COMPUTER ENGINEERING SERIES

3D Video

From Capture to Diffusion

Edited by Laurent Lucas
Céline Loscos and Yannick Remion

iSTE

WILEY

3D Video

From Capture to Diffusion

Edited by
Laurent Lucas
Céline Loscos
Yannick Remion

常州大学图书馆
藏书章

iSTE

WILEY

First published 2013 in Great Britain and the United States by ISTE Ltd and John Wiley & Sons, Inc.

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act 1988, this publication may only be reproduced, stored or transmitted, in any form or by any means, with the prior permission in writing of the publishers, or in the case of reprographic reproduction in accordance with the terms and licenses issued by the CLA. Enquiries concerning reproduction outside these terms should be sent to the publishers at the undermentioned address:

ISTE Ltd
27-37 St George's Road
London SW19 4EU
UK

www.iste.co.uk

John Wiley & Sons, Inc.
111 River Street
Hoboken, NJ 07030
USA

www.wiley.com

© ISTE Ltd 2013
The rights of Laurent Lucas, Céline Loscos and Yannick Remion to be identified as the author of this work have been asserted by them in accordance with the Copyright, Designs and Patents Act 1988.

Library of Congress Control Number: 2013947317

British Library Cataloguing-in-Publication Data
A CIP record for this book is available from the British Library
ISBN: 978-1-84821-507-8

FSC
www.fsc.org
MIX
Paper from
responsible sources
FSC® C013604

Printed and bound in Great Britain by CPI Group (UK) Ltd., Croydon, Surrey CR0 4YY

3D Video

Foreword

The concept of giving 3D sense to flat representations (drawings, paintings, photos and films) has been progressively and deliberately re-examined and considered since the beginning of time. The rock paintings of Altamira (Spain) and Font-de-Gaume (France), for example, provide a fascinating example of the muscular systems of large herbivores. In the Lascaux cave (France), the shape of the rocks has been used to support and even accentuate the painting's form. All ancient art everywhere has, in some way or another, used depth and perspective in its representations, often awkwardly or confused, erroneous, often using more or less shared social codes, but always with the objective of understanding the real world beyond the limits of flat representation.

Formalized understanding of the mechanisms of Quattrocento perspective has largely enabled artists to move away from flat media to new, more accurate methods which have been used widely, often with competing artistic objectives and technical abilities. Complete perspective has therefore become an inseparable part of all pictures to the point of no longer even being a point of discussion: whether boring or shocking, controversial or exposing, it is no longer obvious because it is expected.

The dawn of photography, which by definition respects the canons of perspective, the undoubted problem of traditional representation, allowed artists to move away from this new norm which, over three centuries, had governed real-life representation. Artists can escape the unseen since, for example, space is no longer merely confined to perspective. Braque and Picasso, Klee and Bacon have shown us that this space is not only a matter of geometry but is also richer and holds several mysteries. Beyond perspective, it allows us to see background images and their convergence.

试读结束：需要全本请在线购买： www.ertongbook.com

However, perspective, outside this small artistic field where it has somewhat faded, plays a vital role in our vision, logic and society. Unsurprisingly, the world of photography, as in painting, has quickly sought media which go beyond flat representation. Since the 19th Century, ingenious inventions have provided a third dimension to photography and then, with the dawn of cinema and its younger sibling television, it would not be long until 3D would make an impact, well before the Second World War. Binocular stereovision is the most natural input method for this mode, reliant on various separation means for optical paths, orthogonal polarizations, color decompositions, color flickers through wheels and mirrors and lens networks. Kerr's or Pockel's electromagnetic cells and liquid crystals will be examined later as part of this.

3D has not yet finished developing. Propelled by undeniable economic and social success, it has suffered from a lack of exploration followed by a new found success. The literature is evidence of this and that we are on the brink of a new dawn. However, current technologies are undeniably better than ever. Acquisition, projection, archiving and transmission technologies have come to fruition after long being suspended or in development. It has also been an opportunity and major development for production companies and commercial film distribution organizations, since virtual and augmented reality production has reached previously unseen levels of quality, performance and productivity which are indispensable for ambitious and demanding production sets. The public, expectant and demanding, desires new experiences which can be seen as evidence of the success of these new methods.

Finally, all these factors, which have made this dawn of 3D cinema possible, have played an important role in 3D television because these two fields, cinema and video, have a shared future. The opposing war between them, which has raged for 50 years, to capture an audience seen to favor one over the other, has now disappeared. We only have to think of the success of films on television or the continuation of television series through films. The public is omnivorous, consuming all kinds of images, no longer knowing whether they are from a dark room, a small screen or even a video game. This requires an abundance of pixels, bright, life-like colors and multi-sensory interaction and interactive and 3D animation, particularly when their counterparts exist in real-life but are transformed by video, as discussed in this book.

It is this which allows us to trace the progression of 3D, which has affected the entire chain of production for digital images. This book aims to examine ongoing events and describe their development, with a formal representation of theoretical tools in order to understand the approaches studied. References are provided to allow the reader to further study the developments that these numerous techniques relate to. Another aspect relates to examining all points in the technical chain which today governs 3D television. We will also examine technical tools such as cameras, screens and software. In addition , matching, detection and compression will be studied.

As a complete and complex work, *3D Video* is a welcome to the current efforts and achievements which have accompanied the the emergence of this new addition to our homes, the 3D image.

Henri MAÎTRE
September 2013

Notations

Spaces, sets

$d \in \{2, 3, \ldots\}$	space dimension
$\mathbb{R}^d$	real d-dimensional vector space
$\mathbb{Z}^d$	integer d-dimensional vector space
$][a, b][$	compact interval in $\mathbb{R}$
$\{a, ..., b\}$	discrete interval in $\mathbb{Z}$ or $\mathbb{N}$
$\mathbb{B} = \{false, true\}$	boolean set
$\mathbb{N}_n = \{0, ..., n-1\} \subset \mathbb{N}$	set of n first natural integers
$\mathbb{Z}_{a,b} = \{a, ..., b-1\} \subset \mathbb{Z}$	set of $b - a$ integers connecting from a to $b - 1$

Functional notations

Y^X	set of applications $X \to Y$
$Y[X]$	set of tables or maps on domain X (discrete) with values in Y
$f(b)$	value of the function f in b

$t[b] \equiv t_b$	value of the sample b of the map/table t
$\lfloor x \rfloor$	integer part (the nearest inferior integer) in x
$\lceil x \rceil$	nearest superior integer to x
$\{x\} \equiv x - \lfloor x \rfloor$	fractional part of x
$[x]$	the nearest integer to x: by convention, if $x = k + 0.5$ with $k \in \mathbb{Z}$, $[x] = \begin{Bmatrix} k & \text{if } k >= 0 \\ k+1 & \text{otherwise} \end{Bmatrix}$

Objects

i, j, k, l, m, n	integer numbers
x, y, z	coordinates (integer or real)
t, u, v, λ, μ	real numbers
D, Δ	real lines
P, Π	real planes
$\mathbf{v}, \mathbf{w}$	vectors
$A, B, C, ...$	points in the real affine space $\mathbb{R}^2$ or $\mathbb{R}^3$
AB	bi-point vector ranging from A to B
$\mathbf{M}, \mathbf{A}, \mathbf{B}$	matrices
$\mathbf{R}, \mathbf{T}$	rotation and translation matrices
f, g, h	applications, functions
Φ, Ψ	operators on other sets $\mathbb{R}^d, \mathbb{Z}^d$
ρ, τ	rotation and translation function
G, Γ	graphs
θ, ϕ	angles
ε	threshold

Digital images

$\Omega^{\mathbf{t}} = \prod_{i=1}^{d} \mathbb{N}_{t_i}, \quad \mathbf{t} \in \mathbb{N}^d$	d-dimensional signal indexation domains or spaces with a size of $\mathbf{t} = (t_1, \ldots, t_d) \in \mathbb{N}^d$
$\Omega^d \equiv \Omega^{(nc,nl[,np])} = \mathbb{N}_{nc} \times \mathbb{N}_{nl}[\times \mathbb{N}_{np}]$	for $d = 2[3]$ we will use the numbers of columns, lines and planes by default $(nc, nl[, np])$
$\mathbf{p} = (x, y[, z]) \in \Omega^d$	pixel position ($d = 2$) or voxel position ($d = 3$) within the indexation domain
$\mathcal{C} = \mathbb{X}^{(b)}$ with $\mathbb{X} \in \{\mathbb{N}, \mathbb{Z}, \mathbb{R}\}$	scalar spaces of digital values for sample components. The $b \in \{8, 10, 12, 16, 32, 64, ...\}$ correspond to the number of bits used for coding real values ($\mathbb{R}$) or integer values ($\mathbb{Z}$ or $\mathbb{N}$)
$\mathcal{E} \equiv \mathcal{E}^{(\mathcal{C}_1, \ldots, \mathcal{C}_c)} = \prod_{i=1}^{c} \mathcal{C}_i$ with $\mathcal{C}_i = \mathbb{X}_i^{(b_i)}$	generic vectoral space for samples' digital multi-component value
$\mathcal{S} \in \mathcal{E}[\Omega^d]$	2D or 3D signal (image or volume) table or map of values in $\mathcal{E}$ indexed by Ω^d
$\mathcal{S}[\mathbf{p}] \equiv \mathcal{S}_{\mathbf{p}} \in \mathcal{E}$	the signal sample $\mathcal{S}$ in position $\mathbf{p} \in \Omega^d$

Neighborhoods

$\mathcal{F} = \left\{\mathbf{v} \in \mathbb{Z}^d\right\} \subset \mathbb{Z}^d$	d-dimensional shape, set of vectors $\mathbf{v} \in \mathbb{Z}^d$ coding the positions of each point in the shape in relation to the reference point
$\mathcal{F}_{\mathbf{p}} = \{\mathbf{p} + \mathbf{v} \mid \mathbf{v} \in \mathcal{F}\}$	neighborhood in the shape $\mathcal{F}$ placed in $\mathbf{p} \in \Omega^d$: translated of $\mathcal{F}$ in or by $\mathbf{p}$
$\overline{\mathcal{F}}_{\mathbf{p}} = \mathcal{F}_{\mathbf{p}} \cap \Omega^d = \left\{\mathbf{p} + \mathbf{v} \,\middle\vert\, \begin{matrix} \mathbf{v} \in \mathcal{F} \\ \mathbf{p} + \mathbf{v} \in \Omega^d \end{matrix}\right\}$	neighborhood in the shape $\mathcal{F}$ en $\mathbf{p} \in \Omega^d$, truncated by Ω^d

试读结束：需要全本请在线购买： www.ertongbook.com

Multiviews

A set of N signals (known as views within the context of this book) of the same dimensions d and sizes $(nc \times \; nl \; (\times \; np))$ will both be considered as a table of N signals ($\in \mathcal{E}[\Omega^d][\mathbb{N}_N]$) and as a signal with a superior dimension of $d+1$ ($\in \mathcal{E}[\Omega^d \times \mathbb{N}_N]$).

$\Upsilon^d = \Omega^d \times \mathbb{N}_N$	global indexation space of a set of N views (images or volumes) with a dimension of d
$\mathcal{M} \in \mathcal{E}[\Upsilon^d] \equiv$ $\mathcal{M} \in \mathcal{E}[\Omega^d][\mathbb{N}_N]$	multi-signal, N views indexed by Ω^d with values in $\mathcal{E}$
$\mathcal{M}[i] \equiv \mathcal{M}_i \in \mathcal{E}[\Omega^d]$	digital view number $i \in \mathbb{N}_N$ in $\mathcal{M}$: d-dimensional signal
$\mathbf{r} = (\mathbf{p}, i) \in \Upsilon^d$	multiview sample index: position $\mathbf{p} \in \Omega^d$ and digital image $i \in \mathbb{N}_N$
$\mathcal{M}[\mathbf{r}] \equiv \mathcal{M}_\mathbf{r} \equiv$ $\mathcal{M}[\mathbf{p}, i] \equiv \mathcal{M}_{\mathbf{p},i} \equiv$ $\mathcal{M}[i][\mathbf{p}] \equiv \mathcal{M}[i]_\mathbf{p} \equiv$ $\mathcal{M}_i[\mathbf{p}] \equiv \mathcal{M}_{i_\mathbf{p}}$	different expressions shown as equivalents to reach the value in $\mathcal{E}$ for the sample in position $\mathbf{p}$ in the view i from $\mathcal{M}$. the last (double level indice) should be avoided

Acknowledgments

We are very grateful to those who have contributed to this book through their work and research. We would like to particularly express our gratitude to Henri Maître who has overseen the compilation of this book and has generously given us his help and support. We would also like to show our recognition to ISTE Ltd. and John Wiley & Sons who have greatly assisted as throughout the production of this book. Lastly, we would like to thank all those people and organizations who have allowed us to use their data and/or illustrations within this book.

Several pieces of research data shown in this book have been made possible because of the financial support from the following organizations:

– Regional organizations from three areas: projects including CIA (CPER Nord-Pas De Calais region in France), CREATIS (CPER Champagne-Ardenne region in France) and RUBI3 (Brittany with the Image et Réseaux (Image and Networks) cluster – 2010-13).

– The Agence nationale de la recherche (ANR) (National Agency for Research, France): including projects such as SEMANTIC-3D (AAP RNRT 2002-06), CAM-RELIEF (AAP RIAM 2008-10), FAR3D (AAP CSOSG 2008-2010), COLLAVIZ (AAP COSINUS 2009-12), PERSEE (AAP Blanc 2009-13) and 3D FaceAnalyzer (AAP Blanc Inter. 2011-13).

– Competition clusters and/or Fonds unique interministériel (FUI) (Inter ministerial Funds, France): including the projects FUTURIM@GE (Image et Réseaux (Image and Networks) – 2008-10), Terra Numerica (Cap Digital – 2006-09) and 3DLive (CapDigital, Imaginove, Images et Réseaux – 2009-12).

– The Fond national pour la société numérique (FSN) (National Fund for Digital Society, France): including the RECOVER3D project (2012-2014) for future investments.

– The European Commission: including projects such as 3D Média (FEDER 2008-13), 3D-ConTourNet (ICT COST Action IC1105), HDRi (ICT COST Action IC1005), Ed-cine (FP6 IP), Apidis (FP7 Strep) and JEDI (ITEA2, co financed by DGCIS for the French Ministry of Industry).

We would like to thank these institutions for their past and ongoing support.

Lastly, we would also like to recognize those who, by their patience, understanding and encouragement, have allowed us to bring this project to completion.

Laurent LUCAS, Céline LOSCOS and Yannick REMION
September 2013

Introduction

The extension of visual content to 3D as well as dynamically capturing scenes in 3D to generate an image on a remote site in real time has long been considered merely a part of science fiction. Today they are a reality, collectively referred to by terms such as 3D television (3DTV), free viewpoint TV (FTV) and, more generally, 3D video. This new type of image creates the illusion of a real environment, resulting from continually improving efforts in research and development over a number of years.

Numerous experts believe that 3D represents the future of media, such as television and the Internet, and will in turn improve the quality of visual experiences for the end user. The whole chain of content production must be reconsidered, beginning with recording techniques, since those designed specifically for 3D are far more numerous and varied than those used normally in conventional 2D context. The same can also be said of other aspects, such as, for example:

– the description and representation of scenes according to more or less informative structures, ranging from multiview or multiview-plus-depth videos to 3D digital reconstructed models;

– 3D reconstruction which extracts 3D models in various forms from videos acquired from multiple viewpoints, such as static or animated meshes;

– the compression of representations of scenes created by capture (stereoscopic or multiview videos) or reconstruction (3D models);

Introduction written by Laurent LUCAS, Céline LOSCOS and Yannick REMION.

– 3D display, with or without adaptation/enhancement of content and/or intermediate view synthesis.

The democratization of these technologies needs specifically designed display devices. Stereoscopic or autostereoscopic screens show a heavy tendency toward this while their use for displaying 3D content today still poses a number of problems, showing that all these techniques must yet be perfected to avoid being rejected by the end user due to reasons of poor quality and/or eyestrain.

3D videos therefore cover a multitude of aspects, collectively linking a series of recorded videos to full depth 3D visualizations, potentially using estimations of depth in video sources. The developments examined here are therefore based on methods and tools from highly varied fields, such as applied mathematics, computer imaging, computer graphics, virtual reality, signal processing as well as psychophysics and the psychology of human vision.

In this highly multidisciplinary context, the objective of this book focused on 3D video is twofold since it aims, in addition to summarizing current information about the subject, to provide:

– for students: a solid base enabling readers to carry out activities relating to this topic and to learn the underlying concepts overall;

– for researchers: as complete a reference for this subject as possible which precisely indicates current research and understanding in this field as well as future trends and perspectives.

Its organization into four parts is due to a desire to cover all phases of 3D video by bringing together formal presentations of theoretical tools and developments of more technical or technological aspects. It should be noted that all figures are also available in color at http://www.iste.co.uk/lucas/3D.zip.

The first part of this book runs through the basics of 3D video and the recording of its characterizing multiview videos. This begins with, in Chapter 1, the different fundamental aspects of this technology. Historical and mathematical aspects relating to 3D computer vision and physiology of human vision are thus presented. Chapters 2 and 3 look at technological and methodological problems in relation to capturing images, more specifically in Chapter 3 within a multiview context that characterizes 3D video. The specification of geometric elements relating to the recording and display of

3D media is then examined in Chapter 4. Chapter 5 concludes Part 1 of this book, focusing on the problems of geometric and colorimetric camera calibration.

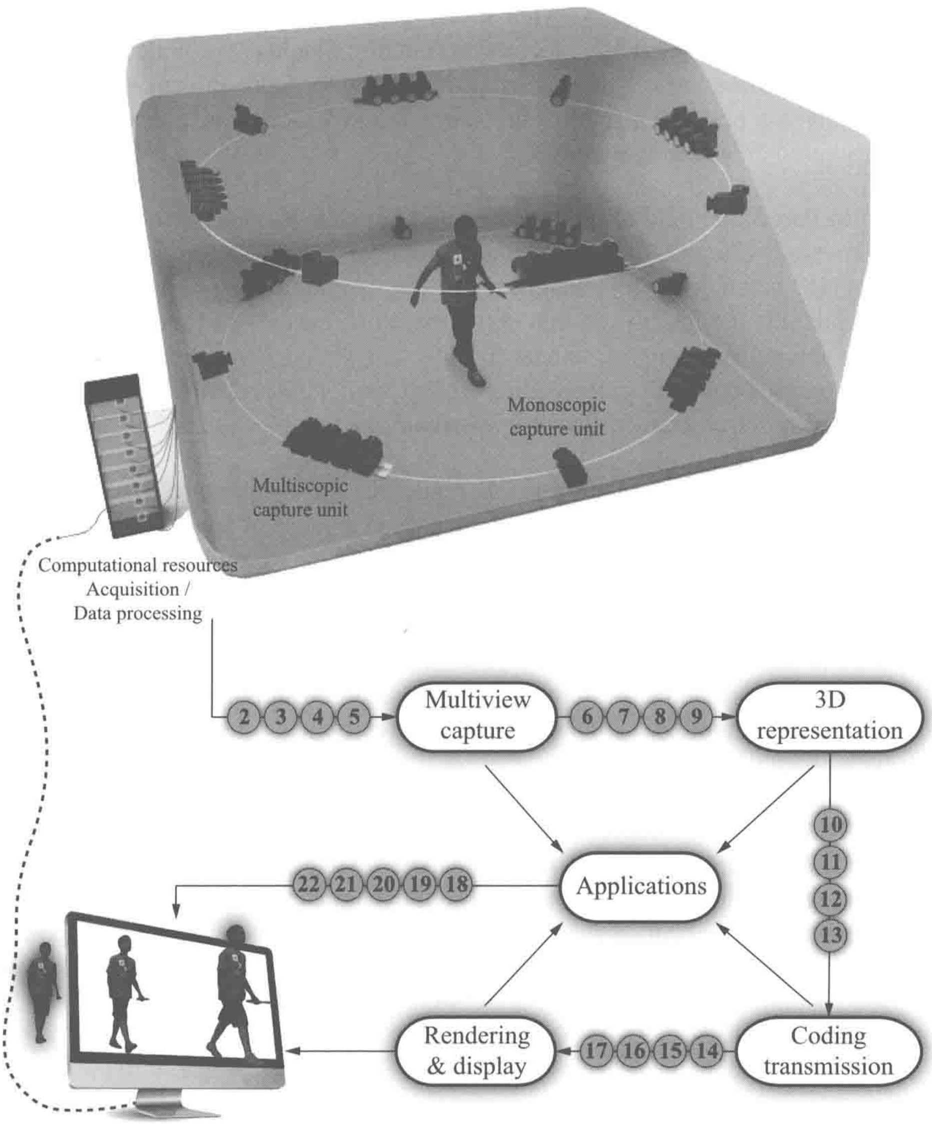

Figure I.1. *Organization of this book: the numbered chips correspond to the different chapters*

试读结束：需要全本请在线购买： www.ertongbook.com