

LOW BIT-RATE FACIAL VIDEO COMPRESSION

by

SHOMA CHATTERJEE

Department of Computer Science and Engineering

Submitted
in fulfillment of the requirements of the degree of Doctor of Philosophy
to the



Indian Institute of Technology, Delhi
June 2001

I. I. T. DELHI.
LIBRARY
Acc. No. TH-2842

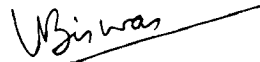
TH
621.397.2:681.3
CHA-L

deo compression

CERTIFICATE

This is to certify that the thesis entitled “Low Bit-Rate Facial Video Compression” submitted by Miss. Shoma Chatterjee in the Department of Computer Science and Engineering, Indian Institute of Technology, New Delhi for the award of Doctor of Philosophy is a record of the original bonafide research work carried out by her under my guidance and supervision.

The thesis and any part thereof has not been submitted to any other university / institution for award of any degree or diploma.



Prof. K.K. Biswas

Deptt. of Computer Sc. & Engg.

Indian Institute of Technology,

New Delhi - 110 016.

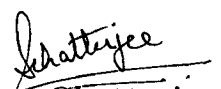
ACKNOWLEDGEMENTS

I wish to express my heartfelt gratitude to my supervisor Prof K.K. Biswas who has been my constant source of encouragement for my work. I am equally thankful to Prof Subhashis Banerjee who was not my formal supervisor, but with whom I have shared many of ideas, and who has been kind enough to spend his valuable time to formulate ideas, critically analyzed the results and offered suggestions to improve my work. I am thankful to the members of my research committee Prof H. M. Gupta, Prof Santanu Chaudhury, Prof P.K. Kalra for spending time to evaluate my work periodically and to offer advice at various stages of my work.

I am also very thankful to Shri K.R. Kaushik, senior technical assistant of Computer Science and Engineering Department, who has guided me in all my lab setups.

I am thankful to the CSIR for providing my financial support for the last year of my Phd work.

June 2001


Shoma Chatterjee

LOW BIT-RATE FACIAL VIDEO COMPRESSION

ABSTRACT

This thesis focuses on compression of facial video sequences to achieve low bit rates. The need for facial video compression arises due to increasing electronic transmission of digital images and facial video sequences over the internet and the growing demand for multimedia applications like on-line classroom lectures and video telephones. Video compression maximizes the amount of information to be transmitted over a given channel bandwidth. Compression of video sequences is also essential for digital transmission in the telephone network, storage of video on digital storage media including conventional storage devices CD-ROM, DAT, tape drives, writable optical drives, as well as telecommunication channels such as ISDNs and local area networks.

Traditional methods of video compression, e.g. MPEG 1-2, H.261, H.263 etc. address the issue of video compression solely in the image domain. In contrast, in the thesis we investigate methods to achieve video compression using techniques of computer vision to estimate 3 dimensional structure of the moving object.

We present an affine structure based global reconstruction scheme in which we utilize the invariant property of affine structure and send the $3D$ affine structure to the decoder at the bootstrapping stage. After that we need to send only the bases parameters which are few in number in order to reconstruct the frame at the decoder.

In our projective view morphing scheme to reconstruct global motion we rectify the image planes to parallel planes before interpolation to minimize the projective distortion in case of non-affine motion.

To reconstruct independent local motion we develop three separate schemes. The first one utilizes $2D$ affine local transformations, the second scheme is snake contour tracker based which develops an on-line database of distinct templates and recognizes current expression and encodes as an index into the database and the third scheme is based on optical flow parameters. Finally we conclude the thesis.

Contents

1	INTRODUCTION	1
1.1	VIDEO COMPRESSION	1
1.2	VIDEO COMPRESSION TECHNIQUES: A SURVEY	3
1.3	CAMERA MODELS AND FEATURE TRACKER	4
1.4	AFFINE STRUCTURE BASED GLOBAL RECONSTRUCTION	5
1.5	VIEW MORPHING BASED FACIAL VIDEO COMPRESSION	6
1.6	LOCAL REGION RECONSTRUCTION	6
2	VIDEO COMPRESSION TECHNIQUES: A SURVEY	8
2.1	INTRODUCTION	9
2.2	SPATIAL REDUNDANCY	10
2.2.1	Transform Coding	11
2.2.2	1-D Transform	11
2.2.3	Determination of Transform Efficiency	14
2.2.4	The Karhunen-Loeve Transform	14
2.2.5	The Discrete Fourier Transform	19

2.2.6	The Walsh - Hadamard Transform	20
2.2.7	The Discrete Cosine Transform (DCT)	21
2.2.8	Wavelet Based Image Coding	23
2.2.9	Fractal Image Coding	25
2.3	CODING REDUNDANCY	26
2.3.1	Theoretical Results from Information Theory	26
2.3.2	The Huffman Coding Algorithm	27
2.3.3	Quantization	28
2.3.4	Vector Quantization	29
2.3.5	Lossless Coding	32
2.4	TEMPORAL REDUNDANCY	33
2.4.1	BTRMEC: Binary Tree Recursive Motion Estimation Coding . .	35
2.4.2	Dynamic 3D Scene Analysis through Synthesis Feedback Control	36
2.4.3	Model Based Facial Image Coding	37
2.4.4	Object-Based Motion Image Compression	38
2.4.5	Physical and Anatomical Models	40
2.4.6	Real-Time Facial Animation	40
2.4.7	Synthesis of Novel Views from a Single Face Image	42
2.4.8	Eigenvector Based Real-Time Structure from Texture	42
2.4.9	Local Parameterized Models of Image Motion	43
2.4.10	Tracking of Facial Images	43

2.4.11	Real Time Tracking and Modeling of Faces	44
2.4.12	Analyzing Facial Expressions for Virtual Conferencing	46
2.4.13	View Morphing	47
2.4.14	Image Metamorphosis with Scattered Feature Constraints	47
2.4.15	Two Dimensional Mesh Based Mosaic Representation	47
2.4.16	Other Schemes	48
2.5	REVIEW OF EXISTING STANDARDS FOR STATIC AND VIDEO COMPRESSION	48
2.5.1	JPEG	48
2.5.2	JPEG 2000	52
2.5.3	MPEG	57
2.5.4	H.261	58
2.5.5	H.263	59
2.5.6	MPEG-4	60
2.6	SUMMARY	63
3	THE CAMERA MODELS AND FEATURE TRACKER	64
3.1	INTRODUCTION	65
3.2	CAMERA MODELS	65
3.2.1	Finite Cameras	66
3.2.2	Principal Point Offset	68
3.2.3	The Projective Camera	71
3.2.4	Cameras at Infinity	73

3.2.5	Orthographic Projection	75
3.2.6	Scaled Orthographic Projection	76
3.2.7	Weak Perspective Projection	76
3.3	PROJECTIVE EPIPOLAR GEOMETRY AND THE FUNDAMEN- TAL MATRIX	77
3.4	AFFINE EPIPOLAR GEOMETRY	82
3.4.1	Relative Motion	84
3.4.2	Affine Epipolar Line	85
3.4.3	The Affine Essential Matrix	88
3.4.4	Affine Structure [79]	89
3.5	SEGMENTATION	91
3.6	AFFINE TRACKING ALGORITHM	93
3.7	PROJECTIVE TRACKING ALGORITHM	97
3.8	RESULTS	99
3.8.1	Results Of Projective Tracking Algorithm	100
3.9	SUMMARY	104
4	AFFINE STRUCTURE BASED GLOBAL RECONSTRUCTION	105
4.1	INTRODUCTION	106
4.2	OVERALL SCHEME	107
4.3	COMPUTATION OF AFFINE STRUCTURE	109
4.4	3D AFFINE RECONSTRUCTION	114

4.4.1	Estimation of the Global Motion Parameters and Global Reconstruction	116
4.4.2	Eigenvalue Compression of Global Motion Parameters	117
4.5	RESULTS OF CASE STUDIES	120
4.5.1	Offline Initialization and Global Reconstruction	120
4.6	CONCLUSIONS	126
5	VIEW MORPHING BASED GLOBAL RECONSTRUCTION	128
5.1	INTRODUCTION	129
5.2	VIEW MORPHING	131
5.2.1	Non-Parallel Views	132
5.2.2	Interpolation in Projective Model	134
5.3	GLOBAL RECONSTRUCTION USING VIEW MORPHING	137
5.3.1	Coder	137
5.3.2	Decoder	143
5.4	RESULTS OF CASE STUDIES	145
5.5	CONCLUSIONS	155
6	LOCAL REGION RECONSTRUCTION	156
6.1	INTRODUCTION	157
6.2	PIECE-WISE 2D AFFINE RECONSTRUCTION OF INDEPENDENTLY MOVING REGIONS	158
6.2.1	Local Reconstruction Scheme	160
6.2.2	Predictive Encoding / Colour Patching	163

- 6.2.3 Results of Case Studies (2D Local Affine Transformation) 164
- 6.2.4 Reconstruction Results after all Three Stages 166
- 6.3 LOCAL RECONSTRUCTION USING SNAKE MODEL 180
 - 6.3.1 “VIEW” Determination 183
 - 6.3.2 Creation of Lip Template Database 185
 - 6.3.3 Snakes: Active Contour Models 186
 - 6.3.4 A Robust Local Tracker using Feedback Analysis and Synthesis 189
 - 6.3.5 Reconstruction of Local Regions 190
 - 6.3.6 Quality of Reconstruction and Bit-Rate 190
- 6.4 OPTICAL FLOW BASED LOCAL RECONSTRUCTION 195
 - 6.4.1 Results of Case Studies 199
- 6.5 CONCLUSIONS 204

7 CONCLUSIONS

207