

Sign Language to Voice Recognition: Hand Detection Techniques for Vision-Based Approach

Noor Saliza Mohd Salleh^{*,1}, Jamilin Jais¹, Lucyantie Mazalan¹, Roslan Ismail¹, Salman Yussof¹, Azhana Ahmad¹, Adzly Anuar¹, Dzulkifli Mohamad²

¹ UNITEN, College of Information Technology, Universiti Tenaga Nasional, Km 7, Jalan Kajang-Puchong, 43009 Kajang, Selangor, Malaysia.

² UTM, Faculty of Computer Science and Information System, Universiti Teknologi Malaysia, 81310 UTM Skudai, Johor Darul Takzim, Malaysia

The use of gestures as means to convey information is an important part of human communication. The automatic recognition of gestures enriches human-computer interaction by offering a natural and intuitive method of data input. Automated systems for aiding those with impaired hearing have recently been one of the major areas of research. In sign language, hand gesture is one of the typical methods of non-verbal communication for human beings and we naturally use various gestures to express our own intentions in everyday life. Study has been made to structure an initial understanding and to identify the next step of function development. Hand detection is a fundamental step in many practical applications as gesture recognition, video surveillance, and multimodal machine interface and so on. For the flexibility and useful features, vision based technique has been proposed for gesture data collection. This paper will present research progress and findings on techniques and algorithms for hand detection as it will be used as an input for gesture recognition process.

Keywords sign language recognition; hand detection; vision-based

1. Introduction

The ability to detect a person unconstrained hand in a natural video sequence has applications in sign language, recognition and human computer interaction.

There are two ways to collect gesture data for recognition. Device based measurement which measures hand gestures with equipment such as data gloves which can archive the accurate position of hand gestures as its position is directly measured. Second is vision-based technique, which can cover both face and hands signer in which signer does not need to wear data gloves device. All processing tasks are solved by using computer vision techniques which are more flexible and useful than prior approach.

Since sign language is gesticulated fluently and interactively like other spoken languages, a sign language recognizer must able to recognize continuous sign vocabularies in real-time. We are trying to build such a system for the Bahasa Melayu Sign Language or Bahasa Isyarat Malaysia (BIM).

A basic definition, gestures are usually understood as hand and body movement which can pass information from one to another. Since we are interested in hand gesture and so the term 'gesture' is always referred to the hand gesture in this paper.

The key points will be described in the following sections. Section 2 and 3 will reviewed some related works that has been done in specific area, sign language recognition and hand detection. Then, our current stage of development, our hand detection technique will be explained and also feature detection that has been extracted from hand segmentation. Current phase of our system development and future direction also will be discussed.

2. Related Works

* Corresponding author: e-mail: noorsaliza@uniten.edu.my, Phone: +603-89212020

Attempts on machine vision-based sign language recognition have begun being published only recently with relevant literature since several years ago.

Most attempts to detect hands from video place restrictions on the environment. For examples, skin colour is surprisingly uniform [1, 2], so colour-based hand detection is possible [3]. However, this by itself is not reliable modality. Hands have to be distinguished from other skin-coloured objects and these are cases of sufficient lighting conditions, such as coloured light or grey-level images. Motion flow information is another modality that can fill this gap under certain conditions [4], but example for non-stationary cameras this approach becomes increasingly difficult and less reliable. Statistical information about hand locations is effective when used as a prior probability [5], but it requires application-specific training.

Eng-Jon Ong and Bowden [1] presented a novel, unsupervised approach to training an efficient and robust detector which applicable of not only detecting the presence of human hands within an image but classifying the hand shape. Their approach is to detect the location of the hands uses a boosted cascade of classifiers to detect shape alone in grey scale image. A database of hand images was clustered into sets of similar looking hands using the k-medoid clustering algorithm that using a distance metric based on shape context. A tree of boosted hand detectors was then formed, consisting of two layers, the top layer for general hand detection, whilst branches in the second layer specialize in classifying the sets of hand shapes resulting from the unsupervised clustering method. Tested the detector with an unseen database of 2509 images has given 99.8% success rate and shape classifier with 97.4% success rate.

Kolsch and Turk did a study on view-specific hand posture detection with an object recognition method recently proposed by Viola and James. First, they demonstrate the suitability of the integral-image approach to the task of detecting hand appearances. Then, the qualitative measure is presented that amounts to an a priori estimate of 'detectability', alleviating the need for compute-intensive training. Finally, parameters of the detection methods are optimized, achieving significant speed and accuracy improvements. They suggested that most convex appearances with internal grey-level variation are better suited the purpose of detection with rectangle feature-classification method.

The Korean Manual Alphabet (KMA) by Jung-Bae Kim, Kwang-Hyun Park and Z.Zenn Bien [3], present a vision-based recognition system of Korean manual alphabet which is a subset of Korean Sign Language. KMA can recognize skin-coloured human hands by implemented fuzzy min-max neural network algorithm using Matrox Genesis imaging board and PULNIX TMC-7 RGB camera.

Feng-Sheng Chen, Chih-Ming Fu, Chung-Lin Huang [4] introduced a hand gesture recognition system to recognize 'dynamic gesture' of which a gesture is performed singly in complex background using 2D video input. The system tracks the moving hand and analyzes the hand-shape variation and motion information as the input to the HMM-based recognition system. They come out with the experimental result of 4-state HMM has proved to generate the best performance for modelling the gesture. Each input image sequence is pre-processed by hand region extraction process for contour information and coding and have been implemented using two methods: (1) only contour information and (2) using combined contour information and motion information. The extracted information is converted to vector sequences and then quantized into symbol sequences for both of the training and recognition processes. Totally 1200 image sequences are collected for 20 different gestures, thus each kind of gesture with 60 sequences in average, in training phase and other 1200 sequences are collected for test. For method (1), recognition rate of using training data for testing is 97%, and the recognition rate of using testing data is 90.5%, meanwhile for method (2), recognition rate of using training data for testing is 98.5% and the recognition rate of using testing data rises to 93.5%.

3. Hand Region Detection

Hand detection is a preliminary step to a number of applications including HCI, surveillance, gesture recognition, hand tracking and understanding human-human interactions.

Hand detection exists in the preliminary stage in image processing. Our researches found out there are several techniques of detecting hand region.

Generally hand detection comprises of subtracting the signers image from the image background. Finding hand region using colour images is not difficult because the nature of skin colour has its own unique value and can easily being processed. Skin colour model will be used to detect pure hand image from complex environment. However, since skin colour is influenced by luminance and shadow, HSL colour model or RGB colour model do not give good performance. To luminance-free image, normalized RGB has been adopted. Some will compare the colours of the extracted skin regions with sample skin colour extracted from training images. These systems fail under poor lighting conditions or for skin colours they are not trained for.

To date, various colour models have been proposed for skin colour detection, e.g., the CbCr colour space is used in [8,9], the RGB colour space is used in [10]. The HS (Hue, Saturation) colour space is used in [11], normalized RGB and HSV (Hue, Saturation, Value) in [12] and etc. Dealing with coloured images will make the process of finding a threshold value to detect the hand become easier. However, dealing with greyscale input image forced us to automate input threshold value to measure the information thus extracted the hand region from the whole image.

In addition, edge detection technique will be applied to separate the arm region from the hand region when the signers wearing short sleeves shirt. For the same scenario, in [12], the hand and arm region is thinned to obtain the skeleton of the region. The hand is then defined as the dense intersections region.

Getting input from all the findings, we have come out with our own steps in order to detect hand region as an input for the next step in images processing which is the recognition part.

4. Current Research

Looking back to the previous discussed planned method, we defined ourselves being in early stage of the image processing level which is hand detection. We are using the Matrox frame grabber in VB.Net platform to initially develop functions to capture input from signers and detect the hand region area. Our constraint is that our camera can only support input image in grey scale value. Therefore we have encountered several problems and facing new challenge on capturing and processing static and online or rather more to say as real-time processing. Thus, we did struggling on finding the best solution to detect hands region with best output quality.

The input images are captured by a Samsung CCTV camera placed on a table using Matrox frame grabber installed in the CPU, running on Intel Xeon CPU with 496 MB of RAM. Each image has a spatial resolution of 256 x 256 pixels and a grayscale resolution of 8 bit. As a result our system can process hand gestures at acceptable speed.

Given a variety of available image processing techniques and recognition algorithms, we have design our preliminary process on detecting the image as part of our image processing part. Hand detection reprocessing flow shows in figure 2.

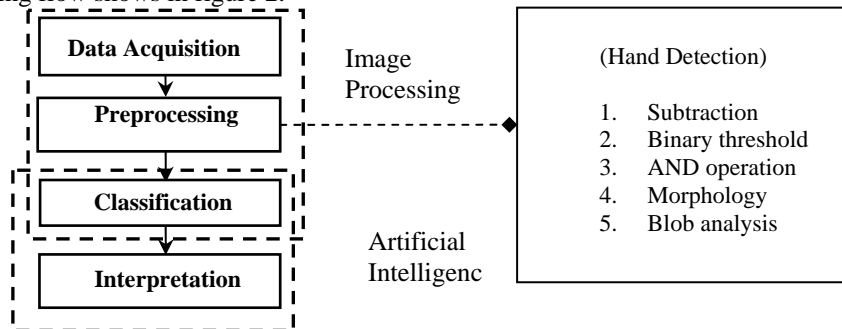


Fig 2 Basic Flow Concept of Recognition System (left). Hand detection pre-processing phase as a sub-phase in the whole recognition system. The 5 steps of hand detection techniques are circle by the straight line.

The system starts by capturing an image without any signers with a still camera setup towards a certain angle with black background. Captured background image then stored into the system in grey scale

image input. The next process will captured the background image with the existence of a stationary signer.

To proceed with hand detection, we extract a signer's region using background subtraction. By having the only image of the signer's, we now proceed with binary threshold. Automated function of input binary value is available as system not dealing with colour image acquisition which is easier to deal with skin colour for hand detection. Then, Bitwise AND operation sets the resulting bit to 1 if the corresponding bit in both image of result from subtraction and binary threshold is 1. Bitwise manipulation enhanced the wanted image of the hand region. Further processing with morphological filters may be used to clean up the segmented hand region. However, due to the nature of same hand and face colour, we might get the results of 3 regions of face and both hands. Blob analysis will identify connected regions of pixels within an image, and then calculated selected features of those regions. For this we use blob analysis and find the maximum two values of regions which are those two hand's blob.

As a result, we successfully detect the hand region. As for time being, attempt of this stage of analysis and experiment being discuss in this paper. The next stage of the image processing is yet to be discussed in future paper. Figure 3 shows the experiment results.

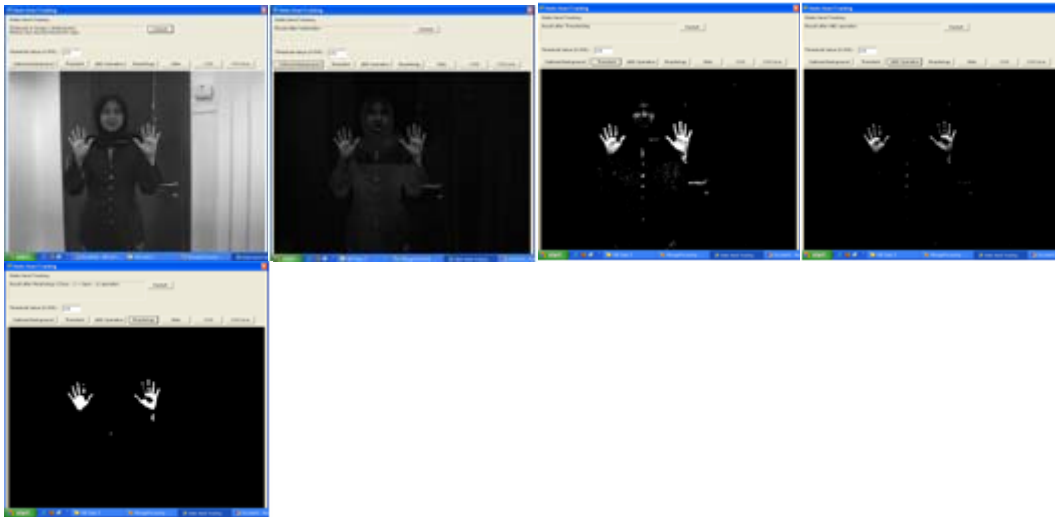


Figure 3 Hand Detection Result

5. Gesture Analysis

Two generally sequential tasks are involved in the analysis (see Fig 4). The first task involves 'detecting' or extracting relevant image features from the raw image or image sequence. The second task uses these image features for computing the model parameters. But in this section we will only discuss on the first task involved as per our current research stage.

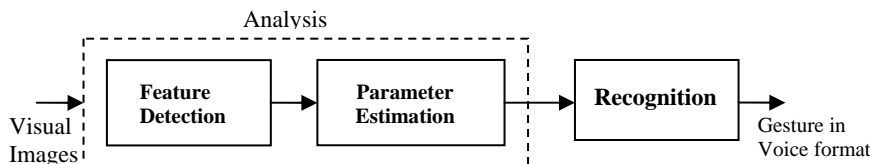


Figure 4 Analysis and recognition of gestures.

5.1 Feature Detection

As per current stage of development, we define 4 features for recognition of BIM words.

1. (X hand, Y hand) : The gravity center position of the hand region
2. A : The area of the hand region
3. Pm : The perimeter of the hand region
4. Θ motion : The direction of hand motion in the image coordinate
5. Θ hand : The direction of hand region in the image coordinate

Those features are calculated using moment derived shape descriptors [15]. The central moments are being calculated by applying the moment function, a built-in function of BlobAnalysis features in Matrox Imaging Library. This BlobAnalysis supports calculation of binary features, which is the result of our hand region detection.

More generally, the definition of moments m_{ij} refers to the pixel location x, y and pixel values $f(x, y)$ is defined by

$$m_{pq} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} X^p Y^q f(x, y) dx dy \quad (1)$$

The first-order moments in x and y, normalized by the area, yield the x and y centroids:

$x_c = \frac{M_{10}}{M_{00}}$ and $y_c = \frac{M_{01}}{M_{00}}$, respectively. These determine the region average location.

Principal major and minor axes are defined to be those axes that pass through the centroid, about which the moment of inertia of the region is, respectively, maximal or minimal [15]. Their direction is given by the expression

$$\frac{1}{2} \left(\frac{\mu_{02} - \mu_{20}}{\mu_{11}} \right) \pm \frac{1}{2\mu_{11}} \sqrt{(\mu_{02}^2 - 2\mu_{02}\mu_{20} + \mu_{20}^2 + 4\mu_{11}^2)} \quad (2)$$

The variables are defined as follows:

$$\mu_{11} = m_{01} - m_{10}m_{01} / m_{00} \quad (3)$$

$$\mu_{20} = m_{20} - m_{10}^2 / m_{00} \quad (4)$$

$$\mu_{02} = m_{02} - m_{01}^2 / m_{00} \quad (5)$$

The moment is given by below definition or equation using the image intensity. The captured binary image exclusively has pixel intensity values of "0" (background) and others for hand region (object).

$$M_{00} = \sum_x \sum_y I(x, y) \quad (6)$$

$$M_{10} = \sum_x \sum_y xI(x, y) \quad (7)$$

$$M_{01} = \sum_x \sum_y yI(x, y) \quad (8)$$

6. Discussion

The results of segmentation and feature detection are performed as explained above. Experimental result of 12 samples of hand image with different position for each left and right hand give the consistent result.

But, there are issues regarding the techniques and quality of the output from pre-processing that need to be further discussed:

6.1 Gray Scale Threshold for Hand Region Detection

Limitation of our Matrox camera giving us difficulties in separating hand region from other image including background. Hand region depends on skin colour non-influenced easily classified by value of RGB colour. But, by having input of grayscale, classifying the hand region are not easy and sometimes disturbed by other region which have similar grayscale value such as signer's forehead or face. Thus, a function of automating the threshold filter value takes place for this initial development to view the output result.

6.2 Quality and Efficiency on Processing for Online Gesture Detection

To implement the same methodology in analyzing real-time image processing, we have to apply multiple buffer programming in our system framework. This is our proposed solution to implement it for the online processing. And as a result, we manage to detect the hand region with additional functions that has to be automated such as threshold value input. This is somehow gives us difficulties which yet we have not faced in the next steps of image processing and even in recognition part. However, for this stage of point, we manage to get the output by implementing the proposed sequence for the image processing.

7. Conclusion and Future Direction

The features of the processed image are now ready to be input into the recognition phase. HMM has been proposed to be used for the recognition technique. But still, the HMM parameter needs to be identified and estimated using the result of the processed images, the features listed in Section 5. How the features are going to be implemented into the HMM are still not being defined yet.

We are currently working on defining all the variables from the sequence captured images in relate to the HMM parameters estimation. Even the features and parameters are ready; the recognition phase using HMM are large hurdles for actual deployment of such a system.

We hope understanding the HMM and estimation of the parameters can be achieved as our target for the next phase.

Acknowledgements This project was fully supported by Ministry of Science, Technology and the Environment under grant 07-99-03-10010-EAR.

References

- [1] M. J. Jones and J. M. Rehg, *Int. Journal of Computer Vision*, Jan 2002, 46(1): pp 81-96
- [2] D. Saxe and R. Foulds, In *Proc. IEEE Intl. Conference on Automatic Face and Gesture Recognition*, Sept. 1996, pp 379-384
- [3] X. Zhu J. Yang, and A. Waibel, , In *Proc IEEE Intl. Conference on Automatic Face and Gesture Recognition*, 2000.
- [4] R. Cutler and M. Turk, , In *Proc IEEE Intl. Conference on Automatic Gesture Recognition*, April 1998, pp 416-421.
- [5] T. Kurata, T. Okuma, M. Kouroggi, and K. Sakaue, In *Second Intl. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-time Systems*, July 2001.
- [8] R. L. Hsu, M. Abdel-Mottaleh and A. K. Jain, , *IEEE Tram. on Pattern Airalysir and Machine Intelligence*, vol.2, NOS, May 2002, pp.696-706.
- [9] J. Yang and A. Waibel, *Proc. of Third Workshop OII Applications of Computer Vision*, 1996, pp.142-147.
- [10] J. Fritsch, S. Lang, M. Kleinhagenbrock, G. A. Fink and G. Sagerer, *IEEE Int. Workshop OIL Robot and Human Interactive Communication*, September 2002.
- [11] N. Tanibata, N. Shimada and Y. Shirai, *Proc. ofInt. Conf on Virion Interface*, pp.391-398.2002.
- [12] N. Soontranon, S. Aramvith and T. H. Chalidabhongst, *Intematorial Symposium on Communications and Information Technologies 2M14 (ISCLT 2004) Sapporo, Japan, October 26- 29: 2004.*
- [13] M. Seul, L. O'rgoman, M. J. Sammon, *The Press Syndicate of The University of Cambridge*, 2000.