ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

Human-Computer Interaction based on Real-time Motion Gesture Recognition

Garvit Arya, Manisha Singh, Mayank Gupta

B.Tech Students, Department of Computer Science and Engineering, Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi, India

ABSTRACT: Humans interact with computers in many ways and the interface between themis crucial in facilitating this interaction. Human–Computer Interaction researches the use of computer technology, focusing particularly on the interfaces between humansand computers.In this paper, we propose a robust approach to detect live motion and gesture identification.We demonstrate that use of frame by frame difference with a background subtraction algorithm allows us to have a robust and fast foregroundclassification.

KEYWORDS: Motion Detection, Frame Differencing, Background Subtraction, Gesture Identification, Human-Computer Interaction.

I. INTRODUCTION

A. Overview

With the massive influx of computers in our society, human-computer interaction (HCI) has become an increasingly important part of our daily lives. Processing a video stream to segment foreground objects from the background is the first critical step in many computer vision applications.

One of the artificial vision goals is to emulate certain features of the human visual system, such as the skill of recognizing objects and tracking their movement in a complex environment. The first step in most of the tracking applications is to detect moving object, classifying object pixels and gathering them in connected areas usually known as “blobs”. This reduces the problem complexity giving a global perception of the scene.

B. Detection Techniques

Detection is an inherent component of any efficient tracking algorithm. The detection process helps identifying and localizing moving objects within any environment. The simplest way of accomplishing detection is through building a representation of the background and comparing each new frame with this representation. This procedure is known as background subtraction [13].

Background subtraction (BGS) is a commonly used technique to segment foreground objects. The popularity of BGS largely comes from its computational efficiency, which allows applications such as human-computer interaction, traffic monitoring and video surveillance to meet their real-time goals.

C. Critical Situations Faced By Detection System

Any motion detection system based on BGS needs to handle a number of critical situations such as:

- Gradual variations of the lighting conditions in the scene.
- Small movements of non-static objects such as tree branches and bushes.
- Noise in the image, due to a poor quality image source.
- Movements of objects in the background that leave parts of it different from the background model.
- Sudden changes in the light conditions.
- Multiple objects moving in the scene both for long and short periods.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

- Shadow regions that are projected by foreground objects and are detected as moving objects.

D. Scope

- This project has a vast arena of development and is very useful for a hands-free approach.
- It will be useful in media player and while reading documents or files. A simple gesture could pause or play the movie or increase the volume even while sitting afar from the computer screen.
- Controlling Mouse Movements via Gesture Recognition System can be used on laptops and desktops with the help of a webcam.
- It will be useful for teachers and lecturer for giving presentations in classroom when computer is connected to a projector. A simple gesture could forward the presentation slides even while standing afar from the computer screen.
- One could easily scroll through an eBook or read an email even when eating something.

II. RELATED WORK

The two core issues forming the basis of a gesture recognition system are foreground (moving objects) extraction, and consistent object tracking.

The foreground segmentation aims to extract moving regions of interest, though moving shadows are often detected as well which are unwanted and should be removed.In gesture recognition domain and for static cameras, the popular approach to addressing the issue of foreground extraction is 'Background Modeling' [1]. The background model can either be acquired in advance (Haritaoglu et al., 2000) or estimated adaptively online (Wren et al., 1997). The features/statistical models used have been well-studied in the literature, often depending on the type of imaging sensors or scene properties under study. Some notable examples of feature/model combination are: RGB colour/mixture of Gaussian model (Stauffer and Grimson, 2000) [14]; YUV colour and depth/MoG model (Harville et al., 2001); Chromaticity and gradient/single Gaussian model (McKenna et al., 2000) [15]; normalised rg/kernel density function (Elgammal et al., 2002).

The aim of object tracking is to associate in consecutive frames the objects considered to be the same so as to identify their respective moving paths (Forsyth and Ponce, 2003). A good deal of effort has been expended so far in designing suitable matching strategies and similarity metrics to establish the correspondence, and in many cases multiple object feature descriptors have been used to improve the robustness and performance of a tracker (Javed and Shah, 2002; Lipton et al., 1998; Zhou and Aggarwal, 2001). Objects may appear and disappear in the scene due to entry, exit or various (inter-object or structure) occlusions [9].In our paper, we propose a robust approach to detect live motion and consistent gesture tracking and identification.

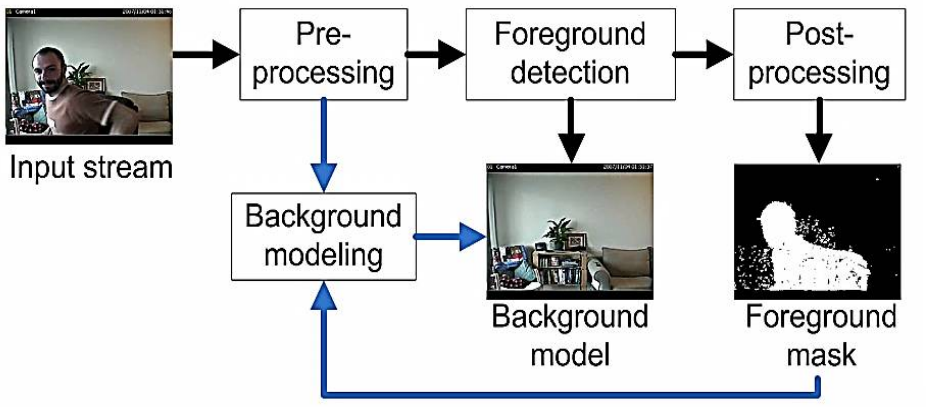

Fig 1. Data flow diagram of a typical BGS algorithm.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

III. BGS ALGORITHMS

There are two main categories of background modeling technique which are:

A. Recursive Techniques

Recursive techniques maintain a single background model that is updated with each new video frame. These techniques are generally computationally efficient and have minimal memory requirements. The disadvantage of this technique is if there is any error at the background model, it needs longer time to disappear.

a) Running Gaussian Average (RGA):

If we consider the background to be nearly static, then the main source of variation in a pixel's color will be due to camera noise. Since camera noise is usually modeled as being Gaussian, it is natural to model each pixel in the background model as a Gaussian distribution [14].

b) Gaussian Mixture Model (GMM):

In order to model multi-modal backgrounds, each channel of a pixel is modeled as a mixture of K Gaussians. It uses a fixed number of Gaussians to model each pixel.

c) GMM With Adaptive Number Of Gaussians (AGMM):

An interesting extension of GMM has recently been proposed by Zivkovic [15], which shows how to automatically adapt the number of Gaussians being used to model a given pixel. This reduces the algorithm's memory requirements, increases its computational efficiency and can improve performance when the background is highly multi-modal [15].

d) Approximated Median filtering (AMF):

All of the above methods model pixels using Gaussian distributions. An alternative uses a recursive filter to estimate the median. The major strengths of this approach are its robustness to noise, and simplicity. A no-table limitation is that it does not model the variance of a pixel.

B. Non-Recursive Techniques

Non-recursive technique is a very adaptive technique that uses sliding window approach to estimate the background and store the previous frame in a buffer. It maintains a buffer of N previous video frames and estimate a background model based solely on the statistical properties of these frames. This causes non-recursive techniques to have higher memory requirements than recursive techniques. However, since they have explicit access to the most recent n video frames they can model aspects of the data that is not possible with recursive techniques.

a) Median Filtering:

Median filtering sets each channel of a pixel in the background model to be the median value as determined from the buffer of video frames. Extending the buffer to include the last background model value makes the algorithm more robust to noise when small buffer sizes are used.

b) Mediod Filtering:

Instead of independently finding the median of each channel, the mediod of a pixel can be estimated from the buffer of video frames as proposed by Cucchiara [13]. This has the advantage of capturing the statistical dependencies between color channels. Mediod filtering is the only background modeling technique that does not treat each color channel independently. A short-coming of this approach is that it does not produce a measure of variance.

c) Eigen Backgrounds (EigBG):

All the other approaches we have considered model each pixel in the background model independently. This approach captures spatial correlations by applying principal component analysis to a set of video frames that do not contain any foreground objects. This result in a set of functions of which only the first D functions are required to capture the

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

primary appearance characteristics of these frames. A new frame can then be projected into the Eigen space defined by these D functions and then back projected into the original image space. Since these functions only model the static part of the scene when no foreground objects are present, the back projected image will not contain any foreground objects. As such, it can be used as a background model. The major limitation of this approach is that computing functions requires a set of video frames without foreground objects. It is not clear how the functions can be updated over time if foreground objects are continually present in the scene.

IV. SYSTEM DESIGN

A. Idea

i. The system continuously streams video from the webcam and processes the frames of the video to recognize the gesture.
ii. This is done by subtracting the RGB value of the pixels of the previous frame from the RGB values of the pixels of the current frame.
iii. Then this image is converted to Octachrome (8 colors only - red, blue, green, cyan, magenta, yellow, white, black). This makes most of the pixels neutral or grey.
iv. The non-grey pixels that remain represent proper motion and noise is eliminated.
v. A database is provided with the system which contains a set of points for each gesture.
vi. As the user performs the gesture, a set of points are generated, using the average of x & y coordinates of non-grey pixels in each frame, which are matched with the gestures in the databases to find the best match.
vii. To match the gestures the points are appropriately scaled as per their standard deviation and then corresponding points of the user's gesture and that from the database are compared.
viii. The gesture which has the least sum of squares of the differences between the corresponding points is returned as the match for the gesture.
ix. According to the gesture recognized, certain set of commands are executed, like executing a keystroke or a particular system command.

B. System Layout

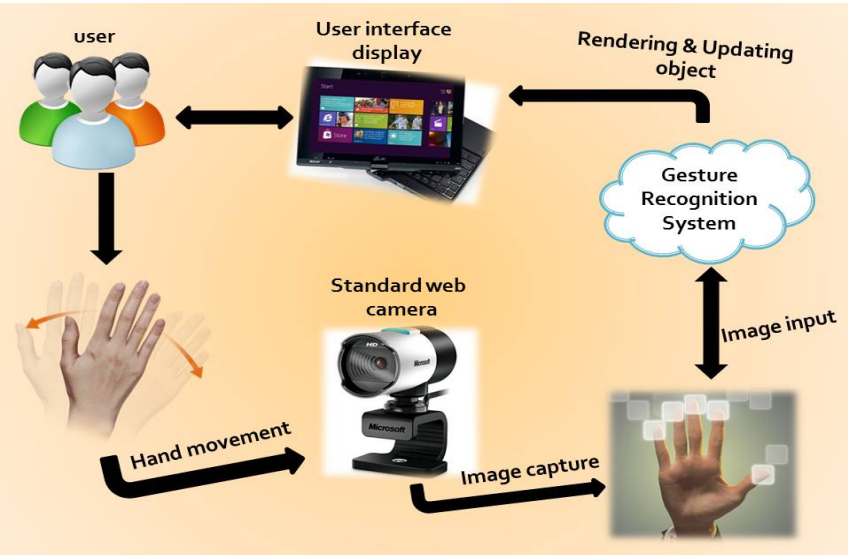

Fig 2. Block diagram of the Gesture Recognition System.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

C. Working

i. We have used the OpenCV library for handling and manipulating input from the webcam.
ii. The algorithm eliminates the static background and hence it can be operated in a place where there is not much movement in the background.
iii. Code is optimized to reduce noise due to change in light.
iv. The movements of the head while performing the gesture are eliminated.
v. We have included further modifications to eliminate noise so that only the moving object (hand or finger) may be interpreted.
vi. The interpreted gesture is scanned against a set of known gesture to find which gesture matches the best.
vii. The action, either a system command or a keystroke, corresponding to the keystroke is then performed accordingly.
viii. Few basic gestures are incorporated for basic operations, while the user is given the choice of adding other gestures.
ix. To simulate keystrokes we are using the X11 library.

D. System Constraints

- It works best when the scene is not overly busy, or is 'non-crowded' and the static and dynamic occlusions are relatively short.
- When the sensor camera vibrates considerably or moves frequently, this solution will fail.

V. EXPERIMENTAL RESULTS

Experiment is conducted using a webcam video with resolution of 640 x 480. Our algorithm is tested using a diverse set of 7 indoor video sequences to track gesture corresponding to letter "M" as shown in Fig 7. These sequences present a significant challenge as they contain moving background elements (Fig 4), objects moving at varying speeds and objects of varying sizes. They also exhibit examples of shadows (Fig 6), slowly varying lighting conditions and absence of threshold lighting condition (Fig 5) that can give rise to the foreground aperture problem.

Fig 3. Actual Image Frame

Fig 4. Moving background

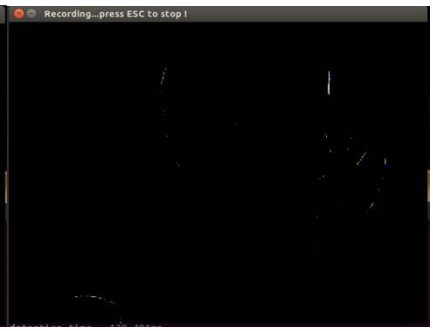

Fig 5. Low light condition

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

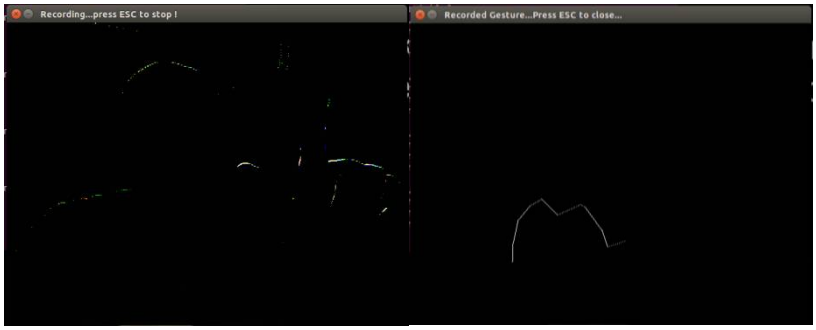

Fig 6. Presence of Shadows

Fig 7. Gesture "M" Recorded correctly

The test results and accuracy corresponding to various constrains and conditions are tabulated below:

TEST CASE ID	CONDITION TESTED	NUMBER OF TIMES TEST IS REPEATED	NUMBER OF CORRECT OUTCOMES	ACCURACY (%)
1	Ideal Environment	20	18	90
2	Moving Background	10	3	30
3	Objects Moving At Varying Speeds	10	7	70
4	Objects Of Varying Sizes	10	6	60
5	Presence Of Shadows	10	7	70
6	Slowly Varying Lighting Conditions	10	8	80
7	Absence Of Threshold Light	10	5	50

Table 1. System test results under various conditions

VI. CONCLUSION

Human-computer interaction is still in its infancy. Visual interpretation of hand gestures would allow the development of potentially natural interfaces to computer controlled environments. In response to this potential, a number of different approaches to video-based hand gesture recognition have grown tremendously in recent years.Thus there is a growing need for systematization andanalysis of many aspects of gestural interaction.

We have presented in this paper a real-time and robust solution, which primarily addresses thepractical challenges in the detection of live motion and gesture identification. The efficiency,robustness and applicability of the system have been demonstrated in a range of visual scenes involving complex object motion and localized illumination changes. The system can be further extended to incorporate mouse movements as well as still gestures.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 3, March 2016

ACKNOWLEDGEMENT

This paper is based on Bachelor of Technology Major Project for Computer Science and Engineering at Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi. The authors are indebted to the guide**Dr. Namita Gupta**, Head of Department - CSEfor her help and support during the course of the major project.

REFERENCES

[1] T. Horprasert, D. Harwood and L.S. Davis, “A statistical approach for real-time robust background subtraction and shadow detection”, IEEE ICCV’99 Frame-Rate Workshop, Corfu, Greece, September 1999.

[2] Haritaoglu, D. Harwood and L. Davis, “Who, when, where, what: A real time system for detecting and tracking people”, Third Face and Gesture Recognition Conference, pages 222–227, 1998.

[3] Y. Kameda and M. Minoh, “A human motion estimation method using 3-successive video frames”, ICVSM, pages 135–140, 1996.

[4] V. Pavlovic, R. Sharma and T. Huang, "Visual interpretation of hand gestures for human-computer interaction: a review", IEEE Trans. Pattern Analysis & Machine Intelligence, vol. 19, no. 7, p. 677-695, 1997.

[5] P. Spagnolo, T. D'Orazio, M.Leo and A.Distante, "Moving Object Segmentation by Background Subtraction and Temporal Analysis", Image and Vision Computing, vol. 24, p. 411-423, May 2006.

[6] H.H. Kenchannavar, Gaurang S. Patkar, U.P. Kulkarni and M.M. Math, "Simulink Model for Frame Difference and Background Subtraction comparison in Visual Sensor Network", The 3rd International Conference on Machine Vision (ICMV 2010), Hong Kong China, 2010.

[7] G. Riva, F. Vatalaro, F. Davide and M. Alaniz, “Ambient Intelligence: The Evolution of Technology, Communication and Cognition towards the Future of HCI”, IOS Press, Fairfax, 2005.

[8] K. Rangachar and R. C. Jain, “Computer Vision: Principles”, IEEE Computer Society Press, 1999.

[9] Donovan H. Parks and Sidney S. Fels, “Evaluation of Background Subtraction Algorithm with Post-Processing”, IEEE Fifth International Conference on Advanced Video & Signal Based Surveillance, pg.192-199, 2008.

[10] A. Murata, “An experimental evaluation of mouse, joystick, joycard, lightpen, trackball and touchscreen for Pointing - Basic Study on Human Interface Design”, Proceedings of the Fourth International Conference on Human-Computer Interaction, pg. 123-127, 1991.

[11] K. Dawson-Howe, “Active surveillance using dynamic background subtraction”, Tech. Rep. TCD-CS-96-06, Trinity College, 1996.

[12] Toshiyuki Yoshida, “Background Differencing Technique for Image Segmentation Based on the Status of Reference Pixels”, IEEE International Conference on Image Processing (ICIP), Vol.5, pg. 3487 – 3490, 2004.

[13] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati. “Detecting moving objects, ghosts and shadows in video streams”, IEEE Trans. on Pattern Analysis and Machine Intel., pg. 1337-1342, 2005.

[14] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland. “Pfinder: real-time tracking of the human body”, IEEE PAMI, pg. 780–785, 1997.

[15] Z. Zivkovic and F. van der Heijden, “Efficient adaptive density estimation per image pixel for the task of background subtraction”, Pattern Recognition Letters, pg. 773–780, 2006.

BIOGRAPHY

Garvit Arya is pursuing his Bachelor of Technology degree in Computer Science and Engineering at Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi, India. He will complete his degree in July 2016. His research interests include Vision Based Computer Interaction, Advanced Human Computer Interfaces, Data Science, Computer Networks, Algorithms, Underwater Sensors, Robotics, Internet of Things, Cyber Security and Forensics.

Manisha Singh is pursuing her Bachelor of Technology degree in Computer Science and Engineering at Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi, India. She will complete his degree in July 2016. Her research interests include Human Computer Interaction and Web Development.

Mayank Gupta is pursuing his Bachelor of Technology degree in Computer Science and Engineering at Maharaja Agrasen Institute of Technology, GGSIPU, New Delhi, India. He will complete his degree in July 2016. His research interests include Human Computer Interaction and Android Development.