TWO DIMENSIONAL UPPER HUMAN BODY POSE MODELLING SYSTEM

ROSALYN R PORLE

SCHOOL OF ENGINEERING AND INFORMATION TECHNOLOGY
UNIVERSITI MALAYSIA SABAH

2011

TWO DIMENSIONAL UPPER HUMAN BODY POSE MODELLING SYSTEM

ROSALYN R PORLE

THESIS SUBMITTED IN FULFILLMENT FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SCHOOL OF ENGINEERING AND INFORMATION TECHNOLOGY

UNIVERSITI MALAYSIA SABAH

2011

DECLARATION

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

23 September 2011

ROSALYN R PORLE
PS05-008-013

CERTIFICATION

NAME : ROSALYN R PORLE

MATRIC NO. : PS05-008-013

TITLE : TWO DIMENSIONAL UPPER HUMAN BODY POSE MODELLING SYSTEM

DEGREE : DOCTOR OF PHILOSOPHY (COMPUTER ENGINEERING)

VIVA DATE : 15 JULY 2011

DECLARED BY

MAIN SUPERVISOR Signature

Prof. Dr. Ali Chekima ____________________

CO-SUPERVISORS

Dr. Farrah Wong Hock Tze ____________________

Prof. Dr. G. Sainarayanan ____________________

ACKNOWLEDGEMENT

Bismillahirrahmanirrahim,

Alhamdulillah, my gratitude to Allah SWT, for His blessings and with His willing, I was able to complete this thesis. I am also gratified that throughout the process of completing this thesis, I have acquired new knowledge and discovered newfound wisdom useful, I am sure, for my future undertakings. Nevertheless, the completion of this thesis would not have been completely possible without the continuous support, undivided supervision, and inspiring motivation from various esteemed individuals.

Therefore, first and foremost, my heartfelt gratefulness is dedicated to Associate Professor Dr. Ali Chekima, my Main Supervisor, for his genuine support and dedicated guidance in the length of the entire research until the successful completion of this thesis.

I convey my deepest gratitude to Dr. Farrah Wong Hock Tze and Professor Dr. G. Sainarayanan, whom as my co-supervisors have dutifully shared valuable ideas and exceptional suggestions in strengthening the core of my research.

My sincere appreciation goes to my colleagues, Dr. Thangamani Uthirapathi, Dr. M. Muralindran, Mr. Edward Wong, Mr. Samuel Lim and Ms. Sheila Oi Yip for their time, assistance and contribution to the research. This appreciation is also extended to the volunteers from the SKTM undergraduates for their time and support in providing and obtaining the research database.

Last but not least, without the prayers, support and love from my family, I will not be where I am or what I am today, thus I thank my parents, brothers and sister for being there for me through my ups and downs, in sickness and in health, without reservation.

God bless, Wasalam.

ROSALYN R PORLE

23 September 2011

ABSTRACT

TWO DIMENSIONAL UPPER HUMAN BODY POSE MODELLING SYSTEM

Human Body Pose Modelling detects the human body parts and estimates their size, position and orientation in image sequences and then represents them using a specified model. It has wide applications such as user interface system, intelligent visual surveillance system and motion analysis in sports and medical applications. The human body can be divided into three main parts, which are the head, the torso and the limbs. In this thesis, the head, the torso and the arms of the human body are detected from the captured video. In this system, the human subject is assumed to be alone, in standing position, facing the video and wearing a short-sleeve clothe. The body parts are estimated and then represented in an image using rectangle shape. The input image sequences, which are acquired from the video, are processed using Single Level Haar Wavelet Transform decomposition. From the decomposition result, three features; namely the silhouette, the edge and the colour are extracted. Silhouette extraction, wavelet-based edge extraction and histogram analysis methods are employed to extract these features respectively. Rectangular shape is used as a model for each targeted body part. The parameters of the model such as the corner position, angle of rotation, width and length are computed using pose estimation methods. Four methods are proposed to estimate the head and the torso. These methods are Windowing Method I, Windowing Method II, Histogram Analysis I and Histogram Analysis II. All these methods use the extracted silhouette feature. Then, the arms pose is estimated using template-based matching. The arms pose are classified into non-occluded and occluded pose. Depending on this classification, different input features are applied. Silhouette or edge feature can be used for the non-occluded pose estimation. Meanwhile, colour feature is used for occluded pose estimation. The developed system is tested using 30 image sequences of different users. The best method for the silhouette extraction achieved 88.62 percent of correct silhouette detection and 98.68 percent of correct non-silhouette detection. For the colour extraction, the best colour format achieved 55.01 percent of correct skin colour detection and 97.71 of correct non-skin colour detection. For the pose estimation, the best pose estimation method for specified body parts achieved more than 90 percent of correct estimation. The proposed methods deals with: 1) partial occlusion, e.g. when the torso is not detected or partially detected, the size and position of the torso can still be estimated; 2) self occlusion, e.g. when the upper and lower arms are occluded with each other; and 3) hand crossing, e.g. when the right and left arms crossed with each other.

ABSTRAK

Pemodelan Pergerakan Tubuh Manusia mengesan bahagian tubuh manusia dan menganggar saiz, posisi dan orientasinya dalam rangkaian imej dan mewakilinya dengan mengunakan model tertentu. Ia mempunyai aplikasi yang luas seperti sistem antaramuka pengguna, sistem pengawasan visual cerdas dan analisa gerakan bagi tujuan sukan dan perubatan. Tubuh manusia boleh dibahagikan kepada tiga bahagian utama iaitu kepala, badan dan anggota lain. Di dalam tesis ini, kepala, badan dan tangan manusia akan dikesan dari video rakaman. Manusia di dalam rakaman tersebut dianggap bersendirian, dalam posisi berdiri, menghadap ke arah video dan memakai pakaian berlengan pendek. Bahagian tubuh akan dianggar dan diwakilkan dalam imej dengan menggunakan bentuk segiempat tepat. Rangkaian input imej yang diperoleh daripada video akan diproses menggunakan penguraian jelmaan Haar wavelet tahap tunggal. Dari hasil penguraian tersebut, tiga ciri iaitu silhuet, sisi dan warna akan diekstrak. Pengestrakan silhuet, pengekstrakan sisi berdasarkan wavelet dan teknik analisa histogram akan digunakan untuk mengestrak ciri-ciri yang disebutkan. Bentuk segiempat tepat akan digunakan sebagai model untuk setiap bahagian badan sasaran. Parameter model seperti kedudukan bucu, sudut pusingan, lebar dan panjang akan dikira dengan menggunakan teknik taksiran pergerakan. Empat teknik dicadangkan untuk taksiran ini iaitu teknik tetingkap I, teknik tetingkap II, analisa histogram I dan analisa histogram II. Semua teknik ini dapat dilaksanakan menggunakan sifat silhuet. Selepas itu, taksiran pergerakan lengan dilaksanakan dengan menggunakan pemadanan berdasarkan templat. Pergerakan lengan diklasifikasikan kepada pergerakan terhalang dan tak terhalang. Bergantung kepada klasifikasi ini, input yang berbeza akan digunakan. Sifat-sifat silhuet dan sisi boleh digunakan untuk taksiran pergerakan tak terhalang. Manakala sifat warna digunakan untuk taksiran pergerakan terhalang. Sistem yang dibangunkan diuji dengan menggunakan 30 rangkaian imej yang berbeza pengguna. Teknik terbaik untuk pengestrakan silhuet mencapai 88.62 peratus bagi pengesan silhuet yg betul dan 98.68 peratus bagi pengesan bukan silhuet yg betul. Untuk pengektrak warna, format warna terbaik mencapai 55.01 peratus bagi pengesan warna kulit yang betul dan 97.71 peratus bagi pengesan bukan warna kulit yang betul. Untuk taksiran pergerakan, teknik taksiran pergerakan yang terbaik bagi bahagian-bahagian badan tertentu mencapai lebih daripada 90 peratus taksiran yang betul. Sistem ini mengendalikan 1) separa halangan, sebagai contoh apabila badan tidak dikesan atau separa dikesan, saiz dan lokasi badan masih dapat ditaksirkan; 2) halangan sesama, sebagai contoh apabila lengan atas dan bawah terhalang sesama sendiri; dan 3) tangan bersilang, sebagai contoh apabila lengan kanan dan kiri bersilang antara satu sama lain.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

DWT	Discrete Wavelet Transform
HSC	Human Silhouette Classification
LL1	Lower left 1 region
LL2	Lower left 2 region
LL3	Lower left 3 region
LL4	Lower left 4 region
LR1	Lower right 1 region
LR2	Lower right 2 region
LR3	Lower right 3 region
LR4	Lower right 4 region
lF	Low pass filter
hF	High pass filter
RGB	Red, Green and Blue colour
ROI	Region of Interest
TN	True Negative
TP	True Positive
UR1	Upper right 1 region
UR2	Upper right 2 region
UR3	Upper right 3 region
UR4	Upper right 4 region
UL1	Upper left 1 region
UL2	Upper left 2 region
UL3	Upper left 3 region
UL4	Upper left 4 region

LIST OF SYMBOLS

μ_i	Class mean
ω_i	Class probabilities
θ_X	Rotational angle
$\sigma^2{}_B$	Between-class variance
A_{CIELAB}	*A* component of the *CIELAB* colour space
aH	Head Area
aO_i	Area of the labelled object
A_{SCT}	*A* component of the *SCT* colour space
B_{CIELAB}	*B* component of the *CIELAB* colour space
$bottomHS$	Bottom of the human silhouette
b_{rgb}	Blue component of the *rgb* colour space
B_{SCT}	*B* component of the *SCT* colour space
c	Column
$\bar{c}\,O_i$	Column centroid of the labelled object
$\bar{c}\,O_R$	Column centroid of the reference object
$cH2$	Leftmost head position
$cT2$	Leftmost torso position
$cH1$	Rightmost head position
$cT1$	Rightmost torso position
$cwhh2H$	Column histogram for the region above the shoulder
$cwhHS(c)$	Column histogram of the human silhouette
$d\,\bar{c}\,O_{i\_R}$	The difference of the column centroid between the labelled and the reference object
$dpHS(p1,p2)$	The distance transform of the human silhouette image
$d_{max}HS$	The maximum distance transform value
$d_{max}I_{MD}$	The maximum intensity differences between two image
ewT	Estimated width of the torso

$f(t)$	Function of t
fC	Frequency of colour histogram
$fcwhh2H$	Frequency of $cwhh2H$
$fcwhHS$	Frequency of the column histogram, $cwhHS(c)$
$f_{max}C$	Maximum frequency of colour histogram
$f_{max}cwhHS$	Maximum frequency of the column histogram, $cwhHS(c)$
$f_{min}C$	Minimum frequency of colour histogram
$f_{max}cwhh2H$	Maximum frequency of $cwhh2H$
$frwhHS$	Frequency of the row histogram, $rwhHS(r)$
$frwhirH$	Frequency of row histogram for the initial region head
g_{rgb}	Green component of the rgb colour space
g_{TSL}	Green component of the TSL colour space
$h2H$	Height of the region above the shoulder
hA	Arm Height
H_b	Histogram bin in colour histogram
hH	Head Height
hHS	Height of the human silhouette
H_{HSI}	Hue component of the HSI colour space
$h_{max}HS$	Maximum preferred height of the human silhouette
$h_{min}HS$	Minimum preferred height of the human silhouette
hO_i	Height of the labelled object
hT	Torso Height
htC	High threshold in colour histogram
i	Label number
I_{BS}	Background Subtraction image
I_C	Current image
I_c	Colour component image
I_{EF}	Extraction Function image
ihH	Initial height of the head
I_{HS}	Human silhouette image

I_{HSI}	Intensity component of the *HSI* colour space
i_{max}	Maximum number of labels in the image
I_{MD}	Output image for the Motion Detection
I_R	Reference image
I_{WA}	Approximation image component
I_{WB}	Blue colour component of the approximation image
I_{WD}	Diagonal image component
I_{WG}	Green colour component of the approximation image
I_{WH}	Horizontal image component
iwH	Initial width of the head
I_{WL}	Grey colour approximation image
I_{WR}	Red colour component of the approximation image
I_{WV}	Vertical image component
L_{CIELAB}	Luminance component of the *CIELAB* colour space
lmcwhHS	Last column position of the column histogram
lorHS	The left arm region
L_{SCT}	Luminance component of the *SCT* colour space
ltC	Low threshold in colour histogram
L_{TSL}	Luminance component of the *TSL* colour space
mW	Wavelet Transform Modulus
O_{HS}	Object of the human silhouette
O_i	Labelled object
O_{MD}	Output image of motion detection
O_R	Reference object
p	Point *p*
P_{LLAC}	Axis origin for the lower left arm
P_{LRAC}	Axis origin for the lower right arm
P_{ULAC}	Axis origin for the upper left arm
P_{URAC}	Axis origin for the upper right arm
r	Row

RH	Head ratio
$rH1$	Top head position
$rH2$	Bottom head position
R_{RbT}	The weight factor for $d_{max} I_{MD}$
R_f	Weight factor for frequency of colour histogram
$RlorHS$	Ratio of the left arm region
$rmcwhHS$	First column position of the column histogram
$rorHS$	The right arm region
r_{rgb}	Red component of the *rgb* colour space
$RrorHS$	Ratio of the right arm region
$rT1$	Top torso position
$rT2$	Bottom torso position
$RtrHS$	Ratio of above the torso region
r_{TSL}	Red component of the *TSL* colour space
$rwhHS(r)$	Row histogram of the human silhouette
$rwhirH$	Row histogram for the initial region head
S	Structuring element
S_{HSI}	Saturation component of the *HSI* colour space
S_{TSL}	Saturation component of the *TSL* colour space
tI_b	Threshold value of the background image
tI_{MD}	Threshold value of Motion Detection
$topHS$	Top of the human silhouette
$tpH(r,c)$	Total pixels for head estimation
$tprH$	Total pixels in the head region
$tprnH$	Total pixels for the new head estimation
$tpT(r,c)$	Total pixels for torso estimation
$trHS$	Above the torso region
T_{TSL}	Tint component of the *TSL* colour space
wA	Arm Width
wHS	Width of the human silhouette
$w_{max}HS$	Maximum width of the human silhouette

$w_{min}H$	Minimum width of the head
wO_i	Width of the labelled object
wO_R	Width of the reference object
vhO_i	Height validation of the labelled object
wH	Head Width
wHS	Width of the human silhouette
$whtT$	High threshold for the width of the torso
$wltT$	Low threshold for the width of the torso
$w_{max}HS$	Maximum width of the human silhouette
$w_{min}HS$	Minimum width of the human silhouette
wT	Torso Width
vwO_i	Width validation of the labelled object
X	X component of the *rgb* colour space
Y	Y component of the *rgb* colour space
Z	Z component of the *rgb* colour space

CHAPTER 1

INTRODUCTION

1.1 Human Body Pose Modelling System

Human Body Pose Modelling is a system that detects the human body parts and estimates their position and orientation in an image sequence using a specified model. This research is an active and growing research area in the last two decades, motivated by the need of man-machine systems to detect, recognize, track, model and analyze human automatically in order to interact effortlessly with user. The Human Body Pose Modelling is applicable in various areas, such as Gesture Identification for User Interface System, Human Tracking for Intelligent Visual Surveillance System and Motion Analysis in sport and medical fields.

Human body refers to the physical structure of a human. The human body is segmented into three main parts, which are the head, the torso and the limbs. The limbs comprise the arms and legs. Each of these parts is further subdivided into smaller parts; for example the arms are subdivided into upper arms, lower arms and the hands; and the legs are subdivided into thighs, calves and feet. Pose here refers to a particular position and orientation of an object (Forsyth and Ponce, 2004). Human Body Pose specifically refers to the movement and position of the whole human body or a particular human body part. Whereas, model refers to the representation or a simplified version of something (Cambridge International Dictionary of English, 1995). In the case of Human Body Pose Modelling, a model is used to directly or indirectly describe and represent the human movement in an image sequence.

The Human Body Pose Modelling is a combination of three major components which are: Human Body Detection, Model Construction and Pose Estimation. The first component, Human Body Detection, aims to detect the human body parts in an image sequence. The type of video camera(s) used for the system and the feature(s) that is extracted from the acquired image sequence will influence the detection method. The second component, Model Construction, aims to represent the human body parts in an image sequence using a specified model. The human body parts can be represented using Stick Figures, 2D Contours or Volumetric Models. The model either represents the whole human body or particular human body parts. The last component, Pose Estimation, configures the model so that it accurately reflects the position of the targeted human body parts. The Human Body Detection and the Pose Estimation are directly influenced by the model used in the system. Generally, the more complex the human body model, the more accurate the estimation result, hence it requires more expensive computation.

1.2 Objectives of the Thesis

This thesis has three objectives. The main objective of this thesis is to model the Upper Human Body Pose using 2D Contours. In order to model the Upper Human Body Pose, multiple features are extracted and then incorporated for detecting the human body in an image sequence. One of the features, the colour feature, is used especially for Skin Colour Segmentation. The performance of the Skin Colour Segmentation in six colour spaces will be evaluated in this thesis. Details on the objectives are given as follows:

i. To model the Upper Human Body Pose in an image sequence using 2D Contours

The task of modelling the Human Body Pose is not straightforward, as human movement is non-rigid. The variety of body pose is unlimited and the size of body parts is person dependent. Furthermore, the occurrence of occlusion, deformable body parts, shadows and varying illumination in the image sequence add to the problems in the system. In this thesis, some of these issues are addressed. The system focuses on modelling the upper human body parts, which includes the head, the torso, the upper arms and the lower arms. 2D Contours are used to represent the human body parts. The modelling method is performed in a hierarchical manner, such that the estimated position and size of the reference body part influences the position and size of the other body parts.

ii. To extract and incorporate multiple features for Human Body Detection

Image sequence offers a wealth of information to aid in the analysis of a real world scene. Silhouette, edge and colour are some examples of features that are available in an image sequence. Due to the richness of information, the image processing requires an enormous amount of computation. Various methods have been established by researchers; either by applying a single feature or by integrating several essential features in the system. A single feature may not be sufficient to address Pose Estimation issues, especially the occlusion issue, which deemed not favourable for a robust system. In this thesis, multiple features from the image sequence are extracted. These features are either independently applied or fused together to detect the human body parts.

iii. To evaluate the performance of the Skin Colour Segmentation in six colour spaces

Colour is useful for detecting the human arms, especially when the arms are occluded with the other body parts. In developing Skin Colour Segmentation, two main issues are addressed: the first issue is the selection of colour space, and the second issue is the selection of Skin Colour Segmentation method for the system. The selection of the colour space is essential since it has significant effect on the Skin Colour Segmentation. The default colour space that is widely used in the machine image processing field is the *RGB* colour space. There are other colour spaces that are applied for skin colour detection besides the *RGB* colour space. Despite having so many colour spaces, neither a single nor a perfect colour space can be applied for segmentation task that can represent a real world scene (Battle *et al.*, 2000). In this thesis, a Skin Colour Segmentation method is proposed to detect human arms, and the performance of this method is evaluated using six different colour spaces. The selection of the best colour space for the proposed method is determined based on the experimental results.