WEBCAM BASED VIRTUAL KEYBOARD

R. JAYACHANDRAN, R. SACHIDHANANDHAM

Abstract - Historically, keyboards are designed for text character printing and text character entry later on in the attached devices. Therefore, keyboard is communication mechanism between human beings and the computing machines and is used to enter textual data and commands. Virtual keyboard are generally assumed to produce the same output as the conventional QWERTY layout produces. The utilization of virtual keyboard appears in space saving situations or requirement in soft programmability of keys or systems avoiding mechanical failure or in movement situations where usability of standard keyboard is limited. Utilization of virtual keyboards in space saving situation is enormous. (e.g.) Text entry in PDA's and cellular phones etc. PDA's and cellular phones don't have a standard keyboard available with them due to limited key on their text entry keypads. Soft key programmability of keyboard is also a considerable factor in the design of virtual keyboards.

Keywords - Posterior Probability, Feature Vector, Highest Posterior Probability, Array of Pre-stored Key, Personal digital assistant (PDA), Computer vision(CV).

I. INTRODUCTION

THE Virtual keyboard have been designed without mechanical keys. These kinds of keyboards are utilized in hostile environments. Virtual keyboards find their position in transport environments. E.g., rail, plane or automotive. Virtual keyboards are also designed for public kiosks and here those designs are suitable which avoids mechanical failure. Industrial environments and medical environments also require the special requirements from keyboard designs.

Problem Statement-To design a vision based Virtual Keyboard which detects interrupt as key recognition instead of mechanical transducer operations of key pressing. Mono-vision video of hand posture for pressing the keys is analyzed. The analyzed hand posture is taken into account under various transactions to estimate the key pressed. Mechanical transducers does two operations for key estimation (key press and key release concepts), while Virtual Keyboard requires only key press operation to estimate the key and not key release operation.

DrawBacks-The cost of the most successful Virtual Keyboard design is high due to design and expensive technology. The set up cost ranges from Rs.5000 and above.

System Requirements

Hardware Requirements-

Webcam	: Pixel ratio of 320 x 240

Software Requirements-

Operating System	: Windows
Media Software	: Java Media Framework

Gesture Architecture - Like any other pattern recognition system, gesture recognition consists of three components.
Gesture acquisition and preprocessing
Gesture feature extraction and representation
Gesture recognition and classification

Gesture acquisition and Preprocessing-Variety of transducers or are available for gesture acquisition ranging from discrete components such as magnetic sensor, position sensor, accelerometer sensor to fully developed functional units such as gesture gloves and more sophisticated systems such as mono vision, stereo vision and range sensors etc. Technologically these sensors employ different methodology ranging from basic physics principles to sophisticated imaging and lasers.
Vision sensors are installed in mainly two configurations they are mono-vision and stereo vision. Mono vision sensors incorporate one sensing camera naming CCD (charge coupled device) or CMOS (complementary metal oxide semiconductor) with multiple possible interfacing such as USB 2.0, Camera link, Ethernet, etc, for their video signal transmission. Similar kind of acquisition sensors are utilized for stereo vision. However, the primary difference exists in the further interpretation of stereo-imaging.

Gesture Future Extraction And Representation-Feature extraction for the purpose of gesture recognition consists of segmentation of image components that contribute to the formation of gesture inputs. Both raster (skin tone blobs, colored gloves, etc.) and vector information (joint geometry, facial animation parameters, etc..) from the basis of feature extraction. Some commonly features for gesture recognition are given below:
Image moments
Skin tone blobs
Colored markers
Geometric features
Multi scale shape characterization
Motion History Images (MHI)
Motion Energy Images(MEI)
Shape Signatures
Polygonal approximation-based Shape Descriptor
Shape descriptor based upon regions and graphs

Gesture Recognition and Classification- Finally, gesture recognition and classification stage classifies the reported features belonging to certain pre-stored category. Following are the list of gesture recognition or classification methods proposed in the literature so far:
Hidden Markov Model (HMM)
Viola and Jones algorithm
Dynamic Programming
Fuzzy Interference Engine
Bayesian Classifier
Template Matching

Hidden Markov Model-Dynamic gestures prolong over certain duration of time, due to which gestures usually appear in the form of sequences or spatiotemporal information. Ultimately some kind of sequences or spatiotemporal matching process is required for successful gesture recognition. Theoretically, a sequence can be characterized as being generated by some parametric random process can be successfully accomplished by Hidden Markov Model (HMM).

Dynamic Programming-Dynamic programming is the method of solving the problems exhibiting the properties of overlapping sub-problems and optimal sub structure that takes much less time than naïve methods. The term was originally used in the 1940's by Richard Bellman to describe the process of solving the problems where one needs to find the best decisions one after another. By 1953, he had refine this to modern meaning. The word "Programming" in "Dynamic Programming" has no particular connection to computer programming at all. A program is, instead the plan for action is produced. For instance, a finalized schedule of events at an exhibition is sometimes called a program. Programming in the sense, finding the acceptable plan of action.

Bayesian Classifier-Bayesian classifier is one of the basic classification methods from supervised classification category. Baye's rule is stated as follows:

$$P(C/x) = (p(C) \times p(C/x)) / p(x)$$

where p(C) is the prior probability, which in particular can be said the prior probability of gesture. P(x\C) is the class likelihood and is the conditional probability that an event belonging to see has observation x. Likelihood for gesture recognition can be specified as the conditional probability that gesture belonging to class C as the feature vector x. p(x) is the evidence in the sense that a particular feature vector for some gesture appears with this probability. Finally posterior probability p(C/x) is calculated by combining the prior. Likelihood and evidence. For multiple classes, the posterior probability can be calculated as

Finally for the minimum error Bayesian classifier selects the class with the highest posterior probability i.e.
Select C if P (C\x) = max

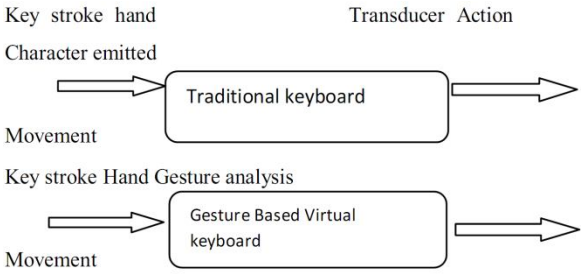
Template Matching- Template Matching is most general class for recognition of a pattern in the pre-stored patterns . Pattern matching operates at both levels raw shapes are recognized or features are initially extracted and later on matched with the pre-stored classes. Pattern matching operates both in spatial as well as in the frequency domain.

Viola Jones Object Detection Framework-The Viola-Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. This algorithm is implemented in OpenCV as cvHaarDetectObjects().

Real Time Mono Vision Gesture Based Virtual Keyboard-Real Time Mono Vision Gesture Based Virtual Keyboard System paper presents a novel mono-vision virtual keyboard design for consumers of mobile and portable computing devices such as PDA's, mobile phones etc. Fuzzy for each symbol (rather than cording methods) inherently approaches to gesture recognition are developed to reveal the realization of soft keyboard. Key pressed over the printed sheet keyboard by analyzing the hand and finger gesture captured in the video sequence. Real time system is developed by integrating camera with PDA in the application environment. Reliable results are experienced by the implementation o f the proposed real time mono vision gestured virtual keyboard system.
In this project a camera for video capture. Novel gesture recognition based virtual keyboard system is designed. A gesture may be defined as the physical movement of hands, arm, face and body with the intent to convey information or command. Gesture recognition consists the tracking of human movement and interpretation of that movement as semantically meaningful commands, Gesture recognition has the potential to be a natural and powerful tool for intuitive interaction between the human and computer. Gesture recognition has been successfully applied in virtual reality, human computer interaction, game control, robot interaction, remote controlling of home and office appliances, sign language, activity recognition, human behavior, and training systems etc. Gesture recognition system is designed in four stages: gesture acquisition, feature extraction, classification, and learning. Gesture acquisition is accomplished by position sensors, motion / rate sensors, and digital imaging. Feature extraction and classification are real time stages to analyze the

acquired gesture while learning stage is off-line activity to learn the relationship between gesture and information or command.a novel gesture recognition based virtual keyboard system which replicates the transducer based keyboard system. Gesture acquisition is accomplished by a mono vision sensor. Suppose the output of the keyboard system is defined as

C= {*c1, c2, ...,cL*} where c1 ='A', C2='B' etc… where as L=63, it's the total no of keys on the keyboard.



Comparison between Traditional Keyboard and Virtual Keyboard.

Proposed work-In Proposed topology, a new perspective to view the problem of Virtual Keyboard is done by using a simple mono vision camera. As the project is fully based on software centric and not hardware centric, therefore project cost is drastically low.

Steps involved in the proposed system:

- Human hand finger movement is analyzed from a video sequence.
- Background objects are eliminated (other than the finger) using Threshold algorithm.
- Finger tip is analyzed and processed using Edge Detection Method.
- Finally, key is evaluated from the process of Edge Detection Method.

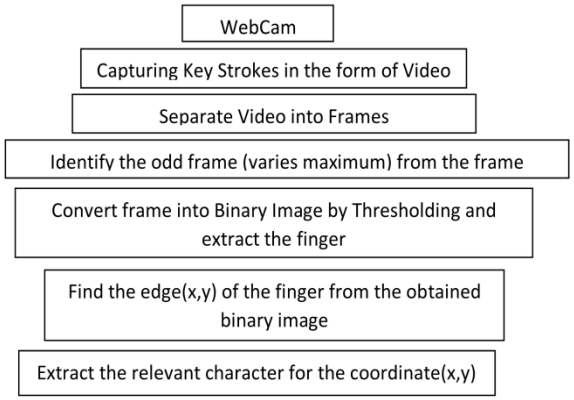
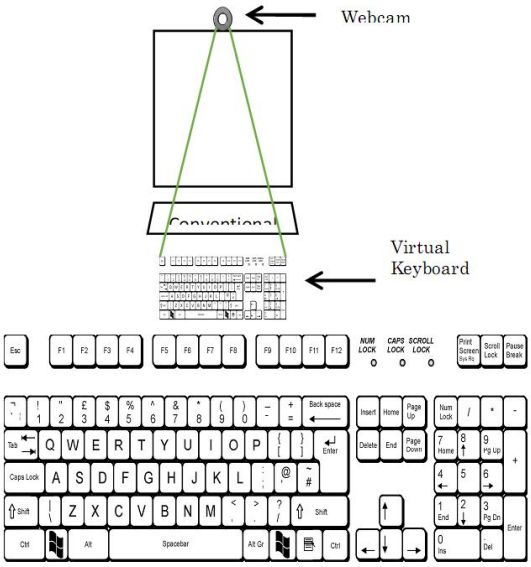


Figure Structure of the Keyboard system

Pictorial Representation-A Camera is mounted over any desired location. But the only criteria are that the camera should focus the entire keyboard layout.

Practical Setup



Modules-**Proposed Virtual Keyboard system includes the following modules.**

1. Image Acquisition
   - Video capture
   - Frame extraction
2. Interrupt Detection from Frame sequence
3. Finger extraction using Threshold Algorithm
4. Finger tip analysis by Edge detection Method
5. Key Extraction

Threshold- Thresholding is the simplest method of image segmentation. From a grayscale image, thresholding can be used to create binary images. During the thresholding process, individual pixels in an image are marked as "object" pixels if their value is greater than some threshold value (assuming an object to be brighter than the background) and as "background" pixels otherwise. This convention is known as *threshold above*. Variants include *threshold below*, which is opposite of threshold above; *threshold inside*, where a pixel is labeled "object" if its value is between two thresholds; and *threshold outside*, which is the opposite of threshold inside . Typically, an object pixel is given a value of "1" while a background pixel is given a value of "0." Finally, a binary image is created by coloring each pixel white or black, depending on a pixel's labels.

Threshold Selection-The key parameter in the thresholding process is the choice of the threshold value (or *values*, as mentioned earlier). Several different methods for choosing a threshold exist; users can manually choose a threshold value, or a thresholding algorithm can compute a value automatically, which is known as *automatic thresholding*.A simple method would be to choose the mean or medianvalue, the rationale being that if the object pixels are brighter than the background,

they should also be brighter than the average. In a noiseless image with uniform background and object values, the mean or median will work well as the threshold, however, this will generally not be the case. A more sophisticated approach might be to create a histogram of the image pixel intensities and use the valley point as the threshold. The histogram approach assumes that there is some average value for the background and object pixels, but that the actual pixel values have some variation around these average values. However, this may be computationally expensive, and image histograms may not have clearly defined valley points, often making the selection of an accurate threshold difficult. One method that is relatively simple, does not require much specific knowledge of the image, and is robust against image noise, is the following iterative method:

1. An initial threshold (T) is chosen; this can be done randomly or according to any other method desired.
2. The image is segmented into object and background pixels as described above, creating two sets:
1. G_1 = {f(m,n):f(m,n)>T} (object pixels)
2. G_2 = {f(m,n):f(m,n)$\leq$T} (background pixels) (note, f(m,n) is the value of the pixel located in the m^{th} column, n^{th} row)
3. The average of each set is computed.
1. m_1 = average value of G_1
2. m_2 = average value of G_2
4. A new threshold is created that is the average of m_1 and m_2
1. T' = $(m_1 + m_2)/2$
5. Go back to step two, now using the new threshold computed in step four, keep repeating until the new threshold matches the one before it (i.e. until convergence has been reached).

Adaptive Thresholding - Thresholding is called **adaptive thresholding** when a different threshold is used for different regions in the image. This may also be known as *local* or *dynamic* thresholding.

Edge Detection Mechanism-Edge detection is a fundamental tool in image processing and computer vision, particularly in the areas of feature detection and feature extraction, which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally which has discontinuities.

The purpose of detecting sharp changes in image brightness is to capture important events and changes in properties of the world. It can be shown that under rather general assumptions for an image formation model, discontinuities in image brightness are likely to correspond to:

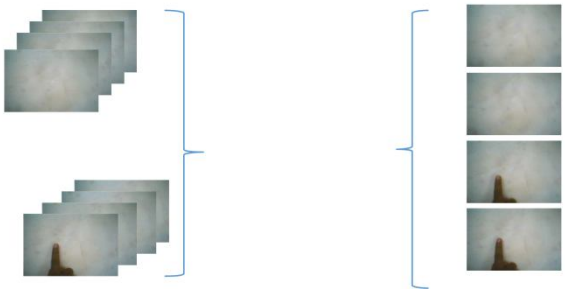
- discontinuities in depth,
- discontinuities in surface orientation,
- changes in material properties and
- Variations in scene illumination.

In the ideal case, the result of applying an edge detector to an image may lead to a set of connected curves that indicate the boundaries of objects, the boundaries of surface markings as well as curves that correspond to discontinuities in surface orientation. Thus, applying an edge detection algorithm to an image may significantly reduce the amount of data to be processed and may therefore filter out information that may be regarded as less relevant, while preserving the important structural properties of an image. If the edge detection step is successful, the subsequent task of interpreting the information contents in the original image may therefore be substantially simplified. However, it is not always possible to obtain such ideal edges from real life images of moderate complexity. Edges extracted from non-trivial images are often hampered by *fragmentation*, meaning that the edge curves are not connected, missing edge segments as well as *false edges* not corresponding to interesting phenomena in the image – thus complicating the subsequent task of interpreting

Image auquisition-Image acquisition is first step of this project where acquisition process is done by using simple mono vision camera. Simple mono vision camera is not a special one just like ordinary camera such as external camera or web cam integrated. Camera should provide atleast 320x240 pixels of image. Camera pixel should be 1.3 or above. More the pixel greater the accurate rate and decreases the error rate.

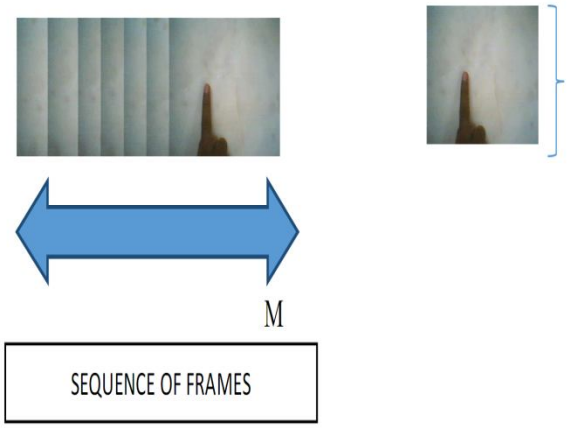
Camera default device driver is required for hardware overriding .When multiple cameras are connected, a window is opened to select required camera. "vfw://0" is the key word used to detect multiple connected cameras.

Acquisition is in the form of the video. Video is sequence of multiple frames. When video splitted into frames, each frames deserved as images 10-15 frames are created for every second. During the process of image acquisition, all the obtained images are stored in any desired location if necessary.



Snapshot of Image Acquisition

Interrupt Detection in frame Sequence-At regular intervals (about 10 to 15 times every second), the current frame from video is compared with previous images and grab the image which has maximum variation.



Snapshot of Interrupt Detection

Finger Extraction Using Threshold Algorithm- Image from the preprocessing sector comprises Red, Green and Blue model i.e., RGB color model. RGB color model is removed from each image and only black and white color model remains. These two colors are called as binary colors and finally a binary image is created.

Thus by converting a colored image into binary image helps in calculating and analyzing the fore ground objects. Ultimately background object details are quietly often reduced to greater level.

Combining all these red >16, green >8, blue>0 produces binary image. The block diagram of the Finger extraction using Threshold Algorithm is shown in the figure

RGB IMAGE &BINARY IMAGE

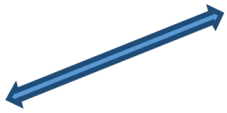
Finger Tip analysis By Edge Detection Method- Edge detection is a fundamental tool in image processing and computer vision, particularly in the areas of feature detection and feature extraction, which aim at identifying points in a digital image at which the image brightness changes sharply or, more formally which has discontinuities. The purpose of detecting sharp changes in image brightness is to capture important events and changes in properties of the world.

The output image from the process of thresholding is in the form of binary image. The obtained image contains more details about the object and lesser details about the background. Only black pixels are taken into account. First black pixel layer is analyzed from the binary image. The vector value of the pixel is calculated. Corresponding value for the vector is pre-stored. Then the key is evaluated.

Pesudocode For Edge Detection Method-

BINARY IMAGE
RED COLOURED LINE DEPICTS THE EDGE OF THE FINGER
Snapshot of Finger tip analysis by Edge Detection Method

Key Extraction-Once when the Virtual Keyboard model is implemented, evaluated key or the model should act as like as traditional keyboard i.e., when key is evaluated, it should made readily available with all the text editors and other applications. This is done by overriding the evaluated key through hardware. Robot package in java does this process easily. Importing Robot package, Robot class is easily called with an object. This object is called by two default functions such as keyPress and keyRelease with a single integer argument. But in case of this Virtual Keyboard only keyPress function in sufficient for Hardware overriding.



Respective x and y coordinate value is calculated
Snapshot Key Extraction

Advantages-Opens a new door to keyboard based applications.

- Games can make the maximum utilization of the keyboard by displaying only those keys that are used in the game.
- Multilingual support: Since the key-displays are reconfigurable, there is no language barrier any more.
- Touch screen is similar to this implementation, but they do require additional effort and are not ergonomically comfortable. User doesn't have to raise his arm to the monitor every time to use it.
- Keyboards with any resolution can be built according to user's choice.
- ***Application-***Any place where keyboard is used.
- As an alternative keypad for mobile phones or smart devices which have a frontal camera. This is possible if the software is converted to J2ME.
- As an input device for gaming.
- At places where a computer or device has multi-lingual users like in net cafes.
- Highly comfortable usage in areas like ATMs, Hospital Bill Checking, Railway Reservation Center, etc.

CONCLUSION

Results showed a very reliable and practical system. The proposed system is less cost due to its software centric mechanism rather a hardware centric mechanism. Performance of the system had been tested over Personal Computer. The data set involved in the development of the system can be easily altered as user requests. Standard data set style is implemented. Response time for Key Extraction is less quite compared to system like Finger-Joint Gesture Wearable Keypad, Thumbcode, etc.

***Future Enhancement*-**Failure rate of the system entirely depends on the light intensity. The system works on dim light also but failure rate is high due to shadowing effects. Future work relies on high efficiency of the system with rid of light intensity

REFERENCES

[1] DSI Datotech Systems Inc.: "*Multi-Point Touchpad*", http://www.dato.com

[2] F. Evans, S. Skiena, and A. Varshney (2006), "*VType: Entering text in virtualWorld*", International journal of human computer studies.

[3] M. Goldstein and D. Chincholle (1999),"*Finger-Joint Gesture Wearable Keypad"*, in Second Workshop on Human Computer Interaction with Mobile Devices.

[4] M. Fukutomo, Y. Tonomura (1997), *"Body coupled fingering: wireless wearable keyboard "*, ACM CHI.

[5] M. Kolsch, M. Turk (2004), *"Robust Hand Detection"*, 6th IEEE international conference on Automatic Face and Gesture Recognition .

[6] M. Kolsch and M. Turk (2002),*"Keyboards without keyboards: a survey of virtual keyboards"*.

[7] M. M. Cerney, J. M. Vance (2005), *"Gesture recognition in virtual environments: a review and framework for future development"*, Iowa state university human computer interaction technical report ISU-HCI 2005-01, Mar. 28.

[8] M. B. Caglar, N. Lobo (2006), *"Open hand detection in a cluttered single image using finger primitives"*, Computer vision and pattern recognition workshop (CVPRW'06).

[9] R. Rosenberg (1998), *"Computing with mice and keyboards: text and graphic input devices for mobile computing"*, PhD Thesis, Dept. of Computer Science, University college London.

[10] SamsungScurry;PCWorldarticleat http://www.pcworld.com/news/article/0,aid,70568,00.asp

[11] Senseboard, http://www.senseboard.com

[12] Virtual Devices, Inc., *VKey;* http://www.virtualdevices.net/press_release.htm

[13] **V. R. Pratt (**2006),**"thumbcode: a device-independent digital sign language, "http:// boole.stanford.edu/thumbcode**. Page 6.**IEEE Transactions on Consumer** Electronics, Vol. 52, No. 4, Nov.

[14] Z. Zang, Y. Wu, Y. Shan, and S. Shafer (2001), *"Visual panel: Virtual mouse, keyboard and 3d controller with an ordinary piece of paper"*,ACM workshop on perceptive user interfaces.

★★★