ภาคผนวก ก
หนังสืออนุมัติจริยธรรมการวิจัย

แบบเสนอขออนุมัติจริยธรรมการวิจัยจากคณะกรรมการจริยธรรมการวิจัย

ชื่อ: นาย ธรรมศาสตร์ วิสุทธารมณ์
ผศ.ดร. พรชัย มงคลนาม
ผศ.ดร. โจนาธาน โฮยิน ชาน

หัวข้อวิจัย: การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้องคิเนค

วันที่: 7 กรกฎาคม2557

สรุปย่อโครงการวิจัย:

ในงานวิจัยนี้มีจุดประสงค์เพื่อประยุกต์เทคโนโลยีสารสนเทศในการช่วยติดตามพฤติกรรมของผู้สูงอายุในระหว่างการรับชมโทรทัศน์โดยใช้กล้อง Kinect ในการวิเคราะห์การเคลื่อนไหวของผู้สูงอายุ กล้อง Kinect จะทำหน้าในการค้นหาร่างกายมนุษย์ และ ทำการวิเคราะห์จุดต่างๆบนร่างกาย ซึ่งจุดบนร่างกายที่ผ่านการวิเคราะห์จากกล้อง Kinect นั้นจะอยู่ในรูปแบบข้อมูลพิกัดสามมิติ ซึ่งประกอบด้วย ข้อมูลแกน X, Y และ Z โดยในงานวิจัยชิ้นนี้จะนำเอาเข้ามูลที่สามารถบันทึกได้จากกล้อง Kinect ไปทำการวิเคราะห์ท่าทางของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ ซึ่งในการทดลองของงานวิจัยชิ้นนี้จะเลือกใช้กลุ่มตัวอย่างเป็นกลุ่มผู้สูงอายุจำนวน 10 คน ซึ่งเป็นบุคคลที่มีอายุตั้งแต่ 60 ปีบริบูรณ์ โดยประกอบด้วยเพศชาย 5 คน และ เพศหญิง 5 คน ซึ่งมีความสูง และ รูปร่างแตกต่างกันออกไปเพื่อให้เกิดความหลากหลายทางข้อมูล ในการทดลองจะทำการบันทึกข้อมูลพฤติกรรมการแสดงออกของท่าทางจริงของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ โดยท่าทางที่งานวิจัยให้ความสนใจเป็นท่าทางพื้นฐานในการรับชมโทรทัศน์ เช่น การนั่ง การยืน การนอน การกอดอก และ การยกมือเป็นต้น โดยผู้เข้าร่วมงานวิจัยทุกคนจะต้องทำการรับชมโทรทัศน์เป็นเวลา 5 นาที เป็นจำนวน 5 ครั้ง ในแต่ละครั้งสามารถแสดงท่าทางได้อย่างอิสระ และ ไม่มีการลำดับท่าทางในระหว่างการรับชมโทรทัศน์ โดยผู้เข้าร่วมทำการทดลองจะต้องทำท่าทางต่าง ๆ ตามที่งานวิจัยนี้ได้กำหนดไว้ให้ครบทุกท่าทาง ซึ่งข้อมูลที่ได้จากการบันทึกท่าทางของผู้เข้าร่วมงานวิจัยทั้ง 10 คนจะถูกนำมาประยุกต์ใช้กับหลักการการทำเหมืองข้อมูล (Data mining) โดยในงานวิจัยนี้เลือกแบบจำลอง (model) เพื่อทดลองทำการเปรียบเทียบผลลัพธ์ของแบบจำลองที่เหมาะสมกับงานวิจัย ซึ่งประกอบด้วย Neural Networks, Support Vector Machine, Naïve Bayes, Decision-tree, Logistic regression และ Random forest โดยแบบจำลองที่ให้ผลลัพธ์ดีที่สุดจะถูกเลือกมาใช้ในการพัฒนาระบบแยะแยกท่าทางในระหว่างรับชมโทรทัศน์ โดยระบบที่ถูกพัฒนาเสร็จสมบูรณ์จะถูกนำไปติดตั้งตามบ้านผู้สูงอายุเพื่อใช้ในการช่วยติดตามพฤติกรรมของผู้สูงอายุในระหว่างรับชมโทรทัศน์

ประเด็นเชิงจริยธรรมและการดำเนินการ

1. การยินยอมเข้าร่วมในการวิจัย
   - กลุ่มตัวอย่างทั้งหมด 10 คน จะได้รับหนังสือขอความยินยอมเข้าร่วมในงานวิจัย และ ได้รับการบอกกล่าวรวมถึงอธิบายว่าการเข้าร่วมในงานวิจัยนี้เป็นไปโดยความสมัครใจ และ หากผู้เข้าร่วมงานวิจัยไม่ลงนามในหนังสือยินยอมเข้าร่วมในงานวิจัย ข้อมูลที่เก็บรวบรวมได้จะไม่นำมาเป็นส่วนหนึ่งของการวิเคราะห์ข้อมูล
2. สิทธิส่วนบุคคล/ การรักษาข้อมูลเป็นความลับ
   - ชื่อนามสกุลจริงของผู้เข้าร่วมงานวิจัยทุกคนจะถูกใช้แทนด้วยนามสมมติ ซึ่งจะไม่สามารถระบุถึงตัวตนของบุคคลที่เข้าร่วมการทดลองได้
3. การปกป้องผู้เข้าร่วมงานวิจัยจากความเสี่ยงต่อการได้รับผลกระทบเชิงลบจากการทำวิจัย
   - กรณีผู้เข้าร่วมงานวิจัยได้รับอุบัติเหตุจากการทดลองผู้เข้าร่วมงานวิจัยจะได้รับค่าใช้จ่ายในการรักษาพยาบาลจากผู้วิจัย

หนังสืออนุมัติจริยธรรมการวิจัย

ชื่อโครงการวิจัย การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้องคิเนค
ชื่อผู้วิจัย นาย ธรรมศาสตร์ วิสุทธารมณ์
ที่อยู่ที่ติดต่อได้ คณะเทคโนโลยีสารสนเทศ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี เบอร์โทรศัพท์ 087-470-8106
อีเมล์ Thammarsat@gmail.com

โครงการวิจัยดังกล่าวนี้ได้รับการพิจารณาตรวจสอบ และได้รับการอนุมัติจริยธรรมการวิจัยที่ถือปฏิบัติตามกฎเกณฑ์และมาตรฐานของคณะกรรมการจริยธรรมการวิจัยคณะศิลปศาสตร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีแล้ว

คณะกรรมการจริยธรรมการวิจัย

ลงนาม Richard Watson Todd
ชื่อ รองศาสตราจารย์ ดร.ริชาร์ด วัตสัน ทอดด์
วันที่ 24 ก.ค. 2557

ลงนาม
ชื่อ ผู้ช่วยศาสตราจารย์ ดร.ปัทมวรรณ ซิลลิ
วันที่ 24 ก.ค. 2557

ลงนาม PASSANAN ASSAVARAK
ชื่อ ดร.ภาสนันทน์ อัศวรักษ์
วันที่ 24 ก.ค. 2557

หนังสือยินยอมเข้าร่วมในการวิจัย

ชื่อโครงการวิจัย การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้องคิเนค

ชื่อผู้วิจัย นาย ธรรมศาสตร์ วิสุทธารมณ์

ที่อยู่ที่ติดต่อได้ คณะเทคโนโลยีสารสนเทศมหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรีเบอร์โทรศัพท์ 087-470-8106

อีเมล์ Thammarsat@gmail.com

ในงานวิจัยนี้มุ่งเน้นการประยุกต์ระบบเทคโนโลยีสารสนเทศในการติดตามพฤติกรรมของผู้สูงอายุในระหว่างการรับชมโทรทัศน์ ซึ่งกลุ่มเป้าหมายคือบุคคลที่มีอายุตั้งแต่ 60 ปีบริบูรณ์ ซึ่งในงานวิจัยนี้ผู้วิจัยจะทำการจำลองสถานการณ์การรับชมโทรทัศน์ โดยต้องการให้ผู้เข้าร่วมงานวิจัยทำการแสดงกิจกรรมต่าง ๆ ในระหว่างการรับชมโทรทัศน์ จำนวน 5 ครั้ง ในแต่ละครั้งใช้เวลาประมาณ 5 นาที โดยใช้กล้อง Kinect ในการบันทึกพฤติกรรมของผู้เข้าร่วมงานวิจัย โดยผู้เข้าร่วมงานวิจัยสามารถแสดงท่าทางต่างๆตามที่ได้กำหนดไว้อย่างอิสระ ซึ่งผู้เข้าร่วมงานวิจัยจะต้องแสดงทุกท่าทางให้ครบตามที่งานวิจัยได้กำหนดไว้ โดยท่าทางที่งานวิจัยให้ความสนใจจะเป็นเพียงท่าทางพื้นฐานซึ่งจะไม่ส่งผลเสียต่อร่างกายของผู้เข้าร่วมงานวิจัย อย่างไรก็ตามในกรณีที่ผู้เข้าร่วมงานวิจัยได้รับการบาดเจ็บจากการร่วมทำการทดลอง ทางผู้วิจัยจะเป็นผู้ออกค่าใช้จ่ายในการรักษาพยาบาลให้แก่ผู้เข้าร่วมงานวิจัย

ในกรณีที่ผู้เข้าร่วมงานวิจัยมีความต้องการถอนตัวจากการทดลอง ผู้เข้าร่วมงานวิจัยสามารถถอนตัวจากการทดลองได้ตลอดเวลา และ ข้อมูลที่ได้ทำการบันทึกไว้แล้วของผู้เข้าร่วมงานวิจัยที่ได้ถอนตัวจากการทดลอง ผู้วิจัยจะไม่นำข้อมูลเหล่านั้นมาใช้ต่อในงานวิจัย

ทั้งนี้ข้อมูลส่วนบุคคลอันได้แก่ใบหน้า ชื่อ และ นามสกุล ของผู้เข้าร่วมงานวิจัยทุกคนจะไม่ถูกนำเสนอสู่สาธารณะเพื่อเป็นการคงไว้ซึ่งสิทธิเสรีภาพส่วนบุคคล

ข้าพเจ้าซึ่งได้ลงนามที่ด้านล่างของหนังสือฉบับนี้ ได้รับคำอธิบายอย่างชัดเจนจนเป็นที่พอใจจากผู้วิจัยถึงวัตถุประสงค์และขั้นตอนการวิจัย และประโยชน์ซึ่งจะเกิดขึ้นจากการวิจัยเรื่องนี้แล้ว

ข้าพเจ้าเข้าร่วมการวิจัยครั้งนี้ด้วยความสมัครใจ และข้าพเจ้ามีสิทธิ จะถอนตัวออกจากการวิจัยเมื่อไรก็ได้ตามความประสงค์ ข้าพเจ้ายินดีเข้าร่วมการวิจัยครั้งนี้ ภายใต้เงื่อนไขที่ระบุไว้ในเอกสารข้อมูลสำหรับกลุ่มประชากรหรือผู้มีส่วนร่วมในการวิจัย

____

ลงนามผู้มีส่วนร่วมในการวิจัย

____

ตัวบรรจง

____

วันที่

คำรับรองจากผู้วิจัย

ข้าพเจ้ารับรองว่า ข้อมูลที่ได้รับจะถูกใช้เพื่อวัตถุประสงค์ในการวิจัยตามที่ได้ระบุไว้เท่านั้นเว้นเสียจากจะได้รับความยินยอมให้ใช้ได้ในวัตถุประสงค์อื่น และข้าพเจ้าจะเก็บรักษาข้อมูลนี้เป็นความลับและไม่เปิดเผยข้อมูลเฉพาะตัวของผู้มีส่วนร่วมในการวิจัยเว้นเสียจากจะได้รับความยินยอมให้ดำเนินการได้

กรุณาลงนามหน้าถัดไป

____

ลงนามผู้วิจัย

____

ตัวบรรจง

____

วันที่

ภาคผนวก ข
ผลงานที่ได้รับการตีพิมพ์

Postural Classification using Kinect

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan Hoyin Chan

The 2014 International Computer Science and Engineering Conference (ICSEC2014)

July 30 - August 1, 2014

2014 International Computer Science and Engineering Conference: ICSEC 2014 InternationalTrack

Postural Classification using Kinect

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan Hoyin Chan
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok 10140, Thailand
{pornchai | jonathan}@sit.kmutt.ac.th

***Abstract*—This research focuses on the comparison of posture recognition, using a data mining classification approach on the skeleton data stream obtained from Kinect camera. We classified four standard postures including *Stand, Sit, Sit on floor* and *Lie Down*. We compared six classifiers, namely, decision tree, neural network, naïve Bayes, support vector machine, logistic regression and random forest in order to find a suitable classifier. Our best results can correctly classify the postures with 97.88% accuracy, 97.40% sensitivity, and 0.991 ROC area under curve using Max-Min normalization with a decision tree classifier on *four* transformed attributes. Our future work will use the knowledge obtained to classify a wider range of postures of the elderly while watching television, to be a part of a bigger effort to monitor and study elderly behavior at home.**

Keywords— data mining; elderly; Kinect camera; postural classification

I. Introduction

There are two definitions of ageing society as defined by the United Nations: when people in the society aged 65 years old or over comprise more than 7%, or 60 years old or over comprise more than 10% of a population. From the second definition, the world became an ageing society since 2000 when the number of people aged 60 years old or over made up about 10% of the world's population [1-2] this number is expected to rise to 21.1% by year 2050, as can be seen in Table I. Thailand, in particular, became an ageing society since 2007 when it had about 11.0% ageing population [3]. Although the ageing population increases globally, elderly people are being taken care less and less. Our main reason is that more and more family members work outside of the home. In addition, physical decline and less financial support are the other factors that largely keep the elderly people at home. They tend to be alone in daytime at best or alone the whole day at worst. One technology that could be used to help to improve this quality of life and to help family members to understand them better is a software system that could track daily activities engage them in various social activities. Specifically, tracking postures of the elderly people while watching television is of our interest and is the focus of our work. We would like to learn what most elderly people spend a substantial time at home with doing while watching television.

Like many researchers who pay attention to behavioral tracking of elderly people. In this paper, we compare the efficacy of classification accuracy in postural classification, using Kinect camera.

978-1-4799-4963-2/14/$31.00 ©2014 IEEE

TABLE I. Percentage of population aged 60 years or over [2].

Region/Country	Year 1950	Year 2000	Year 2050
World	8.2%	10.0%	21.1%
Asia	6.8%	8.8%	22.6%
Europe	12.1%	20.3%	36.6%
China	7.5%	10.1%	29.9%
India	5.6%	7.6%	20.6%
Japan	7.7%	23.2%	42.3%
Thailand	5.0%	8.1%	27.1%
USA	12.5%	16.1%	26.9%

The knowledge and methodology obtained from this experiment is used to classify the basic postures of elderly while watching television, which is a part of tracking elderly behavior.

II. Literature Review

A. Relationship between Elderly People and Televisions

Health and financial problems of elderly people tend to keep them staying more at home and meeting fewer people. When they meet fewer people and spend less time socially interacting, they start to lose social contacts, become more isolated and unaware of current affairs. In addition, they may spend more time alone because the household family members are too busy working outside. As a result, watching television has become a preferred activity for elderly people because it can help them see the world from their homes. Roles of televisions were studied by Reid [4] in "Lifeline or Leisure?: TV's Role in the Lives of the Elderly". He compared television to a window that made them see many things happening around the world, and preserved the sense of participation in society and resisted the feelings of isolation and loneliness [5]. Several researchers have found elderly people love to spend many hours per day in front of televisions [6-8]. In addition, many researches focus on developing a television system appropriate for elderly people. For instance, in 2007 Nakajima et al. [9] developed a simple and inexpensive system or telemonitoring system of television's operating state for remotely located families to improve smooth communications between elderly people, who live alone, and their families. In 2012, Choomkasean et al. [10] proposed a conceptual model of a multimedia delivery to televisions of the elderly people. The system could receive multimedia to the elderly people via televisions, which is used to remind them of upcoming events or to allow them to conveniently communicate with their family members and friends. Consequently, it helped sustain relationships between the elderly people and their family members and friends, which could promote mental health in the long run and improve the quality of the elderly.

B. Gestural and Postural Detection

Two dimension image processing is the most popular technique used to detect human gestures or postures using a simple Web camera or a video camera. However, the technique still has the problem of separating a desired object from a complex background. On the other hand, 3D imaging device can provide a better way to extract a human body out of a complex background. However, it has been in limited use for many years mainly due to its high cost and low accuracy in detecting and extracting desired objects.

In 2010, Kinect cameras and Kinect library functions were released by Microsoft [11-13]. It is a game controller that is controlled by gestures and spoken commands of the user. Kinect library functions can separate a human body from a complex background in any kind of lighting conditions and analyze each presumed point of the human body joints in order to build a visual body structure (skeleton). Each joint consists of a 3D coordinate of X, Y, and Z position. Kinect camera has two competing library functions: OpenNI and Microsoft SDK, which is chosen for this research due to its effectiveness. Microsoft SDK library has two detection modes. One is a default Stand tracking mode; this mode can detect all body joints (20 skeletal joints). Another is a Seated tracking mode; this mode can detect the upper shoulder joints (10 skeletal joints) as shown in Fig. 1. The Kinect system includes an RGB camera, a depth sensor and a multi-array microphone as shown in Fig. 2. The RGB camera is placed between the depth sensors. The camera is used to record 2D image. The depth sensor is used to record depth image for analysis of the skeleton. In addition, the depth sensor system in Kinect camera is of infrared type. Therefore the Kinect camera still works with 3D mode in dark areas. The multi-array microphone is placed horizontally on the Kinect camera body which is used to record sound commands from users.

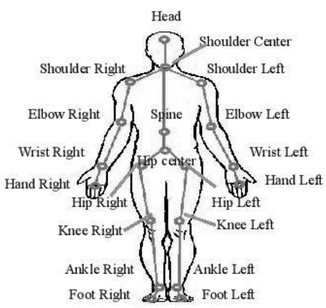

Fig. 1. Twenty Skeletal Joints.

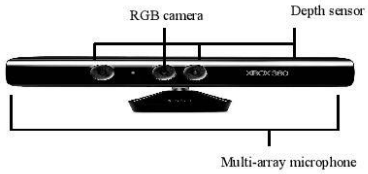

Fig. 2. Kinect components [13].

The ability of Kinect in extracting a 3D human skeleton and its competitive price when compared with other 3D imaging devices are the main reasons why many researches use it to capture and analyze human gestures and postures.

In 2012 Patsadu et al. [14] presented human gesture recognition with Kinect system using a data mining classification technique. Their data set was extracted from input of twenty body-joint positions. The average accuracy of all classification methods was 93.72%, and that can confirm capability of the Kinect camera for use for human gesture recognition. Also in 2012, Liu et al. [15] introduced a view independent posture recognition framework by presenting a body orientation classification system. The posture was used in classification of 5 standing postures. Their results showed that the framework was robust to variant viewpoint with high accuracy.

In 2013, Kaenchan et al. [16] proposed automatic data gathering with multiple Kinect cameras from several viewpoints to create complete skeleton to solve the problem that some parts of the human body may be obscured by objects.

In 2014, Paliyawan et al. [17] presented a system for monitoring office workers in order to prevent Office Workers Syndrome using Kinect camera. The system can alert a user when it is time to relax and provide a daily summary report used to track working behavior of the user.

C. Behavioral Tracking

Behavioral tracking of elderly people can help family members, physicians, and other health practitioners understand the elderly people better. Knowing more about the behaviors and activities of the elderly has been one of the most sought-after studies in imminent ageing societies. Technology could be used to monitor and keep records of the daily lives of the elderly, especially while being at home, to help us better understand and innovate technologies to serve them.

In 2010, Cardile et al. [18] introduced a computer vision-based architecture for remote elderly monitoring, which was built upon a network of wireless camera sensors for tracking activities of the elderly people. The work was motivated by the necessity of unobtrusive remote caring systems for elderly people.

In 2012, Fahim et al. [19] developed an application to track daily life activities tracked by smart phones, which could send messages to remind elderly people to complete some remaining work, take drugs, schedule some activities, etc. In addition, the system can track activities of the elderly people in each day to provide useful data to family members and care givers. Also in 2012, Correa et al. [20] proposed a navigation system for elderly care applications based on wireless sensor networks, which monitored emergency situations of the elderly people. The system added some functionalities in order to obtain a full health care system for increasing standard of lives.

In 2013, Kim et al. [21] proposed a real-time emergency alarm system that could monitor motions of elderly people. The system could send out alert messages and images quickly to the designated people when an emergency situation occurred to the elderly people who lived alone.

III. PROCESS OVERVIEW

Our experiment was separated into two parts. Part one is data preprocessing that is done prior to postural classification. Part two is classification that would compare results from each candidate classifier models. An outline of the experimental procedure is shown in Fig. 3. We used Kinect XBOX 360 and Microsoft SDK version 1.8 working in default standing mode that can give raw data of 20 joint positions. In the experiment, we used only 10 joints including head, shoulder left, shoulder right, hip left, hip right, hip center, knee left, knee right, foot left and foot right with data transformations instead of using all available 20 joints.

The data used in our training set are collected from ten people (5 males and 5 females) with heights between 150-180 centimeters and with various body sizes. Each subject was asked to perform different postures (*Stand, Sit, Sit on Floor,* and *Lie Down*) for 4-6 minutes per round. The subjects were asked to be natural and try not to be stationary only for the duration to emulate the real television watching process. Five rounds were performed for each subject in front of the Kinect cameras, which were about 1.8-3.0 meters away. The Kinect cameras captured about 25-31 frames per second. Fig. 4 shows four different postures of a sample subject. The training data set consists of 30,088 instances (each instance comprises the four attributes mentioned in Section IV), with 7,697 instances for *Stand*, 8,848 instances for *Sit*, 5,819 instances for *Sit on Floor*, and 7,724 instances for *Lie Down*.

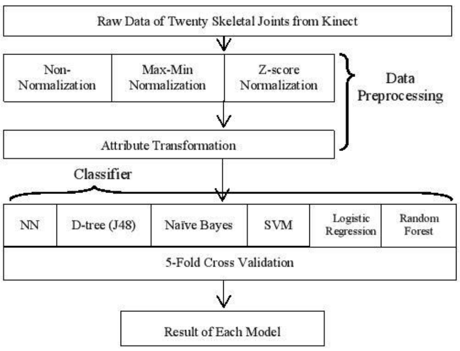

Fig. 3. Process Overview.

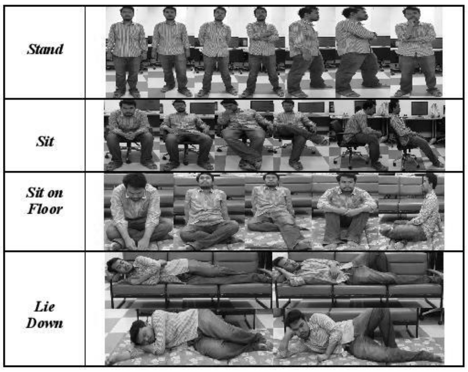

Fig. 4. Different Posture Examples.

IV. DATA PREPROCESSING

In the step of data preprocessing, we compared two techniques of normalization to one without normalization of the raw data. The two normalization techniques used were Max-Min and Z-score. The corresponding normalization equations are shown as (1) and (2), respectively.

$$J' = \frac{J - \mathrm{Min}}{\mathrm{Max} - \mathrm{Min}}, \quad (1)$$

$$J' = \frac{J - \bar{J}}{\sigma}. \quad (2)$$

where J' is the normalized data; Min and Max of (1) are the minimum and maximum values in each skeleton joint position training set; $\bar{J}$ and σ in (2) are the mean and the standard deviation of each skeleton joint position, respectively.

The data used in classification have four transformed attributes, including angle of left knee, angle of right knee, distance from hip to a room floor, and the aspect ratio. These attributes are provided as follows.

A. Aspect ratio of height and width of the depth image

Each posture gives a different aspect ratio of the width and the height as shown in Fig. 5. Therefore, this ratio was selected to be one of the attributes for classification.

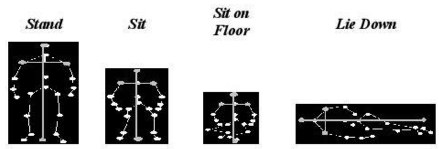

Fig. 5. Aspect Ratio of Each Posture.

The aspect ratio values as in (3)-(5) are derived from the ratio of distance between maximum and minimum values of x and y positions of only 5 joints including head, shoulder left, shoulder right, foot left and foot right.

$$\text{Width} = \text{Max x} - \text{Min x}\ , \quad (3)$$

$$\text{Height} = \text{Max y} - \text{Min y}\ , \quad (4)$$

$$\text{Aspect Ratio} = \frac{\text{Width}}{\text{Height}}\ . \quad (5)$$

However, using the aspect ratio only cannot give a high accuracy. In this research, we added three attributes including angle of left knee, angle of right knee, and distance from hip to a room floor. Angles of left knee and right knee can help to differentiate between *sit* and *stand*. When people stand, the angles of their knees are larger. On the other hand, when they sit, the angles of their knees are smaller as shown in Fig. 6. In addition, we used distance from hip to room floor for helping to classify *sit* and *sit on floor*. When a subject changes posture to *sit on floor*, the distance of hip to room floor will decrease as shown in Fig. 7.

B. Angles of left knee and right knee

Angles, θ, of the left knee and the right knee are used to analyze *Sit* or *Stand* postures. Both angles of the left knee and right knee are computed using (6)-(8), where X, Y, and Z are the coordinates of the knee joint and hip joint positions.

$$A = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}\ , \quad (6)$$

$$B = |Y_1 - Y_2|\ , \quad (7)$$

$$\theta = \sin^{-1}\left(\frac{A}{B}\right). \quad (8)$$

C. Distance from hip to room floor

Distance from hip to a room floor is calculated using (9), where A, B, C, and D are coefficients of the floor plane, and X, Y, and Z are the coordinates of the hip center.

$$\text{Distance} = \frac{AX + BY + CZ + D}{\sqrt{A^2 + B^2 + C^2}}\ . \quad (9)$$

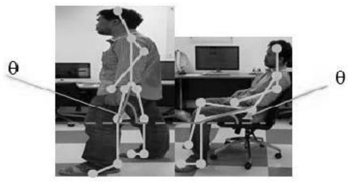

Fig. 6. Angles of left knee and right knee.

Fig. 7. Distance from hip to a room floor.

V. CLASSIFICATION DATA

Postural analysis used attribute data from the data recorder. We selected 6 models including neural network (NN), decision tree (D-tree), naïve Bayes, support vector machine (SVM), random forest, and logistic regression to compare the quality of data analysis using 5-fold cross-validation. The Weka data mining tool was used to implement these six models.

A. Classification Models

1) Neural Network

In this research we used simple multilayer perceptron (MLP) which uses back-propagation algorithm in learning [22-23] for predicting class member of postures (*Stand, Sit, Sit on Floor,* and *Lie Down*). Structural of neural network in our experiment have three layers (input layer, hidden layer, and output layer) with 4, 10, and 4 nodes, respectively.

2) Decision Tree

D-tree is used to classify data from class label, which yields output as a flowchart-like tree structure [14, 17, 22-23]. The J48 is used in our work to classify the data as a set of decision nodes and leaf nodes. Each leaf node shows a class outcome label of postures.

3) Naïve Bayes

Naïve Bayes is a statistical classifier based on the Bayes's theorem, which predicts class membership based on conditional probability [22-23]. The nodes in a Bayesian model are created from the given training data. Building model times of naïve Bayes are extremely fast when compared to other methods.

4) Random Forest

Random forest is a classifier consisting of a collection of tree structured classifiers, which uses the tree bagging algorithm in learning model for predicting class member [24]. In our experiment, we used the default number of trees provided in Weka.

5) Logistic Regression

Logistic regression is the well-known statistical model that uses logistic function for predicting class member by comparing the probability values between the categorical dependent variable and independent variable of each instance [24]. In our experiment, we used the multinomial logistic regression algorithm in learning for predicting class member.

6) Support Vector Machine

SVM can classify both linear and nonlinear data [22-23]. The SVM learner also supports multiple-class problems by computing the hyper-plane between each class and the rest. We used SVM polynomial kernel to classify class member of postures. In our experiment, we used SVM with polynomial kernel to classify data set. The degree of polynomial kernel was 3.

VI. Results

This section compares the results of the six classification methods used in classifying different postures. In the experiment we used 5-fold cross-validation to determine performance of each method with the results as given in Table II. Differences of the building model time of each classification method are not significant in this experiment. The wrong predictions mainly occurred with *Lie Down* because Kinect library function sometimes gave error joint detections on the body while people lay down.

From Table II, the neural network, D-tree, random forest and support vector machine methods almost have same results and give better results than the naïve Bayes and logistic regression methods; and Max-Min normalization is the best one of all results for the D-tree. Therefore, we select the D-tree with Max-Min normalization in our postural classifications.

As a comparison, we also used data of twenty skeletal joints to build classification models without transformation. The results are shown in Table III. It is clear that using only transformed attributes based on domain knowledge provided better performance.

TABLE II. Result of each model with data transformation.

Non-Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.981	97.20 %	97.68 %
D-tree (J48)	0.971	97.30 %	97.32 %
Naïve Bayes	0.869	89.90%	89.90 %
SVM	0.967	97.10 %	97.47 %
Logistic Regression	0.961	96.00%	95.95 %
Random Forest	0.975	97.80 %	97.75 %
Max-Min Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.983	97.60 %	97.87 %
D-tree (J48)	0.991	97.40 %	97.88 %
Naïve Bayes	0.994	95.70 %	85.69 %
SVM	0.971	97.00 %	97.52 %
Logistic Regression	0.970	96.80 %	96.78 %
Random Forest	0.975	97.90 %	97.88 %
Z-score Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.981	97.00 %	97.68 %
D-tree (J48)	0.969	97.30 %	97.32 %
Naïve Bayes	0.899	89.00 %	89.90 %
SVM	0.969	97.00 %	97.50 %
Logistic Regression	0.961	96.40 %	95.95 %
Random Forest	0.975	97.80 %	97.80 %

TABLE III. Results with twenty skeletal joints without data transformation.

Non-Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.943	94.25 %	94.20 %
D-tree (J48)	0.981	96.20 %	96.16 %
Naïve Bayes	0.917	77.50 %	77.50 %
SVM	0.945	95.42 %	95.30 %
Logistic Regression	0.957	84.10 %	84.13 %
Random Forest	0.997	94.50 %	90.01 %
Max-Min Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.944	94.35 %	94.35 %
D-tree (J48)	0.981	96.20 %	96.19 %
Naïve Bayes	0.917	77.50 %	77.50 %
SVM	0.952	95.70 %	95.47 %
Logistic Regression	0.841	95.70 %	84.13 %
Random Forest	0.937	92.40 %	91.43 %
Z-score Normalization			
Model	ROC Area	Sensitivity	Accuracy
NN	0.945	94.25 %	94.25 %
D-tree (J48)	0.985	97.00 %	97.03 %
Naïve Bayes	0.918	77.70 %	77.74 %
SVM	0.940	95.60 %	95.20 %
Logistic Regression	0.841	92.70 %	84.13 %
Random Forest	0.965	92.80 %	91.46 %

VII. Conclusion and future work

In this paper, we presented a system for postural detection and classification of elderly people while watching televisions. We used Kinect camera to record postures of elderly people. The input data have four attributes (angles of the left knee and right knee, distance from hip to a room floor, and aspect ratio). The appropriate distance of the camera from the person is about 1.8 to 3.0 meters. In the experiment, we selected six models to compare the results of postural classification and selected optimal suitable classifier by using 5-fold cross-validation. The six classification methods were NN, SVM, D-tree, logistic regression, random forest and naïve Bayes. The most suitable model was found to be D-tree with Max-Min normalization of the transformed.

In the near future, we plan to include a wider range of postures such as half-lying down, slouching, and so on, into our postural classification system so that we could learn more about the behaviors of the elderly people while they watch televisions at home. These obtained data would be very useful to their family members, physicians, and other health practitioners. Moreover, we could include a facial recognition system to make us understand the elderly people's expressions and emotions even better, say, while they watch comedy shows or suspense movies. Above all, more and more technologies could be used to monitor and keep records of the daily lives of the elderly people, especially while being at home, to help us better understand and innovate technologies to serve them. At the end of the day we may ask ourselves, "Are we prepared for the imminent ageing society yet?"

Acknowledgment

This work is supported by the Higher Education Research Promotion and National Research University Project, Thailand's Office of the Higher Education Commission, and School of Information Technology at KMUTT. We thank all volunteers and the SIT's D-Lab students and staff for their kind help.

References

[1] United Nations, "World Population Ageing: 1950-2050," [Online] www.un.org/esa/population/publications/worldageing19502050, [May 1, 2014].

[2] United Nations, "World Population Ageing 2007," [Online] www.un.org/esa/population/publications/WPA2007/wpp2007.htm, [May 1, 2014].

[3] Thailand's National Statistical Office, "Information Services," [Online] service.nso.go.th/nso/nso_center/project/table/files/S-elderly/2550 /000/00_S-elderly_2550_000_000000_00200.xls, [May 16, 2014].

[4] K. Reid, "Lifeline or Leisure?: TV's Role in the Lives of the Elderly," [Online] http://www.medialit.org/reading-room/lifeline-or-leisure-tvs-role-lives-elderly, [May 1, 2014].

[5] M. Graney and E. Graney, "Communications Activity Substitutions in Aging," Journal of Communication, Vol. 24, Issue 4, 1974, pp. 88–96.

[6] S. Kensaku, "Older People and Television Viewing in Japan," Journal of NHK Broadcasting Studies, No. 8, 2010, pp. 63-94.

[7] B. Oestlund, B. Jönsson and P. Waller, "Watching Television in Later Life: A deeper understanding of the meaning of TV viewing for design in geriatric contexts," Scandinavian Journal of Caring Sciences, Vol. 24, Issue 2, 2010, pp. 233-243.

[8] N. Sachiko, K. Toshiyuki and M. Emi, "Television as the Most 'Relaxing' Medium: From the Time Use Survey on Television and Moods," Journal of NHK Broadcasting Studies, No. 8, 2010, pp. 95-122.

[9] K. Nakajima, H. Matsui and K. Sasaki, "Telemonitoring system of television's operating state for remotely located families," Information Technology Applications in Biomedicine (ITAB2007), the 6th International Special Topic Conference on, Tokyo, Japan, Nov. 8-11, 2007, pp. 186-188.

[10] J. Choomkasean, P. Mongkolnam and J. H. Chan, "Multimedia Delivery for Elderly People: A Conceptual Model," The 7th International Conference on Advances in Information Technology (IAIT2012), Bangkok, Thailand, Dec. 6-7, 2012, pp. 58-69.

[11] Microsoft, "MSDN – the Microsoft Developer Network," [Online] http://msdn.microsoft.com, [May 1, 2014].

[12] J. Webb and J. Ashley, "Beginning Kinect programming with the Microsoft Kinect SDK," [E-book], Apress©, pp. 1-321.

[13] Microsoft Develper Network, "Kinect for Windows Architecture" [Online] http://msdn.microsoft.com/en-us/library/jj131023.aspx, [16 June, 2014].

[14] O. Patsadu, C. Nukoolkit and B. Watanapa, "Human Gesture Recognition Using Kinect Camera," Computer Science and Software Engineering (JCSSE), 2012 International Joint Conference on, Bangkok, Thailand, May 30, 2012-Jun. 1, 2012, pp. 28-32.

[15] Y. Liu, Z. Zhang, A. Li and M. Wang, "View Independent Human Posture Identification using Kinect," Biomedical Engineering and Informatics (BMEI), 2012 the 5th International Conference on, Chongqing, China, Oct. 16-18, 2012, pp. 1590-1593.

[16] S. Kaenchan, P. Mongkolnam, B. Watanapa, and S. Sathienpong, "Automatic Multiple Kinect Cameras Setting for Simple Walking Posture Analysis", The 17th International Computer Science and Engineering Conference, Bangkok, Thailand, Sep. 4-6, 2013, pp. 250-254.

[17] P. Paliyawan, C. Nukoolkit and P. Mongkolnam, "Prolonged Sitting Detection for Office Workers Syndrome Prevention Using Kinect," 2014 Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON), 2014 the 11th International Conference on, Nakhon Ratchasima, Thailand, May 14-17, 2014.

[18] F. Cardile, G. Iannizzotto and F. La Rosa, "A Vision-Based System for Elderly Patients Monitoring," Human System Interactions (HSI), 2010 3rd Conference on, Rzeszów, Poland, May13-15, 2010, pp. 195-202.

[19] M. Fahim, I. Fatima, S. Lee and Y. K. Lee, "Daily Life Activity Tracking Application for Smart Homes using Android Smartphone," Advanced Communication Technology (ICACT), 2012 the 14th International Conference on, PyeongChang, South Korea, Feb. 19-22, 2012, pp. 241-245.

[20] A. Correa, A. Morell, M. Barcelo and J. L. Vicario, "Navigation system for elderly care applications based on wireless sensor networks," Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European, Bucharest, Romania, Aug. 27-31, 2012, pp. 210-214.

[21] J. T. Kim, J. Y. Soh, S. H. Kim and K. Y. Chung, "Emergency Situation Alarm System Motion using Tracking of People Like Elderly Live Alone," Information Science and Applications (ICISA), 2013 International Conference on, Suwon, South Korea, Jun. 24-26, 2013, pp. 1-4.

[22] J. Han and M. Kamber, "Data mining concepts and techniques," Morgan Kaufmann Publishers, 2nd Edition, 2006, pp. 285-378.

[23] The University of Waikato, "Weka 3-Data Mining with Open Source Machine Learning Software in Java," [Online] http://www.cs.waikato.ac.nz/ml/weka/, [May 1, 2014].

[24] T. Hastie, R. Tibshirani and J. Friedman, "The Elements of Statistical Learning Data Mining, Inference, and Prediction", 2nd Edition, 2008, pp. 119-603.

ภาคผนวก ค
ผลงานที่ได้รับการตีพิมพ์

Multiple-stage Classification of Human Poses while Watching Television

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan Hoyin Chan

2014 2nd International Symposium on Computational and Business Intelligence (ISCBI-II 2014)

December 7-8, 2014

Multiple-stage Classification of Human Poses while Watching Television

Thammarsat Visutarrom, Pornchai Mongkolnam, and Jonathan H. Chan
Data and Knowledge Laboratory
School of Information Technology
King Mongkut's University of Technology Thonburi
Bangkok 10140, Thailand
E-mail: jonathan@sit.kmutt.ac.th

***Abstract*— We compared the accuracy measure between a single-stage classifier model and a multiple-stage classifier model in postural classifications using Kinect. Postural training sets were collected from Kinect's skeletal data streams, based on some of the common human postures during television watching. Three types of training sets were used, including Kinect's raw skeletal training set, skeletons with attribute selection training set, and skeletal position transformation training set. We selected four learning models, namely, neural network, naïve Bayes, logistic regression, and decision tree, for learning our data sets and classifying a testing set to find the appropriate learning model. The best accuracy value of our experiment was 87.68 % by using skeletal position transformation training set with neural network. In the future, we will apply our technique and methodology to track elderly behaviors while they are watching television.**

Keywords- postural classification; Kinect; data transformation; multiple-stage classifier; television watching

I. INTRODUCTION

Human posture classification is a very challenging research which has been an ongoing study for a long time by researchers in many different fields such as medical science, sport science, and social science. Current technology allows for better human posture classification with a high accuracy. In 2010, the Kinect system was released by Microsoft. Soon afterwards, it had become popular among many researchers for human body detection and human postural classification, largely due to its capability in separating the human body from a complex background. In addition, Kinect has a built-in software library that helps to separate each skeletal joint and enables the analysis of a whole skeleton. It can detect up to twenty joints, as shown in Fig. 1, each with three positional coordinates X, Y and Z. Thus there are a total of 60 raw attributes for each image or frame.

Unlike most other researchers, we use Kinect to classify human postures while watching television. In this paper, we investigate two classification architectures: a single-stage classifier model and a multiple-stage classifier model. Both models are given three types of training sets, including raw skeletal training set, skeleton with attributes selection training set, and skeletal position transformation training set. The data mining techniques used in our experiment include four learning models of back propagation neural network (NN), naïve Bayes (NB), logistic regression (LR), and decision tree (J48).

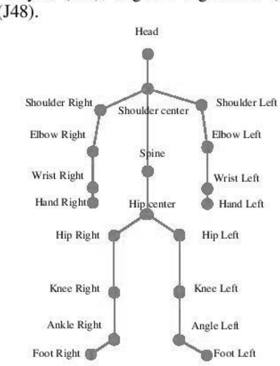

Figure 1. Twenty joints obtained from Kinect.

II. LITERATURE REVIEW

Human posture detection can be separated into two techniques: 2D image and 3D image processing [1-3]. The 2D image processing was popular in the past because it used pixel color information to analyze the data. Consequently, the performance of its postural classification heavily depended on environmental factors such as light conditions and background complexity.

Using 3D image processing can decrease the background complexity problem that occurs with 2D images. The reason being 3D image processing provides depth data that can help to better separate each object. In addition, 3D image processing with infrared sensors could be used in dark rooms, making it suitable for even complex human postural analyses.

In 2003, Cohen and Li [4] presented the human postural classification with SVM modeling using the 3D visual-hull constructed from a set of silhouette input data. The system returned the classified human body postures in the form of thumbnail images.

In 2007, Wu and Aghajan [5] proposed the method of human posture estimation in a multiple camera network by using the concept of an opportunistic fusion framework that comprised three dimensions of space, time, and feature levels to obtain a 3D human skeleton.

In 2010, Kinect [6-9] was launched by Microsoft. It was a 3D camera game controller which was controlled by the human gestures of a player. Currently there are two software libraries for analyzing skeletal joints, i.e., Microsoft SDK and OpenNI. The OpenNI library can analyze 15 main joints of a human body. The Microsoft SDK library has two detection modes. The first mode is the default stand tracking that can analyze 20 joints of the body, and the second mode is the seated tracking mode that can analyze 10 upper shoulder joints of the human body. Kinect has three input components, including depth image sensors, RGB camera, and a multi-array microphone as shown in Fig. 2. With the ability to separate the human body from a complex background and in analyzing the human skeletal joints, Kinect has been applied by many researchers in their studies since its inception.

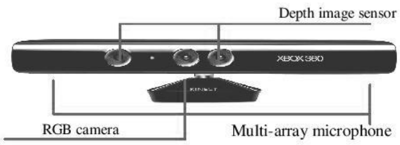

Figure 2. Kinect components [10].

In 2011, Htike and Khalifa [11] presented a real-time gesture classification system to classify the dancing gestures from moving skeletal joint data obtained from Kinect. The accuracy of their system was 96.9% using the approximate 4-second record of the skeletal motion.

In 2014, Visutarrom et al. [12] presented a system for simple postural detection and classification of elderly people while watching television. Their experiments have four standard postures including stand, sit, sit on floor and lie down. They selected six models to compare the results of postural classification and selected the most suitable classifier by using 5-fold cross-validation. The six classification methods used were neural network, support vector machine, decision tree, logistic regression, random forest and naïve Bayes. The best accuracy was 97.88%, obtained by decision tree with Max-Min normalization technique.

In 2014, Dai et al. [13] proposed a machine learning and vision-based method for elderly fall detection using statistical human posture sequence modeling. A series of laboratory simulated falls and activities of daily livings (ADLs) were performed and recorded by Kinect. Hidden Markov Models were used for modeling the fall posture sequences and distinguishing different fall activities and ADLs. The average fall recognition rate was above 80%.

III. PROCESS OVERVIEW

Our experiment can be separated into two parts: data preparation and data classification. The workflow is shown in Fig. 3. Data preparation is used to prepare the training sets before the learning of each model. Data classification is to compare the postural classification result of each training-set against a testing dataset. We use Kinect XBOX 360 with the Microsoft SDK library version 1.8 with the default stand tracking mode to detect a human body having 20 skeletal joints. The distance for test subjects is between 1.8-3.0 meters away from Kinect during the data collection process.

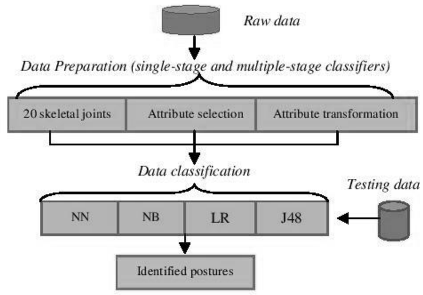

Figure 3. Experimental workflow.

Training sets were recorded from seven subjects, including four males and three females. Their ages are between 26-65 years old, with height between 150-180 centimeters and various body sizes. Our training sets have a total of 110,751 instances. The testing data has 20,868 instances obtained from another three subjects. Their ages are between 30-45 years old, with height between 155-180 centimeters. The details of number of instances in each posture for both training set and testing set are shown in Table I. There are a total of 18 postural class memberships as summarized in Table I. These postures are also illustrated in Fig. 4 as four main class memberships of stand, sit, sit on floor and lie down.

TABLE I. NUMBER OF INSTANCES IN EACH POSTURE.

Posture	Number of instances	
	Training set	Testing set
Normal stand	6584	985
Stand both hands up	6325	1445
Stand left hand up	6575	856
Stand right hand up	5892	1345
Stand lean forward	6320	964
Sit straight back	5981	869
Sit lean forward	5888	1250
Sit lean backward	5123	958
Sit both hands up	5320	1345
Sit left hand up	6120	945
Sit right hand up	5369	1450
Sit on floor straight back	5842	1065
Sit on floor lean forward	6923	1425
Sit on floor lean backward	6320	954
Sit on floor both hands up	6790	1356
Sit on floor left hand up	6952	1475
Sit on floor right hand up	5602	1185
Lie down in various manners	6825	996

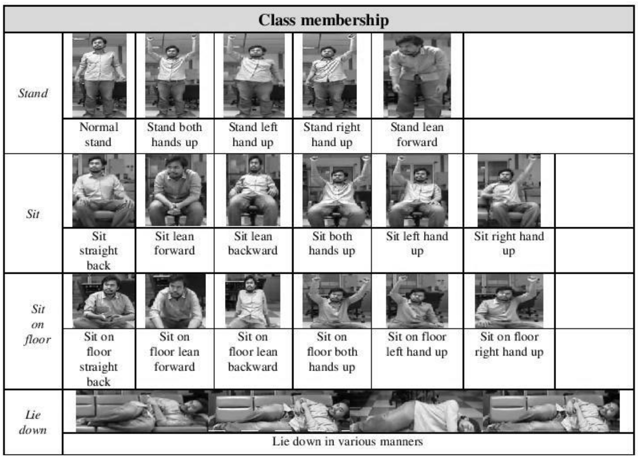

Figure 4. Class membership.

A. *Data preparation*

This step is to prepare the training sets before being used in learning of each model. In our experiment, we considered two types of classifier architectures. First is a single-stage classifier and second is a multiple-stage classifier.

A single-stage classifier model uses all training data packaged in one set. Therefore, the single-stage classifier has only one model in the prediction data as shown in Fig. 5 (a). Consequently, if the training data are large, it is very time consuming in creating the model.

On the other hand, a multiple-stage classifier would partition the data into multiple training sets in the first stage, and that would help to reduce the time to create the model, depending on the number of partitioned training sets. Therefore, a multiple-stage classifier requires many models working together for prediction as shown in Fig. 5 (b).

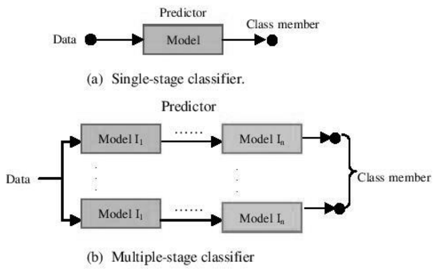

Figure 5. Single-stage classifier and multiple-stage classifier.

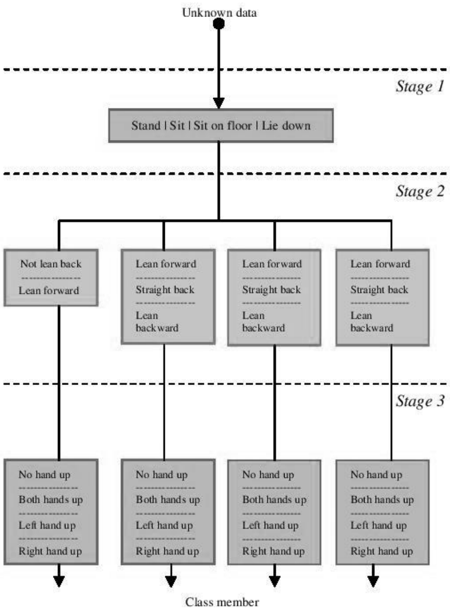

Figure 6. Multiple-stage classifier model.

In this work, we separate the classification task according to target classes of training. The classifier is divided into three stages as shown in Fig. 6. The first stage is used to classify four main postures, including stand, sit, sit on floor, and lie down. The second and third stages are used to classify sub-postures. In particular, the second stage is used to classify the *back* posture. The *back* posture is used to classify lean forward, straight back, or lean backward. The third stage is used to classify *arm* posture. The *arm* posture is used to classify hand-up, including both hands up, left hand up, and right hand up.

Both single-stage classifier and multiple-stage classifier have three types of training data set as follows.

1) Raw skeletal training set is the training data which use all 20 skeletal joint positions. Each joint has three coordinates X, Y, and Z. Each coordinate is considered one attribute; therefore, this training set has 60 total attributes.

2) Skeleton with attribute selection training set is the reduced number of attributes and reduced time for finding relationships between attributes. One advantage of the atrribute reduction (feature selection) is that less time is needed in creating the model. We compare two-attribute selection techniques in order to reduce the dimensionality from the 60 attributes. They are the gain ratio attribute evaluator and the information gain attribute evaluator. The ranker method of feature selection is used in our training set to obtain the top 10, 20, 30, 40, 50 attributes as features. In this work, both the single-stage classifier and the multiple-stage classifier use the same ranker attributes.

3) Skeletal position transformation training set is the training set with transformed data from the skeletal joints. Using raw data may make training set have complexity that affect in learning of classifier and result in error accumulation due to redundant attributes. Attributes transformation is a technique that help to reduce the complexity of attributes in training set. Therefore, we select this technique to compare accuracy with other training sets. Equations used in transformed attributes are the equations that are used to observe the changes of each part of the human body. Transformation data can help to reduce the dimensionality of the problem. In addition, transformation data can help increase the relationships in training and increase the accuracy of prediction of the learning model. We select 12 skeletal joints to create nine attributes in the training set, including shoulder center, shoulder left, shoulder right, elbow left, elbow right, wrist left, wrist right, hip center, hip left, hip right, knee left, and knee right. The nine attributes include angle knee left, angle knee right, aspect ratio of height and width, distance from hip to room floor, back status, hand_SWL, hand_SWR, hand_SEL, and hand_SER.

For stand, sit, sit on floor and lie down postures, they have different aspect ratio of the width and the height. Therefore, the aspect ratio of width and height was selected to be one of the attributes for classification. We found the aspect ratio can help to classify lie down posture from other postures as well. However, using the aspect ratio only cannot give a high accuracy with all postures. We added three transformed attributes including angle of left knee, angle of right knee, and distance from hip to the room floor for improving the prediction of stand, sit and sit on floor. The angles of left knee and right knee can help to identify stand and sit. When people stand, the angles at their knee joints are larger. On the other hand, when they sit, the angles of their knees are smaller. Distance from hip to room floor can help to identify sit and sit on floor. When people sit on floor, distance between hip to room floor will be less than sit on sofa and chair. In addition, we classify hand up and back characteristics of each posture for postural classification to provide even more details. Back status is used to classify the backward characteristics, including lean back forward, straight back and lean back backward. hand_SWL, hand_SWR, hand_SEL, and hand_SER are used to classify hand up of each posture that has 3 characteristics including hand up right, hand up left and both hands up. The equations used to find the values of those attributes are given below in Eqns. (1) – (12).

Aspect ratio of width and height

$$Width = Max_X - Min_X, \tag{1}$$

$$Height = Max_Y - Min_Y, \tag{2}$$

$$Aspect\_Ratio = \frac{Width}{Height}. \tag{3}$$

Max and Min is the maximum and minimum values of X and Y coordinates of only five skeleton joints from head, shoulder left, shoulder right, foot left, and foot right.

Angles of left knee and right knee

$$A = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2 + (Z_1 - Z_2)^2}, \tag{4}$$

$$B = |Y_1 - Y_2|, \tag{5}$$

$$\theta = \sin^{-1}\left(\frac{A}{B}\right). \tag{6}$$

X_1, Y_1 and Z_1 are coordinates of hip left and right. X_2, Y_2 and Z_2 are values of knee left and right. Angles of left knee and right knee can help to identify stand and sit.

Distance from hip to room floor

$$Distance = \frac{AX_3 + BY_3 + CZ_3 + D}{\sqrt{A^2 + B^2 + C^2}}. \tag{8}$$

X_3, Y_3 and Z_3 are the coordinates of the hip center. A, B, C and D are coefficients of the floor plane. Distance from hip to room floor can help to identify sit and sit on floor.

Back

$$Back\ status = Z_4 - Z_3. \quad (7)$$

Z_4 is Z coordinate of shoulder center and Z_3 is Z coordinate of hip center. Back is used to classify the backward characteristics that include lean back forward, straight back and lean back backward.

Hand up

$$Hand\_SWL = Y_4 - Y_6\,, \quad (9)$$

$$Hand\_SWR = Y_5 - Y_7\,, \quad (10)$$

$$Hand\_SEL = Y_4 - Y_8\,, \quad (11)$$

$$Hand\_SER = Y_5 - Y_9. \quad (12)$$

Y_4 and Y_5 are the Y coordinates of shoulder left and shoulder right. Y_6 and Y_7 are the Y coordinates of wrist left and wrist right. Y_8 and Y_9 are Y coordinates of elbow left and elbow right. All attributes in (9)–(12) are used to classify hand up right, hand up left and both hands up. Hand_SWL and Hand_SWR are differences between the Y coordinates of shoulder and wrist of the left hand side and the right hand side, respectively. Hand_SEL and Hand_SER are differences between the Y coordinates of shoulder and elbow of the left hand side and the right hand side, respectively.

B. Data classification

In this work, we select four learning models, including neural network, naïve Bayes, logistic regression, and decision tree; all of which are available in the Weka data mining tool [14].

Neural network [15-16] used in the experiment is the simple multilayer perceptron (MLP) which uses the back-propagation algorithm in learning. Structure of the neural network in the experiment comprises three layers, including input layer, hidden layer, and output layer. The input and output layer values depend on the amount of attributes and the amount of class memberships of each training set. Hidden layer value used is the default value provided by the Weka data mining tool.

Decision tree [15-16] is used to classify data from class label, yielding the output as a flowchart-like tree structure. In the experiment, we used J48 algorithm to classify class membership as a set of decision nodes and leaf nodes. Each leaf node represents a class outcome label.

Naïve Bayes [15-16] is a simple probability classifier based on the Bayes' theorem with the assumption of independence between every pair of features. The nodes in the Bayesian model are created from given training data. Each node counts the number of rows per attribute value per class for nominal attributes and calculates the Gaussian distribution for numerical attributes.

Finally, we also used the multinomial logistic [15-16] regression algorithm for predicting class member and use the default parameter provided in Weka data mining tool.

IV. Results

Table II provides the average accuracy result of each training model, against the testing data. We found that the skeletal training set of both the single-stage classifier model and multiple-stage classifier model produced poor results with the prediction testing set falling between 19% and 26%. Attributes selection of both the single-stage classifier model and multiple-stage classifier model increased the accuracy of the prediction testing set slightly. However, the results were still poor in the range of 15% to 34%, with better performance falling in the top 30 ranked attributes with the values between 22% and 34%. Interestingly, we found Y coordinate was ranked in the top group that was important in our training set and X coordinate was ranked in the unimportant group with the training set. On the other hand, the skeletal position transformation training set produced better values than the skeletal training set and the skeleton with attributes selection training set. The skeletal position transformation training set of the single-stage classifier gave the accuracy results between 40% and 60%, and skeletal position transformation training set of the multiple-stage classifier gave the best results in the range of 79% to 87%. The learning model that gave the best results was neural network with skeletal position transformation training set using a multiple-stage classifier, with the average accuracy result of 87.68%.

The accuracy of each individual postural prediction using the skeletal position transformation training set, against the testing set, is shown in Table III. The results show that the performance of the individual posture models is quite similar except for the noticeably poorer accuracy for the lie down posture. The best classifier was neural network in all cases. This is consistent with previous works in medical data classification using classic machine learning techniques that show neural networks performs at least as well as other techniques and quite often outperforms the others [17]. One possible reason in our case is that there are more adjustable parameters that can be tuned in Weka.

In summary, attributes transformation is the technique that helped reduce the complexity of attributes in training set and increase the accuracy of the data prediction of each model. In addition, the reduced number of attributes helped decrease the time needed to create the classifier model. Most error prediction occurred with the lie down posture. In addition, we found the error of predicting postures occurred while people changed postures, or during the transition period. As a result, postural transitions should also be considered in order to detect and predict accurate postures while watching television.

V. Conclusions and Future Work

In this paper, we are interested in studying the differences of accuracy values of the single-stage classifier model and multiple-stage classifier model of postural classification. Postural data were recorded from Kinect. From the experiment results, the data transformation training set of multiple-stage classifier with the neural network learning model gave the best results.

In the near future, we want to apply our knowledge and methodology found herein for developing the postural classification of elderly people while watching television, which is the main pastime/activity of the elderly. More sub-postures such as stretching hands to the sides may be added. Being able to detect postures and subsequent behaviors can help their family members, physicians, and other health practitioners understand the elderly even more. As a result, better services and technologies could be created and offered to improve their quality of lives.

ACKNOWLEDGMENT

This work has been supported by the Higher Education Research Promotion and National Research University Project, Thailand's Office of the Higher Education Commission, and School of Information Technology at KMUTT. We thank all volunteers and D-Lab students and staff for their kind help and useful discussions.

REFERENCES

[1] Y. Sun, M.H. Li, J.S. Hu and E.L. Wang, "2D RECOVERY OF HUMAN POSTURE," Machine Learning and Cybernetics, 2002. Proceedings. 2002 International Conference on, Beijing, China, Nov. 4-5, 2002, pp. 1638 - 1640.

[2] K. Takahashi and S. Sugakawa, "Remarks on human posture classification using self-organizing map," Systems, Man and Cybernetics, 2004 IEEE International Conference on, Netherlands, Oct. 10-13, 2004, pp. 2623 - 2628.

[3] C.C. Li and Y.Y. Chen, "Human Posture Recognition by Simple Rules," Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on, Southampton, United Kingdom, Apr. 10-12, 2006, pp. 351 - 356.

[4] I. Cohen and H. Li, "Inference of Human Postures by Classification of 3D Human Body Shape", Analysis and Modeling of Faces and Gestures, 2003. AMFG 2003. IEEE International Workshop on, Nice, France, Oct. 17, 2003, pp. 74 – 81.

[5] C. Wu and H. Aghajan, "Model-based Human Posture Estimation for Gesture Analysis in an Opportunistic Fusion Smart Camera Network", Advanced Video and Signal Based Surveillance, 2007. AVSS 2007. IEEE Conference on, London, England, Sept. 5-7, 2007, pp. 453 – 458.

[6] Microsoft, "Kinect for Windows Architecture," [Online] http://msdn.microsoft.com/en-us/library/jj131023.aspx, [Sep. 18, 2014].

[7] M. Ruta, F. Scioscia, M. Di Summa, S. Ieva, E. Di Sciascio and M. Sacco, "Semantic matchmaking for Kinect-based posture and gesture recognition," Semantic Computing (ICSC), 2014 IEEE International Conference on, Hue Royal City, Vietnam, Aug. 1-3, 2012, pp. 309 - 312.

[8] Y. Liu, Z. Zhang, A. Li and M. Wang, "View Independent Human Posture Identification using Kinect," Biomedical Engineering and Informatics (BMEI), 2012 the 5th International Conference on, Chongqing, China, Oct. 16-18, 2012, pp. 1590-1593.

[9] H.P.H. Shum, E.S.L. Ho, Y. Jiang and S. Takagi, "Real-Time Posture Reconstruction for Microsoft Kinect," Cybernetics, IEEE Transactions on, vol. 43, Oct. 2013, pp. 1357 - 1369.

[10] Xbox, "Kinect for Xbox 360," [Online] http://www.xbox.com/en-US/kinect, [Sep. 18 2014].

[11] K. K. Htike and O.O. Khalifa, "Comparison of supervised and unsupervised learning classifiers for human posture recognition", Computer and Communication Engineering (ICCCE), 2010 International Conference on, Kuala Lumpur, Malaysia, May 1-12, 2010, pp. 1-6.

[12] T. Visutarrom, P. Mongkolnam, and J. H. Chan, "Postural Classification using Kinect," 2014 International Computer Science and Engineering Conference, Khon Kaen, Thailand, Jul. 30 - Aug. 1, 2014, pp. 346 - 351.

[13] X. Dai, M. Wu, B. Davidson, M. Mahoor and Z. Jun, "Image-Based Fall Detection with Human Posture SequenceModeling," Healthcare Informatics (ICHI), 2013 IEEE International Conference on, Philadelphia, Pennsylvania, USA, Sept. 9-11, 2013, pp. 376 - 381.

[14] The University of Waikato, "Weka 3-Data Mining with Open Source Machine Learning Software in Java," [Online] http://www.cs.waikato.ac.nz/ml/weka/, [May 1, 2014].

[15] T. Hastie, R. Tibshirani and J. Friedman, The Elements of Statistical Learning – Data Mining, Inference, and Prediction, 2nd ed. Springer, 2008, pp. 119-603.

[16] J. Han and M. Kamber, Data mining concepts and techniques, 2nd ed. Morgan Kaufmann Publishers, 2006, pp. 285-378.

[17] S.Dreiseitl and L. Ohno-Machado, "Logistic regression and artificial neural netowrk classification models: a methodology review," J. Biomed. Inform. vol. 35, pp. 352-359, October 2002.

TABLE II. AVERAGE ACCURACY OF 18 POSTURES USING DIFFERENT FEATURE SELECTION METHODS

Single-Stage classifier					Multiple-stage classifier				
	NN	NB	LGT	J48		NN	NB	LGT	J48
All 20 joints	***25.78***	19.79	22.80	20.89	**All 20 joints**	***26.89***	18.57	20.04	18.65
Gain 10	16.89	15.89	21.90	***20.70***	**Gain 10**	17.89	11.87	18.76	***19.00***
Gain 20	***32.78***	23.45	25.55	25.89	**Gain 20**	***24.56***	17.02	19.01	20.01
Gain 30	***33.89***	26.50	27.45	28.61	**Gain 30**	***34.05***	22.34	28.23	22.45
Gain 40	***28.45***	22.00	25.12	24.87	**Gain 40**	***27.46***	21.03	26.45	20.34
Gain 50	***24.30***	21.00	20.56	20.86	**Gain 50**	***26.04***	20.67	24.08	19.04
Info-Gain 10	***16.70***	15.81	21.90	20.76	**Info-Gain 10**	***22.49***	17.57	21.00	17.48
Info-Gain 20	***32.06***	22.80	25.45	26.89	**Info-Gain 20**	***23.93***	20.89	22.48	19.78
Info-Gain 30	***32.45***	24.32	25.45	24.31	**Info-Gain 30**	***29.59***	21.89	23.90	21.90
Info-Gain 40	***28.76***	22.04	25.50	25.12	**Info-Gain 40**	***25.67***	20.56	22.56	19.87
Info-Gain 50	***23.50***	21.45	19.55	21.56	**Info-Gain 50**	***23.90***	19.34	20.45	19.87
Attribute transformation	***62.78***	50.34	60.47	58.46	**Attribute transformation**	***87.68***	79.89	85.67	84.69

TABLE III. ACCURACY OF EACH POSTURE USING TRANSFORMED ATTRIBUTES

Single-Stage classifier					Multiple-stage classifier				
Posture	NN	NB	LGT	J48	Posture	NN	NB	LGT	J48
Normal stand	**66.23**	52.14	62.04	60.12	Normal stand	**89.36**	83.26	88.65	86.56
Stand both hands up	**66.23**	52.02	61.59	60.37	Stand both hands up	**89.56**	83.25	88.58	86.45
Stand left hand up	**66.78**	52.01	61.89	59.34	Stand left hand up	**89.53**	83.14	88.56	85.14
Stand right hand up	**67.32**	51.48	61.59	59.32	Stand right hand up	**89.66**	82.36	87.48	85.36
Stand lean forward	**61.65**	51.21	61.32	59.38	Stand lean forward	**89.79**	82.34	86.32	85.47
Sit straight back	**60.25**	51.42	61.56	59.12	Sit straight back	**88.69**	80.24	86.64	85.36
Sit lean forward	**61.08**	50.12	61.25	58.36	Sit lean forward	**88.89**	80.53	86.32	85.69
Sit lean backward	**65.36**	51.28	61.01	58.01	Sit lean backward	**87.96**	81.56	86.95	85.36
Sit both hands up	**61.45**	51.24	61.04	58.47	Sit both hands up	**87.21**	79.45	85.36	85.32
Sit left hand up	**62.54**	50.18	60.48	58.21	Sit left hand up	**87.45**	79.36	85.99	85.15
Sit right hand up	**63.51**	50.13	60.48	58.12	Sit right hand up	**86.35**	79.85	85.86	84.26
Sit on floor straight back	**61.79**	50.48	60.78	58.02	Sit on floor straight back	**87.69**	79.45	85.36	84.89
Sit on floor lean forward	**60.35**	50.23	59.62	58.15	Sit on floor lean forward	**87.65**	79.25	84.26	84.56
Sit on floor lean backward	**62.35**	50.02	59.89	58.36	Sit on floor lean backward	**87.36**	78.54	84.38	84.36
Sit on floor both hands up	**61.23**	50.45	59.42	58.69	Sit on floor both hands up	**87.45**	78.56	84.35	84.57
Sit on floor left hand up	**65.25**	50.36	59.24	58.25	Sit on floor left hand up	**87.35**	78.98	84.15	83.53
Sit on floor right hand up	**61.48**	50.23	59.67	58.36	Sit on floor right hand up	**87.39**	78.36	84.28	83.63
Lie down in various manners	**55.21**	41.2	55.63	53.68	Lie down in various manners	**78.98**	69.58	78.64	78.79

ประวัติผู้วิจัย

ชื่อ – สกุล	นาย ธรรมศาสตร์วิสุทธารมณ์
วัน เดือน ปีเกิด	1 กรกฎาคม 2532
ประวัติการศึกษา	
ระดับมัธยมศึกษา	โรงเรียนภูเก็ตวิทยาลัย พ.ศ. 2550
ระดับปริญญาตรี	วิศวกรรมศาสตรบัณฑิต สาขาวิชาวิศวกรรมคอมพิวเตอร์ มหาวิทยาลัยสงขลานครินทร์ พ.ศ. 2554
ระดับปริญญาโท	วิทยาศาสตรมหาบัณฑิต สาขาวิชาวิศวกรรมซอฟต์แวร์ มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าธนบุรี พ.ศ. 2557
ทุนการศึกษา หรือทุนวิจัย	ทุนอุดหนุนสนับสนุนการวิจัยระดับบัณฑิตศึกษา ทบวงมหาวิทยาลัย ปีงบประมาณ 2556

ผลงานที่ได้รับการตีพิมพ์

1. ธรรมศาสตร์ วิสุทธารมณ์, พรชัย มงคลนาม และโจนาธาน โฮ ยิน ชาน, 2557, "การจำแนกท่าทางขณะรับชมโทรทัศน์โดยใช้กล้องคิเนค," **The 10th National Conference on Computing and Information Technology**, ครั้งที่ 10, 8-9 พฤษภาคม 2557, โรงแรมอังสนาลากูนา, อ. ภูเก็ต, จ. ภูเก็ต, หน้า 583-588 .

2. Visutarrom, T., Mongkolnam, P. and Chan, J. H., 2014, "Postural Classification using Kinect," **The 2014 International Computer Science and Engineering Conference (ICSEC2014)**, July 30 - August 1, Pullman Khon Kaen Raja Orchid, Khon Kaen, Thailand, pp. 346-351.

ผลงานที่ได้รับการตีพิมพ์

3. Visutarrom, T., Mongkolnam, P. and Chan, J. H., 2014, "Multiple-stage Classification of Human Poses while Watching Television," **2014 2nd International Symposium on Computational and Business Intelligence (ISCBI-II 2014)**, December 7-8, The Ashtan Sarovar Portico, New Delhi, India, (Accepted).