Learning Non-linear Models of Shape and Motion

A Thesis submitted for the degree of Doctor of Philosophy

By

Richard Bowden

Department of Systems Engineering, Brunel University

October 1999

Abstract

Deformable models have been an active area of research in computer vision for a number of years. Their ability to model non-ridgid objects through the combination of geometry and physics has proven a valuable tool in image processing. More recently a class of deformable objects known as Point Distribution Models or Eigen Models have been introduced. These statistical models of deformation overcome some of the shortfalls of earlier deformable models by learning what is 'allowable' deformation, for an object class, from a training set of examples. This semi-automated learning procedure provides a more generic approach to object recognition, tracking and classification. Their strength lies in their simplicity and speed of operation, allowing the robust ability to model complex deformations in cluttered environments. However, the automated construction of such models leads to a breakdown of the fundamental assumptions upon which they are based. Primarily, that the underlying mathematical model is linear in nature. Furthermore, as more complex objects are considered, these assumptions fail completely and what is produced is an unreliable model.

This work addresses these problems and presents novel techniques for the automated construction and application of non-linear deformable models, which retain the speed, and simplicity of the linear Point Distribution Model. It is further shown how these non-linear models can be augmented with probabilistic temporal constraints, which are essential in object tracking and classification.

This work presents, in essence, three developments to the field. Firstly, a piecewise linear approach to modelling non-linearity is proposed and results demonstrated that show its accuracy in modelling both low and high dimensional datasets with heavy non-linearity. The technique is then extended to the automated construction of models. Secondly, it is shown how the piecewise approach can be augmented with temporal constraints and used in both model prediction, animation and for the support of multiple hypotheses during tracking. It is further shown how these temporal models can be extended to incorporate

information from other sources, providing more reliable tracking in the absence of complete training data. Thirdly, it is shown how elements can be combined statistically and used to infer information about an object from its shape alone. Using human motion capture as an example, it is demonstrated that models can be assembled which allow 3D structural information about body pose and motion to be inferred from a monoscopic image sequence using only natural features of the body as markers.

Acknowledgements

I would like to thank both EPSRC and Foster Findaly Associates Ltd who supported this work financially. Televirtual Ltd for access to the human motion capture data, the Turing Institute for the 3D surface head models and the Centre for Medical Imaging Research at the University of Leeds for access to the MRI datasets of the human hand.

I would like to thank my supervisors Dr Tom Mitchell and Professor Mansoor Sarhadi for having the confidence in my abilities to allow me to pursue the topics which excite me. Tom has been more of a friend than a supervisor during the course of my studies and invaluable as a sounding board for ideas and approaches. Mansoor has been more than generous in both his time and confidence in my abilities and I thank him for the opportunities he has provided me.

I would like to thank my colleagues at Brunel and other institutions for their numerous stimulating conversations and insights into my work, you know who you are. A special word of thanks for Dr Dave Cohen for the many valued discussions related to this work.

Id like to thank my friends and family, especially my parents, for the opportunities they have provided to their children and for teaching us all that our abilities were only limited by our imaginations. For pushing us, understanding us, and helping us to achieve our dreams.

Lastly I would like to thank my wife Beccy for the constant encouragement, proof reading and tolerance of my absence. It is a debt I can never repay. I could not have done it without you.

Declaration

Elements from this manuscript have appeared in, or are about to appear in the following publications.

Non-linear Statistical Models for the 3D Reconstruction of Human Pose and Motion from Monocular Image Sequences. , R. Bowden, T. A. Mitchell, M. Sarhadi, To appear in Image and Vision Computing.

Non-linear Point Distribution Models, R. Bowden, In CVonline: On-Line Compendium of Computer Vision [Online]. R. Fisher (ed). Section 11.3.1.2 , Oct 98.

Reconstructing 3D Pose and Motion from a Single Camera View, R. Bowden, T. A. Mitchell, M. Sarhadi, In Proc. BMVC, John N. Carter & Mark S. Nixon Eds, Uni of Southampton, Vol 2, pp , Southampton, Sept 1998.

Cluster Based non-linear Principal Component Analysis, R. Bowden, T. A. Mitchell, M. Sahardi, IEE Electronics Letters, 23rd Oct 1997, 33(22), pp1858-1859.

Real-time Dynamic Deformable Meshes for Volumetric Segmentation and Visualisation, R. Bowden, T. A. Mitchell, and M. Sahardi. In Proc. BMVC, Adrian F. Clark Ed, Vol 1, pp 310-319, Essex, UK, Sept 1997.

Virtual Datagloves: Interacting with Virtual Environments Through Computer Vision, R.Bowden, A. J. Heap and C. Hart, In Proc. 3rd UK VR-Sig Conference, DeMontfort University, Chris Hand Ed, Leicester, UK, July 1996.

Some elements of this work are similar in nature and content to the work of A. J. Heap. However, this work was done concurrently and in isolation as demonstrated by the publication, Cluster Based non-linear Principal Component Analysis, which was submitted to IEE Electronics Letters on 2nd Sept 1997 and predates the work of the other author.

Table of Contents

Table of Figures

List of Abbreviations

ACM	**A**ctive **C**ontour **M**odel
ARM	**A**ctive **R**egion **M**odel
ASL	**A**merican **S**ign **L**anguage
ASM	**A**ctive **S**hape **M**odel
BSL	**B**ritish **S**ign **L**anguage
CBNLPDM	**C**luster **B**ased **N**on-**l**inear **P**oint **D**istribution **M**odel
CSSPDM	**C**onstrained **S**hape **S**pace **P**oint **D**istribution **M**odel
CCD	**C**harge **C**oupled **D**evice
CONDENSATION	**Con**ditional **Dens**ity Propag**ation**
FCM	**F**uzzy **C** **M**eans (k-means) Clustering Algorithm
HSV	**H**ue **S**aturation **V**alue
HSB	**H**ue **S**aturation **B**rightness
HLS	**H**ue **L**ightness **S**aturation
HSL	**H**ue **S**aturation **L**uminosity
HVC	**H**ue **V**alue **C**hroma
$\mathbf{I}_{\mathbf{x,y}}$	Intensity of pixel at point (x,y)
ISL	**I**nternational **S**ign **L**anguage
MRI	**M**agnetic **R**esonance **I**maging
NLPDM	**N**on-**l**inear **P**oint **D**istribution **M**odel
PCA	**Principal** **C**omponent **A**nalysis
PDF	**P**robability **D**ensity **F**unction
PDM	**P**oint **D**istribution **M**odel
RGB	**R**ed **G**reen **B**lue
ROI	**R**egion **o**f **I**nterest
σ	Standard Deviation
Voxel	**Vo**lumetric **El**ement