AUTOMATIC MARKERLESS POSE ESTIMATION FOR HUMAN IN SHORT SLEEVE OR LONG SLEEVE ATTIRE

LIM SIEW HOOI

THESIS SUBMITTED IN FULFILLMENT FOR THE DEGREE OF MASTER OF ENGINEERING

SCHOOL OF ENGINEERING AND INFORMATION TECHNOLOGY
UNIVERSITI MALAYSIA SABAH
2008

UNIVERSITI MALAYSIA SABAH

BORANG PENGESAHAN STATUS TESIS@

JUDUL :**AUTOMATIC MARKERLESS POSE ESTIMATION FOR HUMAN IN SHORT SLEEVE OR LONG SLEEVE ATTIRE**

IJAZAH :**SARJANA KEJURUTERAAN (PENGLIHATAN KOMPUTER)**

SESI PENGAJIAN : **2006 – 2008**

Saya, **LIM SIEW HOOI**, mengaku membenarkan tesis Sarjana Sains ini disimpan di Perpustakaan Universiti Malaysia Sabah dengan syarat-syarat kegunaan seperti berikut:

1. Tesis adalah hakmilik Universiti Malaysia Sabah,
2. Perpustakaan Universiti Malaysia Sabah dibenarkan membuat salinan untuk tujuan pengajian sahaja,
3. Perpustakaan dibenarkan membuat salinan tesis ini sebagai bahan pertukaran antara institusi pengajian tinggi,
4. TIDAK TERHAD.

Disahkan oleh:

(Penulis: **LIM SIEW HOOI**)

(TANDATANGAN PUSTAKAWAN)

Alamat tetap:

A25, Lorong Semangat 3,
Taman Semangat,
08000 Sungai Petani,
Kedah.

(Penyelia: **MR. LIAU CHUNG FAN**)

Tarikh: 28 Julai 2008

CATATAN: @ Tesis dimaksudkan sebagai tesis ijazah Doktor Falsafah dan Sarjana secara penyelidikan atau disertasi bagi pengajian secara kerja kursus dan penyelidikan, atau Laporan Projek Sarjana Muda (LPSM).

DECLARATION

I hereby declare that the material in this thesis is original except for quotations, excerpts, summaries and references, which have been duly acknowledged.

24 March 2008

Lim Siew Hooi
PK2006-8004

CERTIFICATION

TITLE : AUTOMATIC MARKERLESS POSE ESTIMATION FOR HUMAN IN SHORT SLEEVE OR LONG SLEEVE ATTIRE

DEGREE : MASTER OF ENGINEERING

VIVA DATE : 11 JULY 2008

DECLARED BY

1. SUPERVISOR
   MR. LIAU CHUNG FAN

ACKNOWLEDGEMENTS

I would like to express my deepest sense of gratitude and appreciation to my supervisor, Mr. Liau Chung Fan and my co supervisor, Dr. G. Sainarayanan for their guidance and encouragement during the course of this research work.

I am highly indebted to Mr. Liau Chung Fan for his inspiring and constructive criticism, invaluable suggestion and constant support for the successful completion of this thesis.

I faithfully thank to my co supervisor Dr. G. Sainarayanan for his invaluable suggestion, help and support for successful completion of this thesis.

I would like to thank Dean of School of Engineering and Information Technology, UMS for providing the necessary facilities to carry out this research.

I would like to express my sincere thanks to Dr. G. Balakrishnan for his time, assistance and discussion at the early stage of this research work.

I would also like to express my appreciation to lab assistance of Department of Electrical and Electronic Engineering, SKTM, UMS for the help during the project as well as those who have contributed indirectly towards the completion of this research.

I would also like to acknowledge with thanks to the NSF scholarship offered by the Ministry of Science, Technology and Innovation, Malaysia through UMS to carry out the work reported on this thesis.

Lastly, I wish to thank my family and Tan Chee Leong for the motivation and support throughout the course of my studies.

ABSTRACT

Automatic Markerless Pose Estimation for Human in Short Sleeve or Long Sleeve Attire

Markerless motion capture systems have the potential to provide an inexpensive, non-obtrusive solution for the estimation of body poses. Video analysis of human dynamics has a wide variety of applications in visual surveillance, visual-based human computer interaction, human robot interaction, gesture recognition and analysis, etc. Human pose estimation is a difficult problem, because human bodies are versatile, presenting a wide range of poses and self-occlusion occurrence. The goal of this project is to develop an automatic human pose estimation algorithm to recover the body and limb configurations in 2D from monocular video sequences without any special markers on the body. In this project, two approaches were described to model the human body poses from a monocular video sequence. In the first approach, a cardboard person model was proposed to model the upper part of human body using template matching algorithm. However silhouette feature alone was not enough to recover body configurations of the human when the body parts occluded each other and it was computationally expensive. Thus, a computer vision-based approach was proposed as the second method to automatically detect human body parts and estimate the human body poses from a markerless monocular video sequence. The input image was first segmented using a silhouette extraction function based on the brightness level transformation to extract the moving silhouette patterns (human figures) from a static background for subsequent processing. Human body parts detection was then performed using colour, contours and silhouettes cues. Human body model initialization was performed in a fully automatic way. The only assumption was the person should be in an upright and frontal poses in the video sequence. Circular head fitting was first detected using circular Hough transform method. K-means clustering was then performed on the circular head region to obtain the skin colour distribution of the face. A skin colour model was built from the detected face region and used to find the candidate positions of limbs. The pixel classification performance was measured by using a receiver operating characteristic (ROC) curve. Radon transform was used to obtain more accurate orientation of the upper arms. Various physical and motion constraints regarding the human body were then used to construct the upper body configuration. The relation of the hand position inside the torso region was introduced to estimate the human pose for long sleeve attire's users. The final stage of this project was to recognize the pose performed by the human subject. Nineteen different body poses were considered for classification using feed-forward backpropagation neural network, in which eight features are extracted from each pose. Our algorithm can recognize poses for any person entering the scene in either short sleeve or long sleeve shirt. It could also estimate the human poses even under illumination changes, self-occlusion occurrence and distance variations. There is no need for our system to use skin colour model built from skin pixel database for skin colour detection. Human pose estimation using computer vision-based approach reduces the computational cost by finding correct body part candidates more efficiently.

ABSTRAK

Sistem peralatan penangkapan pergerakan tanpa penandaan mempunyai potensi untuk membekalkan penyelesaian yang lebih murah dan "non-obtrusive" untuk penganggaran postur manusia. Sistem video penganalisa pergerakan manusia mempunyai pelbagai aplikasi seperti "video surveillance", interaksi antara manusia dan komputer, interaksi antara manusia dan robot, sistem pengecaman gerak isyarat dan lain-lain. Sistem penganggaran postur manusia adalah amat rumit kerana badan manusia adalah serba lentur untuk melakukan pelbagai postur dan kejadian "self-occlusion". Tujuan utama projek ini adalah untuk membangunkan satu algoritma penganggaran postur manusia secara automatik untuk menentukan konfigurasi badan dan tangan manusia dalam dua dimensi dengan menggunakan urutan video monocular tanpa sebarang penandaan khas pada badan. Dalam projek ini, dua cara digunakan di dalam pemodelan postur badan manusia melalui urutan video monocular. Cara pertama memperkenalkan model kadbod untuk pemodelan bahagian atas badan manusia dengan menggunakan algoritma "template matching". Akan tetapi, ciri "silhouette" sahaja tidak cukup untuk menentukan konfigurasi badan manusia apabila terdapat bahagian badan yang tersembunyi di belakang bahagian badan yang lain dan ia memerlukan masa yang panjang. Oleh itu, cara berdasarkan penglihatan komputer telah dicadangkan sebagai cara kedua untuk mengesan bahagian badan dan menganggar postur badan manusia secara automatik daripada urutan video monocular tanpa sebarang pakaian khas. Imej gambaran urutan pertama video disegmentasikan dengan menggunakan fungsi "silhouette extraction" yang berdasarkan transformasi peringkat kecerahan pada imej untuk mengasingkan latar depan yang dinamik (manusia) daripada latar belakang yang statik bagi pemprosesan seterusnya. Pengesanan bahagian badan manusia dilakukan dengan menggunakan ciri-ciri warna, kontur dan "silhouette". Initiasi model badan manusia dilakukan secara automatik. Anggapan yang dibuat adalah manusia mesti berhadapan dengan kamera dalam urutan video. Kepala manusia mula dikesan dengan menggunakan cara "circular Hough transform". Selepas itu, "k-means clustering" digunakan pada kawasan bulatan kepala untuk memperolehi taburan warna kulit pada muka. Model warna muka dibina daripada kawasan muka dan digunakan untuk menentukan posisi/kedudukan anggota badan yang lain. Prestasi klasifikasi pixel diukur dengan menggunakan "receiver operating characteristic (ROC) curve". "Radon transform" digunakan untuk memperolehi orientasi bahagian atas lengan tangan. Pelbagai kekangan fizikal dan gerakan berkaitan dengan badan manusia digunakan untuk membina konfigurasi bahagian atas badan manusia. Perhubungan antara kedudukan tangan dan badan (torso) diperkenalkan untuk menganggar postur manusia yang berpakaian lengan panjang. Peringkat terakhir projek ini adalah untuk mengecam postur manusia. Sembilan belas postur badan manusia yang berlainan diklasifikasikan dengan menggunakan "feed-forward backpropagation neural network". Lapan ciri diperolehi daripada setiap postur. Algoritma yang dibangunkan boleh mengecam postur manusia sama ada yang berpakaian lengan pendek atau lengan panjang. Ia juga boleh menganggar postur manusia walaupun terdapat perubahan cahaya, bahagian badan yang tersembunyi dan perubahan jarak antara manusia dan kamera. Sistem ini tidak memerlukan model warna kulit yang dibina daripada pangkalan data piksel kulit untuk mengesan warna kulit. Penganggaran postur badan manusia dengan cara berdasarkan penglihatan komputer mengurangkan masa pemprosesan dengan mencari dan menentukan calon-calon bahagian badan secara lebih cekap.

TABLE OF CONTENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS

ANN	Artificial Neural Network
CR	Classification rate
DOF	Degree of freedom
DSI	Disparity space image
FN	False negative
FP	False positive
FPR	False positive rate
GUI	Graphical user interface
HCI	Human-Computer Interaction
HCII	Human Computer Intelligent Interaction
HMM	Hidden Markov Model
HSV	Hue-Saturation-Value
NCC	Normalized cross correlation
RGB	Red-Green-Blue
rgb	Normalized RGB
ROC	Receiver Operating Characteristic
ROI	Region of interest
SAD	Sum of absolute difference
SSD	Sum of squared difference
SVM	Support Vector Machine
TN	True negative
TP	True positive
TPR	True positive rate

VIP	Video and Image Processing
2D	Two-dimensional
3D	Three-dimensional
6D	Six-dimensional

LIST OF SYMBOLS

A_R	Average aspect ratio of height to width of a face
a, b, c, d, e, f	Coefficients of the ellipse
c_i	Center of the *i* th cluster
d	Disparity value
$D(u,v)$	Distance from point (u,v) to the center of the frequency rectangle
DT(x,y)	Value of the pixel at coordinate (x,y) in the distance transform image
$f(x,y)$	Original image
$F(u,v)$	Fourier transforms of the $f(x,y)$
F(x,y)	Algebraic distance of the point (x,y) to the given conic
$g(x,y)$	Blurred image
$G(u,v)$	Fourier transforms of the $g(x,y)$
G_{opt}	Optimal generalization performance
$h(x,y)$	Inverse Fourier transform of the filter transfer function $H(u,v)$
$H(u,v)$	Gaussian lowpass filter function in frequency domain
H_{hs}	Head to shoulder height
$I_{bg}(x,y)$	Brightness in the background image
$I_{curr}(x,y)$	Brightness in the current image
$I_{seg}(x,y)$	Brightness in the segmented image
I_x, I_y and I_t	Spatiotemporal image brightness derivatives
$\log sig(n)$	Log-sigmoid transfer function
ma	Major axis of the ellipse

mb	Minor axis of the ellipse
p	General conic parameters of the ellipse
$\vec{Q_1}$ and $\vec{Q_2}$	Position vectors of the shoulder points and the point where traversal stops
R	Correlation coefficient
R^2	Coefficient of determination of the regression
t	Output target vector; $t = (t_1, \ldots, t_k, \ldots, t_m)$
$\tan sig(n)$	Hyperbolic tangent sigmoid transfer function
T_L and T_H	Low and high threshold for the face skin blob aspect ratio respectively
T_{opt}	Optimal trainability
u	x-component of image velocity
uarm_height	The width of the arm
uarm_width	The length of the arm
v	y-component of image velocity
v_{ij}	Weight for the connection between neuron i in the input layer and neuron j in the hidden layer
v_{oj}	Bias on hidden unit j
w_{jk}	Weight for the connection between neuron j in the hidden layer and neuron k in the output layer
w_{ok}	Bias on output unit k
W_s	Shoulder width
x	Input training vector; $x = (x_1, \ldots, x_i, \ldots, x_n)$
x_{head}	Top head position in x-coordinate
$\boldsymbol{x}_j$	The jth element to be clustered in feature vector form
x_l	Left x-coordinate of the person bounding box

x_r	Right x-coordinate of the person bounding box
x_L	X-coordinate of the left shoulder
x_R	X-coordinate of the right shoulder
(x, y)	Coordinates of points lying on the ellipse
(xc, yc)	Central position of the ellipse
y_{head}	Top head position in y-coordinate
y_k	Output unit k
y_s	Shoulder points in y-coordinate
z_j	Hidden unit j
α	Level of significance
n	Sample size
ρ	The shortest distance from the line to the origin of the coordinate system
δ_j	Error at hidden unit z_j
δ_k	Error at output unit y_k
μ	True mean
μ_H, μ_S	Mean of the Hue and Saturation components respectively
$\overline{X}$	Sample mean
Φ	Radian angle of the major axis with respect to the x-axis
σ_H, σ_S	Standard deviation of the Hue and Saturation components respectively
σ	Sample standard deviation
σ_g	Standard deviation of the Gaussian curve (cutoff frequency)
θ	Angle the line makes with the x-axis

LIST OF APPENDIX

CHAPTER 1

INTRODUCTION

1.1 INTRODUCTION

Vision-based motion capture systems and automatic human motion analysis have received great interests from computer vision researchers. Human motion analysis concerns with the detection, tracking and recognition of people, and more generally, the understanding of human behaviours from image sequences. Markerless vision-based human motion analysis has the potential to provide an inexpensive, non-obtrusive solution for the estimation of body poses. From an application perspective, computer vision-based methods provide the only non-invasive solution. The significant research effort in this domain has been motivated by a number of potential applications, especially in smart surveillance and Human-Computer Interaction (HCI). Other promising applications include video conferencing, athletic performance analysis, virtual reality, etc. The research area on human motion analysis contains a number of ill-posed problems such as inferring pose and motion of a highly articulated and self-occluding non-rigid 3D object from images.

Three major application areas of human motion analysis are surveillance, control and motion data analysis. Surveillance applications are inspired by the increased awareness of security issues for analysis of human actions, activities and behaviours. Surveillance applications cover the problems related to automatically monitoring and understanding locations where a lot of people pass through. Automated Video Surveillance deals with real time observation of people, leading to tracking and activity analysis of the subjects in the field of view. Human motion analysis is usually used to provide control and command in advanced user interfaces application. The control area relates to applications where the estimated motion or pose parameters are used to provide controlling functionalities. It could be used as an interface to games, virtual environments or animation. The third application area is concerned with the detailed analysis of the captured motion data. This may be used in automatic diagnostics of orthopedic patients or optimization of athletes' performances.

The evolution of user interface witnessed the development from text-based user interface based on keyboard to graphical user interface (GUI) based on mouse. In current virtual environments applications, keyboards, mice, wands and joysticks are still the most popular and dominant devices. However, they are inconvenient and unnatural. The use of human movements, especially hand gestures, has become an important part of Human Computer Intelligent Interaction (HCII) in recent years, which serves as a motivating force for research in pose modelling, analyzing and recognition of human motion. Many techniques developed in HCII can be extended to other areas such as sign-language translation, gesture driven controls and signaling in high-noise environment such as factories and airports (Turk, 1996). The study of automatic modes of interaction between humans and machines occurs at all levels ranging from video games and cameras to pedestrian activities such as interacting with ATMs or "check-in" terminals in airports for security purpose to detect suspicious activities.

To develop such interaction systems, computers must be able to automatically perceive and identify users' communicative actions such as poses and gestures and respond to them accordingly. A very crucial part of all these systems is the input module which is devoted to recognize the human operator in terms of tracking and recognition of human face, arms position, hand gestures and so on. The correct interpretation of human body and arm pose, hand gesture, and facial expressions is a crucial issue in HCII because it represents the main information path in the interface between man and machine.

Typically the input to a human motion capture system is a video sequence, in which the frames capture a person exhibiting some motion. The challenge is then to locate the individual body parts in the video frame and estimate the pose parameters from the relative locations. In commercial motion capture systems, markers are

REFERENCES

Agarwal, A. & Triggs, B. 2006a. A Local Basis Representation for Estimating Human Pose from Cluttered Images. *Proceedings of the 7th Asian Conference on Computer Vision*. **1**: 50-59.

Agarwal, A. & Triggs, B. 2006b. Recovering 3D Human Pose from Monocular Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **28**(1): 44-58.

Amat, J., Casals, M. & Frigola, M. 1999. Stereoscopic System for Human Body Tracking in Natural Scenes. *Proceedings of the IEEE International Workshop on Modeling People at ICCV'99*. 70-76.

Ardizzone, E., Chella, A. & Pirrone, R. 2000. Pose Classification using Support Vector Machine. *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks*. **6**: 317-322.

Attwood, C. I., Sullivan, G. D. & Baker, K. D. 1989. Model-Based Recognition of Human Posture Using Single Synthetic Images. *Fifth Alvey Vision Conference*. 25-30.

Baker, H. H. & Binford, T. O. 1981. Depth from Edge and Intensity Based Stereo. *Proceedings of Seventh International Joint Conference on Artificial Intelligence*. 631-636.

Barron, J. L., Fleet, D. J. & Beauchemin, S. S. 1994. Performance of Optical Flow Techniques. *International Journal of Computer Vision*. **12**: 43–77.

Bharatkumar, A. G., Daigle, K. E., Pandy, M. G., Cai, Q. & Aggarwal, J. K. 1994. Lower Limb Kinematics of Human Walking with the Medial Axis Transformation. *Proceedings of the IEEE Workshop on Motion of Non-Rigid and Articulated Objects*. 70-76.

Bregler, C. 1997. Learning and Recognizing Human Dynamics in Video Sequences. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 568-574.

Brown, M. Z., Burschka, D. & Hager, G. D. 2003. Advances In Computational Stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **25**(8): 993-1008.

Capo, A. J., Varona, J., Hidalgo, M. G., Mas, R. & Perales, F. J. 2006. Automatic Human Body Modelling for Vision-Based Motion Capture. *Proceedings of the 14th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* (*WSCG'2006)*. Plzen, Czech Republic.

Choo, K. & Fleet, D. J. 2001. People Tracking with Hybrid Monte Carlo. *IEEE International Conference on Computer Vision*. **II**: 321-328.

Cornelius, T. L. 1998. *Image Processing and Pattern Recognition (Neural Network Systems Techniques and Applications)*. San Diego: Academic Press.

Darrell, T., Maes, P., Blumberg, B. & Pentland, A. P. 1994. A Novel Environment for Situated Vision and Behavior. *Proceedings of the IEEE CVPR94 Workshop for Visual Behaviors*. 68-72.

Datta, A., Shah, M. & Lobo, N. D. V. 2002. Person-on-Person Violence Detection in Video Data. *Proceedings of the 16th International Conference on Pattern Recognition*. **1**: 433-438.

Davis, J. W. & Gao, H. 2003. An Expressive Three-Mode Principal Components Model of Human Action Style. *Journal of Image and Vision Computing*. **21**(11): 1001-1016.

Delamarre, Q. & Faugeras, O. 2001. 3D Articulated Models and Multi-View Tracking with Physical Forces. *Computer Vision and Image Understanding*. **81**(3): 328–357.

Deutscher, J., Blake, A., & Reid, I. 2000. Articulated Body Motion Capture by Annealed Particle Filtering. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **2**: 1144-1149.

Dhond, U. R. & Aggarwal, J. K. 1989. Structure from Stereo – A Review. *IEEE Transactions on Systems, Man and Cybernatics*. **19**(6): 1489-1510.

Elgammal, A. & Lee, C.-S. 2004. Inferring 3D Body Pose from Silhouettes Using Activity Manifold Learning. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **2**: 681-688.

Fausett, L. 1994. *Fundamentals of Neural Network*. New Jersey: Prentice Hall.

Fawcett, T. 2004. *ROC Graphs: Notes and Practical Considerations for Researchers*. Technical report. Palo Alto, USA: HP Laboratories.

Felzenszwalb, P. F. & Huttenlocher, D. P. 2005. Pictorial Structures for Object Recognition. *International Journal of Computer Vision*. **61**(1): 55–79.

Fitzgibbon, A. W., Pilu, M. & Fisher, R. B. 1996. Direct Least Squares Fitting of Ellipses. *IEEE Proceedings of the 13th International Conference on Pattern Recognition*. **1**: 253–257.

Fleet, D. J. & Jepson, A. D. 1990. Computation of Component Image Velocity from Local Phase Information. *International Journal of Computer Vision*. **5**(1): 77-104.

Forsyth, D. A. & Ponce, J. 2004. *Computer Vision: A Modern Approach*. (Eastern Economy edition). New Delhi, India: Prentice-Hall.

Freeman, W. T., Tanaka, K., Ohta, J. & Kyuma, K. 1996. Computer Vision for Computer Games. *Proceedings of the 2nd International Conference on Automatic Face- and Gesture-Recognition*. 100-105.

Fua, P. 1993. A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features. *Machine Vision Application*. **6**: 35-49.

Fujiyoshi, H. & Lipton, A. 1998. Real-Time Human Motion Analysis by Image Skeletonization. *Proceedings of Fourth IEEE Workshop on Application of Computer Vision*. 15–21.

Gautama, T. & Hulle, M. M. V. 2002. A Phase-based Approach to the Estimation of the Optical Flow Field using Spatial Filtering. *IEEE Transactions on Neural Networks*. **13**(5): 1127-1136.

Gonzalez, R. C. & Woods, R. E. 2002. *Digital Image Processing*. (2nd edition). New Jersey: Prentice-Hall.

Haralick, R. M. 1993. *Computer and Robot Vision*. Massachusetts: Addison Wesley.

Haritaoglu, I., Harwood, D. & Davis, L. S. 2000. W4: Real-time Surveillance of People and Their Activities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **22**(8): 809–830.

Heeger, D. J. 1988. Optical Flow using Spatiotemporal Filters. *International Journal of Computer Vision*. **1**(4): 279-302.

Heisele, B. & Wohler, C. 1998. Motion-Based Recognition of Pedestrians. *Proceedings of the Fourteenth International Conference on Pattern Recognition*. **2**: 1325-1330.

Horn, B. K. P. & Schunck, B. G. 1981. Determining Optical Flow. *Artificial Intelligence*. **17**: 185-203.

Hsu, R.-L., Abdel-Mottaleb, M. & Jain, A. K. 2002. Face Detection in Color Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **24**(5): 696-706.

Hua, G., Yang, M. H. & Wu, Y. 2005. Learning to Estimate Human Pose with Data Driven Belief Propagation. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. **2**: 747-754.

Ioffe, S. & Forsyth, D. A. 2001. Probabilistic Methods for Finding People. *International Journal of Computer Vision*. **43**(1): 45–68.

Jayaram, S., Schmugge, S., Shin, M. C. & Tsap, L. V. 2004. Effect of Colorspace Transformation, the Illuminance Component, and Color Modeling on Skin Detection. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. **2**: 813-818.

Jojic, N., Brumitt, B., Meyers, B., Harris, S. & Huang, T. 2000. Detection and Estimation of Pointing Gestures in Dense Disparity Maps. *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*. 468-475.

Ju, S., Black, M. J. & Yacoob, Y. 1996. Cardboard People: A Parameterized Model of Articulated Image Motion. *Proceedings of the Second IEEE International Conference on Automatic Face and Gesture Recognition.* 38-44.

Kahn, R. E., Swain, M. J., Prokopowicz, P. N. & Firby, R. J. 1996. Gesture Recognition Using the Perseus Architecture. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition.* 734-741.

Kameda, Y., Minoh, M. & Ikeda, K. 1993. Three Dimensional Pose Estimation of An Articulated Object from Its Silhouette Image. *Proceedings of Asian Conference on Computer Vision.* 612-615.

Kanade, T. & Okutomi, M. 1994. A Stereo Matching Algorithm with An Adaptive Window: Theory and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence.* **16**(9): 920-932.

Karaulova, I. A., Hall, P. M. & Marshall, A. D. 2000. A Hierarchical Model of Dynamics for Tracking People with A Single Video Camera. *Proceedings of British Machine Vision Conference.* 262–352.

Kim, S. H., Kim, H. G. & Tchah, K. H. 1998. Object-Oriented Face Detection Using Colour Transformation and Range Segmentation. *IEE Electronics Letters.* **34**(19): 979-980.

Kumar, S. S. & Chatterji, B. N. 2002. Stereo Matching Algorithms based on Fuzzy Approach. *International Journal of Pattern Recognition and Artificial Intelligence.* **16**(7): 883-899.

Kuno, Y., Watanable, T., Shimosakoda, Y. & Nakagawa, S. 1996. Automated Detection of Human for Visual Surveillance System. *IEEE Proceedings of the 13th International Conference on Pattern Recognition.* **3**: 865-869.

Laxton, B. 2007. *Monocular Human Pose Estimation.* Research Examination for the MS degree. San Diego: University of California.

Lee, M. W. & Cohen, I. 2004a. Human Upper Body Pose Estimation in Static Images. *European Conference Computer Vision (ECCV).* **II**: 126–138.

Lee, M. W. & Cohen, I. 2004b. Proposal Maps Driven MCMC for Estimating Human Body Pose in Static Images. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **2**: 334-341.

Lee, M. W. & Nevatia, R. 2007. Body Part Detection for Human Pose Estimation and Tracking. *IEEE Workshop on Motion and Video Computing*. 23-30.

Lee, M. W., Cohen, I., & Jung, S. K. 2002. Particle Filter with Analytical Inference for Human Body Tracking. *Proceedings of the IEEE Workshop on Motion and Video Computing*. 159–165.

Levine, M., O'Handley, D. & Yagi, G. 1973. Computer Determination of Depth Maps. *Computer Graphics and Image Processing*. **2**(4): 131-150.

Lucas, B. D. & Kanade, T. 1981. An Iterative Image Registration Technique with An Application to Stereo Vision. *DARPA Image Understanding Workshop*. 121-130.

Maren, A. J., Harston, C. T. & Pap, R. M. 1990. *Handbook Of Neural Computing Applications*. San Diego, California: Academic Press.

Memon, Q. & Khan, S. 2001. Camera Calibration and Three-Dimensional World Reconstruction of Stereo-Vision using Neural Networks. *International Journal of Systems Science*. **32**(9): 1155-1159.

Moeslund, T. B. & Garnum, E. 2001. A Survey of Computer Vision-based Human Motion Capture. *Computer Vision and Image Understanding*. **81**(3): 231–268.

Moore, D. J., Essa, I. A. & Hayes, M. H. 1999. Exploiting Human Actions and Object Context for Recognition Tasks. *Proceedings of the IEEE International Conference on Computer Vision*. **1**: 80–86.

Mori, G. & Malik, J. 2002. Estimation Human Body Configurations Using Shape Context Matching. *Proceedings of the European Conference on Computer Vision*. **2352**: 666-680.

Mori, G., Ren, X., Efros, A. A. & Malik, J. 2004. Recovering Human Body Configurations: Combining Segmentation and Recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **2**: 326-333.

Morris, D. D. & Rehg, J. M. 1998. Singularity Analysis for Articulated Object Tracking. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 289–296.

Muhlmann, K., Maier, D., Hesser, J. & Manner, R. 2002. Calculating Dense Disparity Maps from Color Stereo Images, An Efficient Implementation. *International Journal of Computer Vision*. **47**(1-3): 79-88.

Oren, M., Papageorigiou, C., Sinha, P., Osuna, E. & Poggio, T. 1997. Pedestrian Detection Using Wavelet Templates. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 193-199.

Park, S. & Aggarwal, J. K. 2003. Recognition of Two-Person Interactions Using A Hierarchical Bayesian Network. *Proceedings of the ACM SIGMM International Workshop on Visual* Surveillance. 65–76.

Pavlidis, I., Morellas, V., Tsiamyrtzis, P. & Harp, S. 2001. Urban Surveillance System: From the Laboratory to the Commercial World. *Proceedings of the IEEE*. **89**(10): 1478-1497.

Poppe, R. 2007. Vision-Based Human Motion Analysis: An Overview. *Computer Vision and Image Understanding*. **108**(1-2): 4-18.

Poppe, R., Heylen, D., Nijholt, A. & Poel, M. 2005. Towards Real-Time Body Pose Estimation for Presenters In Meeting Environments. *Proceedings of the International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision* (*WSCG'2005)*. 41-44.

Porle, R. R., Lim, S. H., Sainarayanan, G., Chekima, A. & Liau, C. F. 2006. Human Pose Classification using Neural Network for Visual Surveillance. *Proceedings of the 3rd International Conference on Artificial Intelligence in Engineering and Technology*. 749-753.

Poynton, C. A. 1995. Frequently Asked Questions About Colour. ftp://www.inforamp.net/pub/users/poynton/doc/colour/ColorFAQ.ps.gz.

Ramanan, D. & Forsyth, D. A. 2003. Finding and Tracking People from the Bottom Up. *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. **2**: 467–474.

Remondino, F. & Roditakis, A. 2003. 3D Reconstruction of Human Skeleton from Single Images or Monocular Video Sequences. *25th DAGM Pattern Recognition Symposium*. 100-107.

Ren, X., Berg, A. C. & Malik, J. 2005. Recovering Human Body Configurations using Pairwise Constraints between Parts. *Proceedings of the 10th IEEE International Conference on Computer Vision*. **1**: 824-831.

Scharstein, D. & Szeliski, R. 2002. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*. **47**(1-3): 7-42.

Siddiqui, M. & Medioni, G. 2006. Robust Real-Time Upper Body Limb Detection and Tracking. *ACM International Workshop on Visual Surveillance and Sensory Networks*. 53-60.

Sigal, L., Bhatia, S., Roth, S., Black, M. J. & Isard, M. 2004. Tracking Loose-Limbed People. *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*. **1**: 421–428.

Sigal, L., Sclaroff, S. & Athitsos, V. 2000. Estimation and Prediction of Evolving Color Distributions for Skin Segmentation under Varying Illumination. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **2**: 152-159.

Sivanandam, S. N., Sumathi, S. & Deepa, S. N. 2006. *Introduction to Neural Networks using Matlab 6.0*. New Delhi: Tata McGraw-Hill.

Sminchisescu, C. & Triggs, B. 2001. Covariance Scaled Sampling for Monocular 3D Body Tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. **1**: 447-454.

Song, Y., Goncalves, L. & Perona, P. 2003. Unsupervised Learning of Human Motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. **25**(7): 814-827.

Stern, H. & Efros, B. 2002. Adaptive Color Space Switching for Face Tracking in Multi-Colored Lighting Environments. *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*. 236-241.

Sul, C., Lee, K. & Wohn, K. 1998. Virtual Stage: A Location-Based Karaoke System. *IEEE Multimedia*. **5**(2): 42-52.

The MathWorks Inc. 1998. *Neural Network Toolbox User's Guide*. Natick, USA.

Toft, P. 1996. *The Radon Transform: Theory and Implementation*. Department of Mathematical Modelling. Denmark: Technical University.

Trucco, E. & Verri, A. 1998. *Introductory Techniques for 3-D Computer Vision*. New Jersey: Prentice Hall.

Tsai, R. Y. 1987. A Versatile Camera Calibration Techniques for High Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation*. **3**(4): 323-344.

Turk, M. 1996. Visual Interaction with Lifelike Characters. *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*. 368–373.

Vezhnevets, V., Sazonov, V., & Andreeva A. 2003. A Survey on Pixel-Based Skin Color Detection Techniques. *Proceedings of the Graphicon-2003*. 85-92.

Viola, P. & Jones, M. 2001. Rapid Object Detection Using A Boosted Cascade of Simple Features. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. **1**: 511-518.

Wang, L., Hu, W. & Tan, T. 2003. Recent Developments in Human Motion Analysis. *Journal of the Pattern Recognition*. **36**(3): 585-601.

Yamamoto, T. & Chellappa, R. 2003. Shape and Motion Driven Particle Filtering for Human Body Tracking. *Proceedings of the IEEE International Conference on Multimedia and Expo*. **3**: 61-64.

Yang, J. & Waibel, A. 1995. *Tracking Human Faces in Real-Time*. Technical report. Carnegie Mellon University.

Yang, J., Lu, W. & Waibel, A. 1998. Skin-Color Modeling and Adaptation. *Proceedings of the Third Asian Conference on Computer Vision*. **II**: 687-694.

Young, T. Y. & Calvert, T. W. 1974. *Classification, Estimation and Pattern Recognition*. New York: Elsevier.

Zarit, B. D., Super, B. J. & Quek, F. K. H. 1999. Comparison of Five Color Models in Skin Pixel Classification. *ICCV'99 International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*. 58-63.