VMware as an Intermediate Platform between Windows 7 and Linux Fedora 15 for Real-Time Malaysian Sign Language Translator

Sara Bilal, Rasheed M. Nassr, and Rini Akmeliawati

***Abstract*—This research introduces a virtual interface between Windows 7 and LINUX Fedora 16 based on Virtual Machine ware (VMware) for real-time Malaysian Sign Language (MSL) translation into text and/or voice (in English). The developed method is based on HTK, Gt2k under LINUX Fedora 16 and VC++ 2010 and OpenCV pre 1.1 library under Windows7. The communication between client (Windows7) and server (LINUX Fedora 16) has been established using VMware. The main significance of this approach is that the best characteristics of both operating systems LINUX Fedora 16 and Windows7 have been utilized. Under Windows7, Visual C++ 2010 combined with OpenCV pre 1.1 library supports video processing algorithms and has a power graphical user interface. Meanwhile, the Gt2k for gesture recognition is fully supported under LINUX. Therefore, a client/server technology has secured much time during the MSL recognition system development and helped in terms of algorithms enhancement.**

***Index Terms*—VMware, VC++, MSL, gesture recognition, HMM.**

I. Introduction

Sign Language (SL) is a highly structured non-verbal language utilizing both manual and non-manual communication. Manual communication consists of movements and orientation of hand/arm that convey symbolic meaning while non-manual communication involves mainly facial expression, head movement, body posture and orientation which help in augmenting the meaning of the manual signs [1].

The original difficulty of SL recognition is aggravated by the fact that variety of SL sets exist in the world for even a single language such as Malay or English. Each country has its own symbols and gestures for SL. Even regions within the same country have their own SL as claimed by the Philippine Federation of the Deaf [2]. Thus, it is impossible for users of different SL groups to understand each other. So, by defining a vision-based system that is capable of providing recognition and interpreting of signs from different regions will be useful for hearing/speech impaired communities. But in general, human to Computer Interaction (HCI) applications, particularly automatic SL recognition is a challenging problem in the domain of image processing and computer graphics whereas tremendous efforts are required for translating the lexical form of hand gestures (signs) and developing the algorithms that scale effectively to large vocabularies.

To recognize and translate a SL, mainly four stages are required to build a complete Automatic Sign Language Translator (ASLT) as shown in Fig. 1.

There are diverse approaches and techniques developed to recognize American SL (ASL) [3], [4], Japanese SL (JSL) [5], etc. But there are limitations and drawbacks in each SL technique in terms of real-time implementation. Further research and investigation are required to address the limitations and improve the existing systems. Furthermore, there are commercial HCI applications, which have reached real-time system implementation level as in [6]-[8]. The commercial products is mainly meant for gaming purposes and extensive use of an advanced hardware (camera or screen) thus cannot be used as a standalone for SL recognition especially in terms of database size and the SL structure.

In this research, a client/server approach between Windows7 and LINUX Fedora 16 with the help of VMware workstation has been developed for real-time SL recognition. The following sections describe the methodology adopted in this work.

II. System Setting

A. Sharing Files between Client and Server

Client/Server is a two separated programming modules with explicitly distinctive tasks which communicate through network [9]. Usually clients can be categorized into two types: thin client and thick client. Thin client is capable to achieve acceptable computation performance over wide area networks [10]. However, for the purpose of this study thick client (VC++ 2010 and OpenCV pre 1.1 library) which residual in Windows7, has been used. The use of thick client is significant is this research during the preprocessing stages of signs' videos which requires a high visualization performance. This video processing functionality may exceed regular functionalities of thin client.

Similarly, in the server side, Gt^2k is a tool for gesture recognition which helps HCI developers to focus mainly on pre-recognition stages rather than spending time and efforts in building the HMM based recognizer.

1) Thick client

In this work, the algorithms for face and hands detection, skin blobs tracking up to feature extraction stage have been

Manuscript received March 1, 2014; revised May 5, 2014. This work was supported by the Research Matching Grant Scheme RMGS 09-03.

The authors are with International Islamic University Malaysia, Faculty of Engineering, Department of Science, Kuala Lumpur, Malaysia (e-mail: sarra@iium.edu.my, rakmelia@iium.edu.my)

completed under Windows7 environment using VC++ version 2010 with the help of OpenCV library version pre 1.1. To train HMM in an offline mode, features' files have been transferred to the GT2k for training a specific number of signs.

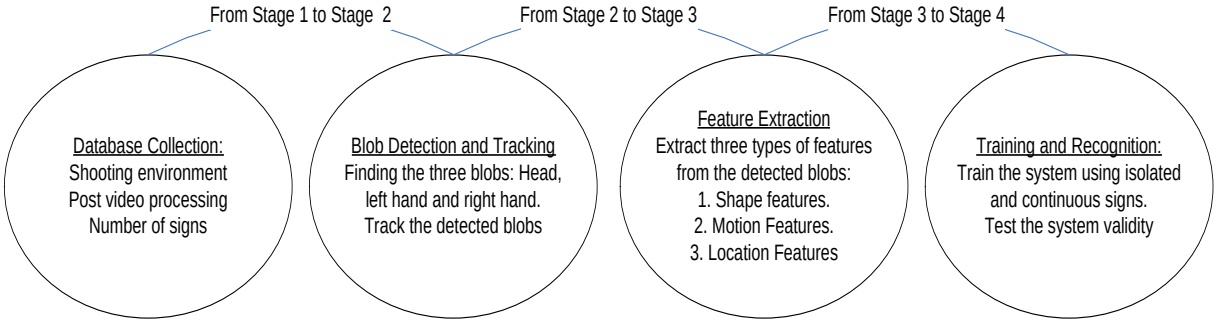

Fig. 1. ASLT system overview.

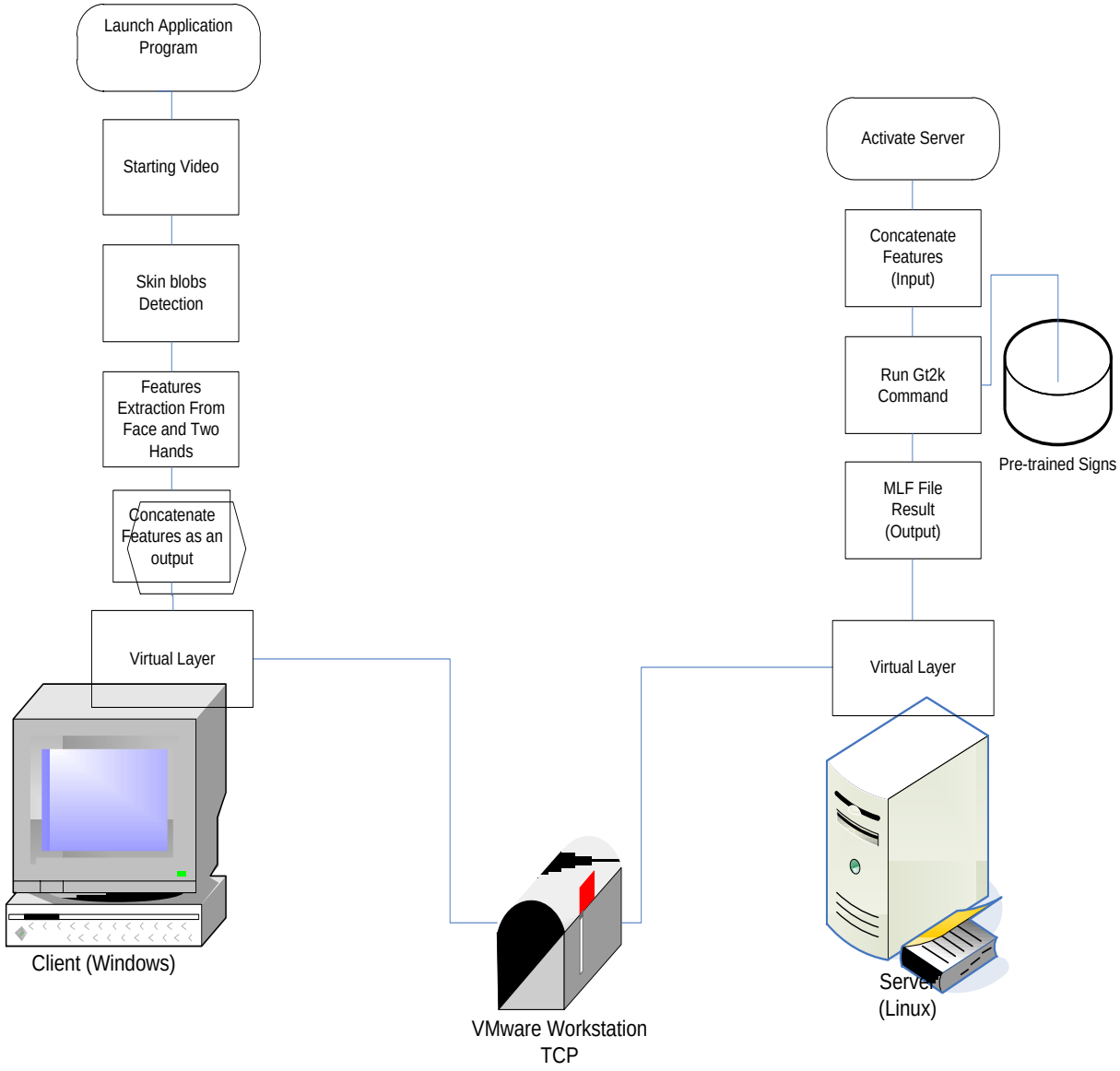

Fig. 2. Overall real-time MSL recognition system.

2) Server for SL recognition

To achieve the SL recognition using a new, non-trained sign; a recognize.sh script can be used from the Gt2k. Four arguments need to be passed; data files 'signs.txt', a file to store the recognition results 'results.txt' which will be created automatically, 'options.sh' and the trained HMM model. HMM model is called new Macros and the output will be in the form of a Master Label File (MLF).

The format of the MLF file will have gesture ranked by likelihood score as shown in Table I which states MLF for five isolated words. The command for recognition using Gt2k is as follows:

Recognize.sh signs. txt result.txt options.sh NewMacros.

As stated previously, establishing network connection is necessary to begin an interaction between client and server. It is required to establish a communication which enables features collector 'client' to send features file to the analyzer 'server'.

B. Thick Client and Server Interface

VMware Workstation was lunched as a media on Windows7 to install Fedora 16. Following that, a transmission Control Protocol (TCP) connection type is created between client and server to transfer the processed video file which is an output from client to be input to the server. Likewise, TCP transfers the resulted output from server to be input for client as illustrated in Fig. 2.

The aforementioned real-time system implementation structure is shown in Fig. 2 and is described as follows:

First, server has to be activated using java scripts as shown in Table II.

TABLE I: MLF RECOGNITION RESULTS

Results from Gt2k Recognizer
#!MLF!#
"/mnt/javacode/test1.txt.rec"
0 172000 car -21368.128906
0 172000 noodles -11883843.000000
0 172000 dog -14732120.000000
0 172000 chicken -18813332.000000
0 172000 school -19096242.000000

TABLE II: SERVER ACTIVATION IN LINUX

Java Script

import java.io.*;
import java.net.*;
public class Fileserver {
private static ServerSocket serv;
private static Socket client;
//private static File myFile;
public static void main(String arg[]){
try {
int port=4444;
serv=new ServerSocket(port);
while(true)
{
try{
//wait for Connection
System.out.println("Waiting for connection on port "+port);
client=serv.accept();
String msg=receiveFile();
client.close();
client=serv.accept();
System.out.println("send results... ");

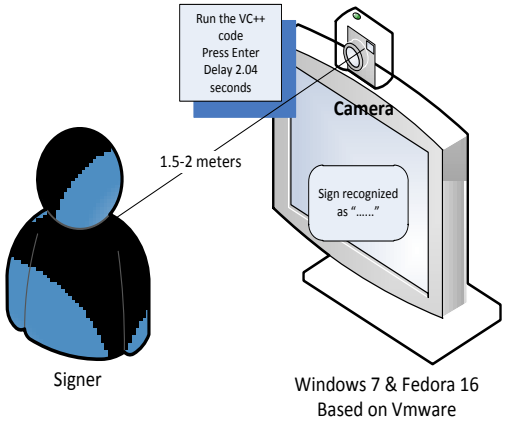

Fig. 3. MSL recognition system interface.

Second, the main program which processes the input video files has to be launched. Following that, features files will be propagated as soon as the client successfully concatenates all required features in one file.

Third, the server will activate the pre-trained Gt2k code as soon as the propagated features files completely received using java scripts.

Fourth, the MLF result as shown in Table I, will be send back to the client when Gt2k finishes analyzing the inputted sign from the client.

Finally, the client will read the 3rd line from the MLF file and display it on the screen. Furthermore the system translates the text into voice using voice recognition libraries from Microsoft. Net.

III. REAL-TIME SYSTEM IMPLEMENTATION

The real-time MSL recognition system has been developed by concatenating the aforementioned system stages as described in Fig. 2. The system has been tested and evaluated during each stage to test the possibility of real-time interface. The interface to the system has been established in Windows7 (client) and the recognition is performed in LINUX Fedora 16 (server) as explained in Section II. The user interface is shown in Fig. 3.

A. System Efficiency and Performance

The MSL recognition system has been tested and evaluated for real-time implementations from video input stage up to the translating of signs into text and/or voice (in English). The system translates the text into voice using voice recognition libraries from Microsoft .Net. Dell Precision T7500 Tower Workstation with 64-bit multi-core Intel® Xeon® processors, each with Intel® Quick Path technology (i.e. each processor core features an integrated memory controller and high-speed interconnects) has been used [11]. The system performance and efficiency has been evaluated based on the following measurements:

1) Processing time: The developed system works on 30 frames/second from Minoru 3D web camera. The processing time has been measured through offline training and real-time recognition as follows:
   a) Training the system offline for 20 isolated signs takes an average time of 1.89 seconds. As well, training the system for sentences of two sets (172 and 202), approximately, an average time of 21.85 seconds has taken place under LINUX Fedora 16.
   b) During the real-time system implementation, the developed system takes about 85-90 ms to process the blob detection and feature extraction stages under Windows7. The overall system processing time starting from video input stage up to the translation of isolated sign into text and/or voice (in English) takes about 2.05 seconds.
2) System Accuracy: Table III shows the average recognition accuracy for 20 isolated signs reached up to 80% and 55% for a total lexicon of 37 words in 20 sentences.
3) Robustness: The most crucial stage for the system to be

robust is the face and hands detection and tracking stage. The proposed system for skin detection is robust because it highlights the skin colour dynamically under different environments. Therefore, the system has reached a level of robustness during the initialization stage which can be considered as the most crucial stage for this work.

4) Cost: The cost has been calculated based on the prices of the software and hardware required to design the whole system. VMware cost about RM 620 (Ringgit Malaysian) [12], where the powerful computer Dell Precision T7500 Tower Workstation that has been used to establish a client/server application with high speed efficiency cost around RM 7,690.
5) End-User: The system is user friendly where only a camera and computer are required with a simple interface as shown in Fig. 3.

B. Recognition Accuracy Comparison of Isolated and Continuous Signs

Various methods have been developed using different approaches for SL recognition as listed in Table III. Each developed system has its own settings, merits and restrictions. In this work, most of the settings such as the environment and the signing flexibility were set to a minimum. This was done intentionally so that a more natural way of communication between the hearing/speech impaired people and hearing ones could be established using the proposed MSL recognition system. Therefore, the proposed system have been built, tested and enhanced based on the collected database. This implies that the SL recognition system could be applicable for a portable device. Table III shows the comparison of the proposed system with some other existing ones. The comparison is done only with the existing systems which have relatively same settings as the proposed one. Using bare hands with no coloured gloves or electronic ones was one of these required conditions. Among these systems as shown in Table 3, Elmezian *et al*., in [13], [14] used a static velocity between the signs which is quite far from our assumption that system should have a natural way of signing during the database collection of isolated signs and their system reached an accuracy of 92.5-98%.

ASLT system in [5] reached up to 73% recognition accuracy, but the researchers used very little vocabulary and limited for only one hand signs.

A huge set of signs, 163 isolated and continuous, were presented in [15], but they had fixed a motion transition between signs during the collection and recognition of the sentences.

The developed MSL recognition system has been tested using 20 words and lexicons of 37 words in 20 sentences (see Appendix A). The system recognition accuracy reached up to 55%-80% for continuous and isolated signs, respectively.

C. Significance of the Proposed Client/Server Technique

In this work, virtual interface between Windows 7 and LINUX Fedora 16 based on VMware for real-time Malaysian Sign Language (MSL) translation has been developed. By using VMware, it is not required to transfer one of the two programs; either the Gt2k from LINUX Fedora 16 to Windows7 or the VC++ 2010 and OpenCV pre 1.1 library from Windows7 to LINUX Fedora 16. In general, VC++ 2010 program cannot be transferred from Windows7 to LINUX Fedora 16 operating system because it is a Microsoft product with a powerful functionality for developers only under Windows, specifically in developing codes for video and image processing area. Qt is another open source product which can execute image and video visualizing under LINUX and could be another alternative for VC++. But Qt is a new programming language with lack of resources for developers [16].

TABLE III: RECOGNITION ACCURACY COMPARISON OF THE DEVELOPED SYSTEM AND THE EXISTING SYSTEMS

No	Works	Methodology	Target language	2/3 D	Restrictions	Accuracy %	Vocabulary size	Pitfalls	Advantages
1	Mahmoud Elmezian et al. [13][14]	Left-right banded Hmm/spatio temporal features	ArSL	3	Static velocity as threshold	98.94 isolated and 95.7continuous	10 (isolated and continuous)	Limited vocabulary	spatio temporal features
2	Kobayashi and Haruyama [5]	PHMM/new Pattern matching method	JSL	2	One hand	73 isolated	6 isolated	Single hand signing/very little vocabulary.	--
3	Holden et al. [15]	HMM/ invariants features	Auslan	2	One hand	99 isolated , 97 continuous	163 (isolated and continuous)	Fixed Motion between signs recognized as a sign	Significant vocabularly size
4	Developed MSL Recognition	HMM/new blob extraction method and new features matching method	MSL	2	The human upper body must be visible in the scene	55continuous 80 isolated	lexicons of 37 words in 20 continuous sentences and 20 isolated words	Fails, if no face or hand detected	-Not restricted by background. -More natural (eg. One and two hands, no pauses between signs, flexible velocity). -Real-time translation of signs into text and/or voice (in English).

Far from the proposed client/server connectionist, there is another alternative method to link these two different operating systems (Windows7 and LINUX Fedora 16) physically. A physical link using cable between client and server computers could be attached. But this method is lacking in terms of the speed of data transfer through the physical line and connection stability. Also the cost is so crucial because two different computers are required to build such system. But the proposed system was launched on one powerful computer as mentioned earlier with the two operating systems (Windows7 and LINUX Fedora 16) in the same computer. This is possible by using the VMware workstation software.

IV. Conclusion

A real-time MSL translator has been developed for lexion of 37 words in 20 sentences and 20 isolated words. The system accuracy has been enhanced through the entire developed stages. For real-time SL recognition, an interface between Windows7 and Linux Fedora has been established. The main significant objective behind this approach is that the best characteristics of both operating systems LINUX Fedora 16 and Windows7 have been utilized. Under Windows7, Visual C++ 2010 combined with OpenCV pre 1.1 library supports video processing and has a power graphical user interface. Meanwhile, the Gt2k for gesture recognition is fully supported under LINUX. Therefore, a client/server technology has secured much time during the MSL recognition system development and helped in terms of algorithms enhancement. For future work, The recognition of continuous SL using HMM needs modification in terms of the sequence observations where larger number of sentences have to be collected only to allow the system to reflect the desired ones. Therefore, a method to reduce the number of database for training the sentences should be developed. The developed MSL recognition system can be built into a portable device such as android systems that can help to minimize the communication barriers between the hearing/speech impaired people and the hearing people.

Appendix A

TABLE A.1:Sample of 202 Sentences Used for Training Signed by Hearing/Speech Impaired Person

Signer 2 Sentences		
bag hit far father	fish like red snake	cat like far bag
bag hit far water	food buy white food	cat lose red scissor
blood buy white person	food see red father	chicken buy red book
blood see far fish	he love white person	chicken lose far water
book buy far school	he see far snake	dog like far snake
book want red dog	husband buy far snake	dog love far pen
bread like yellow chicken	husband like sharp scissor	father like far school
bread love white rice	i hit red chicken	father love yellow person
bus love red cat	i want red cat	fish see white pen
bus want red rice	pen hit red book	food see far rice
cat love white husband	pen want far scissor	food want far bus
cat see white dog	person buy yellow bus	he buy yellow chicken
fish see far snake	Person want far bread	husband hit yellow rice
car see sharp scissor	rice like white blood	i hit far snake
chicken buy white pen	rice see red food	i love red pen
chicken like white rice	school love sharp chicken	pen hit yellow scissor
dog want white snake	school see far school	pen see yellow bread
dog want yellow cat	scissor like yellow water	person hit white rice
father buy red bag	scissor want white fish	person see yellow chicken
father like far book	snake love white water	rice like red food
fish see yellow blood		rice lose red car
food love white bus		
food want red car	snake want yellow bread	school like red bag
he hit yellow cat	they buy yellow dog	school love red school
he love yellow water	they buy sharp father	scissor lose white blood
he want white bag	water hit far dog	snake see white blood
husband love white school	water see sharp cat	snake want yellow food
husband see sharp rice	we buy white water	they lose yellow blood
i hit yellow snake	we see far bread	they want red snake
i like white bread	we see white chicken	water buy red car
i like yellow pen	you hit far fish	water hit white husband
pen like far bread	you hit sharp car	we like white father
pen want white book	you love yellow father	we love far rice
person buy red fish	bag like white pen	you see far fish
person lose far bag	blood lose yellow food	bag love yellow person
rice love white chicken	book see red scissor	bag want white bag
school hit yellow bag	bread see far rice	blood see red blood
school like yellow scissor	bus want red book	book like far bag
scissor buy far husband	car lose white bag	bread lose yellow bread
scissor buy red blood	cat hit yellow dog	bus hit red bus
snake love yellow bread	chicken see yellow cat	bus see white bread
snake want red school	dog buy yellow blood	car see far car
they love far bus	father buy white car	cat want red cat
they love red rice	fish love far water	chicken hit yellow scissor
water hit yellow food	food like yellow chicken	dog like yellow dog
water want far person	he want yellow chicken	dog lose far car
we hit far bus	husband want white father	father lose white fish
we buy red scissor	i lose far water	fish want far husband
you like yellow dog	pen love white husband	food love red chicken
you want red blood	person love red bus	he lose red blood
bag hit white pen	rice hit far school	husband like red food
bag see yellow bus	school buy far fish	i want yellow bus
blood like yellow husband	scissor hit red snake	pen hit far pen
blood want sharp pen	snake like red person	person love white person
book hit white fish	they lose red snake	rice buy red dog
book love yellow school	water want far bread	rice lose far rice
bread hit white dog	we hit white father	school buy far school
bread like red blood	you like far fish	scissor buy white school
bus buy far rice	you love yellow car	snake buy yellow snake
bus love sharp school	bag see yellow book	they see yellow book
car like yellow person	bag want yellow water	water like white water
car see sharp bus	blood love far person	we love far pen
cat buy red car	blood want far husband	you hit white rice
cat love white bag	book buy white dog	you see white scissor
chicken want yellow chicken	book lose white pen	car want yellow bus
dog love red bag	bread buy white father	father like red father
dog want far rice	bread lose red blood	fish like red bag
father hit sharp snake	bus buy white school	
	bus hit white cat	
	car love far fish	

TABLE A.2: 20 Isolated Words Used for Evaluating the System Performance

Isolated signs
Big, blood, car, chicken, coffee, dog, dress, LRT, noodles, school, street, university, bag, beat, green, like, throw, white, yellow, water

Acknowledgment

The authors would like to thank the International Islamic University Malaysia (IIUM) for their support under the Research Matching Grant Scheme RMGS 09-03, and Malaysian Federation of the Deaf (MFD) for providing facilities, including interpreters and signers for the recordings.

References

[1] S. C. Ong and S. Ranganath, "Automatic sign language analysis: A survey and the future beyond lexical meaning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 873-891, 2005.

[2] P. F. O. T. Deaf, "Filipino sign language: A compilation of signs from regions of the Philippines," Part 1. In L. P. Services (Ed.), Quezon City, 2005.

[3] T. Starner, A. Pentland, and J. Weaver, "Real-time American sign language recognition using desk and wearable computer based video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1371-1375, 1998.

[4] T. E. Starner, "Visual recognition of American sign language using hidden Markov models," Technical Report, Tr-306, Cambridge Dept of Brain and Cognitive Sciences: Massachusetts Inst of Tech, 1995.

[5] T. Kobayashi and S. Haruyama, "Partly-Hidden Markov Model and its Application to Gesture Recognition," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 1997, pp. 3081-3084.
[6] E. Krzeslo, X. Baele, G. Pinault, and M. Tombroff. (2007). [Online]. Available: http://www.softkinetic.com/Solutions/iisuSDK.aspx.
[7] G. Shmuel, I. Katz, A. Shenfeld, E. Talmor, E. Aharonson, Z. Katz, and L. Rostock. (2005). [Online]. Available: http://www.eyesight-tech.com/
[8] M. Kolsch and M. Turk, "Robust hand detection," in *Proc. IEEE Intl. Conference on Automatic Face and Gesture Recognition*, 2004.
[9] S. M. Lewandowski, "Frameworks for component-based client/server computing," *ACM Computing Surveys*, vol. 30, no. 1, pp. 3-27, 1998.
[10] A. M. Lai and J. Nieh, "On the performance of wide-area thin-client computing," *ACM Transactions on Computer Systems*, vol. 24, no. 2, pp. 175-209, 2006.
[11] Dell Precision T7500. (2012). *Dell Precision T7500 Tower Workstation*. [Online]. Available: http://www.dell.com/my/business/p/precision-t7500/pd
[12] VMware Workstation 8. (2012). [Online]. Available: http://www.vmware.com/products/workstation/overview.html
[13] M. Elmezain, A. Al-Hamadi, J. Appenrodt, and B. Michaelis, "A hidden Markov model-based continuous gesture recognition system for hand motion trajectory," in *Proc. 19th IEEE International Conference on Pattern Recognition (ICPR)*, 2008, pp. 1-4.
[14] M. Elmezain, A. Al-Hamadi, and B. Michaelis, "Real-time capable system for hand gesture recognition using hidden markov models in stereo color image sequences," *Journal WSCG*, vol. 16, no. 1, pp. 65-72, 2008.
[15] E. J. Holden, G. Lee, and R. Owens, "Australian sign language recognition," *Machine Vision and Applications*, vol. 16, no. 5, pp. 312-320, 2005.
[16] N. Westin. (2008). Opera is not based on Qt. *My Opera*. [Online]. Available: http://my.opera.com/kilsmo/blog/2008/01/29/opera-is-not-based-on-qt.

Sara Bilal obtained her B.Tech in electronic engineering from Sudan University of Science and Technology (SUST) in 2002, Sudan and her master in electrical, electronic and telecommunication engineering from University Technology Malaysia (UTM) in 2005, Malaysia. In 2013, Sara obtained her PhD in mechatronics from International Islamic University Malaysia, IIUM, Malaysia. She is currently an Assistant Professor in the department of science in engineering, IIUM. She was a lecturer at SUST from 2005-2009 and a research assistant (RA) in IIUM from 2009-2012. She is a member of IEEE and signal processing society since 2009 and a member in Intelligent Mechatronics System Research Unit at the International Islamic University Malaysia. Her research interests are in the areas of image and video processing and analysis, computer vision and pattern recognition.

Rini Akmeliawati was born in Jakarta, Indonesia. She obtained the bachelor of engineering (Honours) in electrical engineering from Royal Melbourne Institute of Technology (RMIT) University, Australia in 1997, and her PhD in electrical and electronics engineering from the University of Melbourne, Australia in 2002. She is currently a professor of the Department of Mechatronics Engineering at the International Islamic University Malaysia. Previously, she was a lecturer at Royal Melbourne Institute of Technology (RMIT) University, Australia (2001–2004) and Monash University (2004–2008). Her main research interests include nonlinear control systems theory and applications, intelligent systems and image processing. She has published more than 80 international journal and conference papers. Rini Akmeliawati is a Senior Member of IEEE. She was the chair of IEEE Instrumentation and Measurement Society—Malaysia Chapter 2007–2009, the treasurer and the vice-chair of the same society in 2010 and 2011, respectively, and the secretary of IEEE Control Systems Society—Malaysia Chapter in 2010. She is also currently the Chair of Intelligent Mechatronics System Research Unit at the International Islamic University Malaysia.

Rasheed M. Nassr is a PhD candidate at IIUM, Malaysia, originally from Yemen and he was born in 1980. He is interested in web semantic, HCI and programming. Currently he is working on web semantic project related to Islamic websites. Also, he worked on projects related to sign language, meta-data. He has published many papers related to his PhD which is focusing on web semantic.