and a corrected reference model . .	139
5.7	Age distribution of the test subjects	145
5.8	Test subject rates for evaluation background B_0 and B_1 . .	146
5.9	Averaged ratings for all backgrounds	147
5.10	Test subject ratings for twelve manipulated video sequences .	147
5.11	Comparison of subjects for manipulated videos	149
5.12	Comparison of subjects with and without knowledge	150
5.13	Test subject ratings for all twelve original video sequences .	150
5.14	Comparison of subjects for original videos	151
5.15	Subjects with and without knowledge for original videos . .	152
5.16	Average rating \bar{x} of the test subjects for individual groups .	153
5.17	Comparison of averaged ratings for individual groups	153
5.18	Averaged ratings for nonbriefed subject without knowledge .	154
5.19	Averaged ratings for briefed subject with knowledge	154
5.20	Comparison of the amount of undecidable ratings	156
5.21	Video inpainting with a heterogeneous background	159
5.22	Video inpainting of a coat of arms	160
5.23	Real-time selection results of the fingerprint segmentation .	162
5.24	Video inpainting of a volumetric object	163
7.1	Visualization of the neighborhood $N_s(p)$	176
7.2	Visualization of the subsets T_x and $\overline{T_x}$	178
7.3	Visualization of the rearrangement of the distance measure .	183
7.4	Scheme of the direct and indirect spatial cost	186
8.1	Visualization of the neighborhood $N_a(p, f)$	190
A.1	Patch initialization for the <i>Elephant</i> image.	202
A.2	Patch initialization for the <i>Window</i> image	202
A.3	Patch initialization for the <i>Bungee</i> image.	203

A.4	Patch initialization for the <i>Wood</i> image	203
A.5	Initialization comparison for the <i>Wall</i> image	204
A.6	Initialization comparison for the <i>Biker</i> image	205
A.7	Initialization comparison for the <i>Blobs</i> image	206
A.8	Initialization comparison for the <i>Window</i> image	207
A.9	Initialization comparison for the <i>Wood</i> image	207
A.10	Inpainting result for the <i>Bungee</i> image by Pritch et la.	208
A.11	Inpainting result for the <i>Window</i> image	209
A.12	Inpainting result for the <i>Wood</i> image	210
A.13	Inpainting result for the <i>Universal Studios</i> image	211
A.14	Inpainting result for the <i>Still Life with Apples</i> image	212
A.15	Inpainting result for the <i>Microphone</i> image	213
A.16	Inpainting result for the <i>Dog</i> image	213
A.17	Inpainting result for the <i>Train</i> image	214
A.18	Inpainting result for the <i>Chair</i> image	214
A.19	Inpainting result example with leaves in the background	215
A.20	Constraint image inpainting example	215
A.21	Video inpainting removing a window in a house facade	217
A.22	Video inpainting with ivy plants in the background	219
A.23	Video inpainting with a homogenous background	220
A.24	Video inpainting with a grass background	221
A.25	Video inpainting removing a drain	222
A.26	Video inpainting recovering a straight line	223
A.27	Video inpainting recovering an circular object	224
A.28	Video inpainting with a homogenous background	225
A.29	Video inpainting recovering a volumetric object	226
A.30	Test subject rates for evaluation background B_2 and B_3	228
A.31	Test subject rates for evaluation background B_4 and B_5	228
A.32	Test subject rates for evaluation background B_6 and B_7	229
A.33	Test subject rates for evaluation background B_8 and B_9	229
A.34	Test subject rates for evaluation background B_{10} and B_{11}	230
A.35	The manipulated evaluation video of test background B_2	233
A.36	SPEC 2006 benchmarks	234
A.37	Visualization of the estimated performance increase	234

<http://www.springer.com/978-3-658-05809-8>

Advanced Real-Time Manipulation of Video Streams

Herling, J.

2014, XXIV, 244 p. 150 illus., 20 illus. in color., Softcover

ISBN: 978-3-658-05809-8