

Vision based Motion Estimation for Human Machine Interaction

Prerna D. Uddharwar

Electronics and Telecommunication Department,
Savitribai Phule University of Pune,
Sandip Institute of Technology and Research,
Mahiravani, Nashik

Pravin A. Dhulekar

Electronics and Telecommunication Department,
Savitribai Phule University of Pune,
Sandip Institute of Technology and Research,
Mahiravani, Nashik

ABSTRACT

Detection and estimation of human body motion is a challenging issue for real-time human-machine interaction. Real-time processing and accuracy are key requirements during the designing of the system. The task of detecting and estimating human motion is a very important aspect for various high-level applications. However, many methods suffer because of not having enough robust estimation and proper motion detection. This paper presents a novel human motion detection algorithm that uses a background subtraction based segmentation based on moving blob regions. Considering the accuracy, here in this system a single video camera is employed without any auxiliary marking tools. This approach first obtains a background image through the acquisition and enhancement of video sequences. Then, it obtains a motion image which is then subtracted from the background image to detect the motion. Pre-processing is then applied to the difference image before the major blob is identified. We then calculate the angle of the motion that was detected by the difference image to evaluate the motion effectively. This measured angle is then sent to the hardware control through wireless transmission. Based on the range of the angle, the room lights and fan speed is controlled. Multiple experimental results demonstrate the accuracy of this system.

General Terms

Angle estimation, Switching of room lights

Keywords

Motion detection, background subtraction, motion estimation, real-time human-machine interaction.

1. INTRODUCTION

The study of natural interaction between man and machine is an important and vast research area. An important reason behind this is our widespread dependency on machines in life and the evolution of Human-Machine Interaction in recent years. As most of our work can be very efficiently completed with fewer amounts of efforts and time when we interact with machines, we prefer the interaction to be easily understood and convenient. This especially becomes prominent when the user might not be technologically savvy, or have some disability preventing him from using the available traditional controls and other utilities.

Motion estimation is nothing but the interpretation of human motions via mathematical algorithms [1]. We are trying to make human machine interaction so intelligent that machines would be able to detect the presence of human in its surrounding and will respond to human movements. Human machine intelligent interaction needs vision based motion

estimation which requires many interdisciplinary studies. As human machine interaction is concerned about a human and a machine in conjunction with each other, designing a user interface requires knowledge not only on the human side but also on the machine side. On one side, the knowledge about communication theory, graphic study, social sciences, etc. are required; and on the other hand, techniques in computer graphics, programming languages, operating systems etc. are required [11].

The evolution of user interface has developed from text-based interface that is based on keyboard to GUI based on mouse [22]. In today's world, remote controls, mouse and joysticks are the most popular devices but they are inconvenient and unnatural.

To overcome this difficulty, it is important to design a system that does not need any auxiliary marking tools or any external hardware for controlling and make it efficient to meet the requirement of the natural man machine interaction [7]. We use a single video camera to locate the position of the user, and collect information of his body. With the collected information, we then measure the angle of the detected motion and make the estimation. The task is done through a background subtraction method, which is very convenient for processing in real-time. Furthermore, we apply this estimated motion to control the room lights and fan speed.

2. RELATED WORK

Within the domain of human motion analysis, various surveys have been written. Moeslund et al. [2] grouped application of human motion analysis under these three titles: surveillance, analysis and control. The functional taxonomies used were initialization, tracking, pose estimation and recognition. Human body position estimation and tracking from the input images and video is tackled with auxiliary marking tools [5][6]. These approaches use sensors to mark the location of the hand and the three joints of the arm. Y. Azoz et al. [4] proposed a method for tracking arm dynamics by constraint fusion and multiple cue integration. The cues are color, motion, and shape. J. Yao et al. [3] proposed an approach of the arm gesture detection in a classroom environment using some visual features.

Ronald Poppe [7] introduced the drawback of systems that used attached markers and provided a marker less solution by using cameras as sensors. He also observed that studies can be divided into two main classes: model based and model free approaches.

Pareesh Rawat et al. [8] gave the review of various motion estimation and stabilization techniques. A good comparison of the direct and feature based method are presented. They classified motion estimation technique into two i.e., feature

based approach and direct pixel based approach. Matthew Padiaditis et al. [9] gave the advances in analysis of human motion from video with respect to epilepsy, where human motion is an important element of clinical image. It also described the achievements in vision based motion detection for marker based and marker free systems. R. S. Rakibe et al. proposed method of moving human detection using background subtraction algorithm that will help to find the moving object perfectly. For high accuracy and reliability they used threshold method to detect moving object, background initialization and update the current image in real time [10].

In everyday life, we hope we can control the device without a remote control or without a touch pad. Motivated by these practical requirements, we propose a motion estimation technique to detect the human body and estimation the motion. This estimation can help machines to give smart response to human motion based on the predefined rules.

3. OUR APPROACH

The method is depicted in the flow chart of Fig.1. As can be seen, the algorithm is consisted of four steps: (1) Image acquisition toolbox acquires images and video from any industry standard hardware directly into the MATLAB. The toolbox acquires frames from the video stream and logs the frames to a buffer in memory. (2) The frame conversion method converts the acquired video into frames that are used for carrying out some processes before our main segmentation method. (3) The preprocesses include RGB to grey conversion of the image, image resizing and image filtering. The image pre-processing is used for the enhancement of the image before its main processing. Colours in an image may be converted to gray by calculating the brightness of the hue. Using this value a shade of gray is created that matches the desired brightness. During the resizing of an image, the image to be resized must be specified along with the magnification factor. To enlarge an image, a magnification factor greater than 1 is specified. Where as to reduce an image, a magnification factor between 0 and 1 is specified. The size of the output image can be specified by passing a vector containing the number of rows and columns in the output image. If the produced image with the specified size does not have the same aspect ratio as the input image, the output image will be distorted. If one of the elements in the vector is specified as NaN, imresize calculates the value for that dimension to preserve the aspect ratio of the image. Median filters are quite popular as they are highly capable to provide excellent noise-reduction, with less blurring than the other linear smoothing filters of same size. Median filters are specifically effective in the presence of impulse noise, which can also be called as salt-and-pepper noise because of its appearance as black and white dots on an image. (4) The motion area is detected using background subtraction algorithm. (5) Median filtering is used to remove the noise part. After this the motion region not only includes human motion but it also may include other non human parts. Hence morphological methods are used for taking out non human motion. After this interference of some small parts are eliminated to get a clear and accurate region of human motion. (6) Next we will find the angle of the motion region by using regionprops that measures properties of the motion region. By default it computes the area, centroid and bounding box measurements. By using majorAxisLength, minorAxisLength and Orientation property we get the exact angle of our motion region. (7) We transmit this estimated angle through a zigbee module to the hardware setup. If the

angle matches to our predefined range, the particular light can be turned on and off.

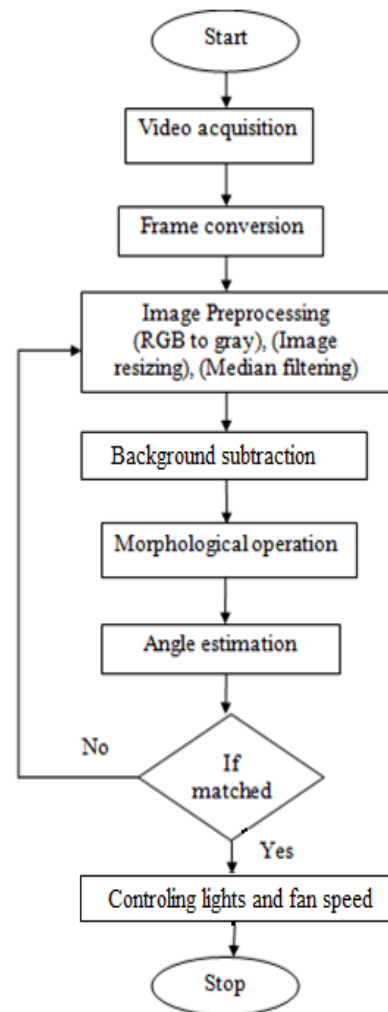


Fig 1: Flow chart for motion detection and estimation method

4. BACKGROUND SUBTRACTION METHOD

The background subtraction is a widely used method for the detection of moving objects in videos from static cameras. The principle used in background subtraction is that the moving objects are detected by the difference of a reference frame and the current frame. This reference frame is called as background image. This background image is nothing but the representation of the scene with no moving objects. The reference image needs to be updated regularly so that it must adjust to the varying conditions and geometrical settings [13].

The key of this method lies in the initialization and updating of the reference image [10].

After the background image $BG(x, y)$ is obtained, subtract the background image $BG(x, y)$ from the foreground image $FG(x, y)$. If the pixel difference is greater than the set threshold T , the result indicates that the pixels are of the moving object. The expression is as follows:

$$DI(x, y) = BG(x, y) - FG(x, y) > T$$

Where, $DI(x, y)$ is the resultant binary difference image. T is the gray-scale threshold.

5. MORPHOLOGICAL OPERATIONS

As the difficulty of the background, the difference image obtained contains the motion region as well as large number of noise [10]. Hence to remove the noise we use median filter with the 3 X 3 window. The median filter replaces a pixel value by median of the gray level in the neighbourhood of that pixel. The original value of the pixel is included in the computation of the median.

A structuring element is used for performing the morphological operations to an input image, creating an output image of the same size. In a morphological operation, each pixel value in the output image is based on a comparison of the current pixel with the corresponding pixel in the input image with its neighbours. By choosing the size and shape of the neighbourhood, you can construct a morphological operation that is sensitive to specific shapes in the input image. The motion region may also include non human motion along with the human body motion. To solve this problem corrosion and expansion operations are used to filter out the non human body motion. Corrosion effectively filters out the non human activity and expansion maintains the shape of human motion.

6. ANGLE ESTIMATION

Angle estimation is the measurement of the properties associated with the motion region. Regionprops (L, Properties) measures properties for each labelled region in the label matrix L. Positive integer elements of L corresponds to the different regions. For example, the set of elements of L equal to x corresponds to region x; the set of elements of L equal to y corresponds to region y; and so on.

Properties can be a list of strings, an array of cells containing strings, the single string 'all', or the string 'basic'. If properties are 'all', all the shape measurements are computed by the regionprops, listed in Shape Measurements. If called with a grayscale image, regionprops also returns the pixel value measurements, listed in Pixel Value Measurements. If properties are not specified or if it is specified as 'basic', regionprops computes only the basic properties such as 'Area', 'Centroid', and 'BoundingBox' measurements. We can calculate the following properties on N-D inputs: 'Area', 'BoundingBox', 'Centroid', 'FilledArea', 'Image', 'PixelIdxList', 'PixelList', and 'SubarrayIdx'.

We first find the bounding box of the motion region in the image. The bounding box is the smallest rectangle containing the region i.e., a 1-by-Q *2 vector

Where, Q is the number of image dimensions: ndims(L), ndims(BW), or numel(CC.ImageSize).

BoundingBox is [ul_corner width], where, the upper-left corner of the bounding box is specified as [ul_corner] and the width of the bounding box along each dimensionwidth is in the form [x_width y_width....].

Next, we measure the centroid. Centroid is a 1-by-Q vector that specifies the centre of mass of the region. The horizontal coordinate (or x-coordinate) of the center of mass is specified by the first element of Centroid, and the second element is the vertical coordinate (or y-coordinate). We also measure the area using the Area prop. It gives the actual number of pixels in the region. The MajorAxisLength specifies the length (in pixels) of the major axis of the ellipse that has the same normalized second central moments as the region. Similarly the MinorAxisLength specifies the length (in pixels) of the minor axis of the ellipse that has the same normalized second

central moments as the region. Both the above properties are supported only for 2-D input label matrices. Lastly we use the orientation prop that gives the angle (in degrees ranging from -90 to 90 degrees) between the x-axis and the major axis of the ellipse that has the same second-moments as the region.

7. CONTROLLING THE ROOM LIGHTS AND FAN SPEED

There are two main parts of the hardware, the transmitter block and the receiver block. The transmitter block contains the PC to which a zigbee module is attached through which the measured angle of the motion region is transmitted to the receiver part.

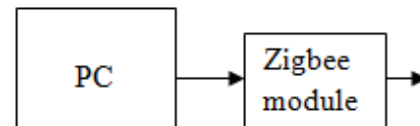


Fig 2: The transmitter block

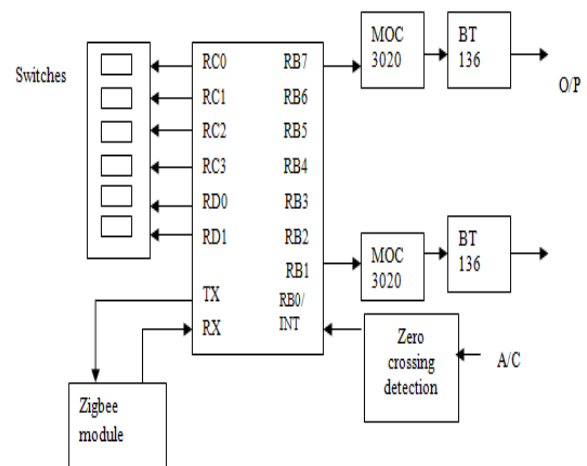


Fig 3: The receiver block

The receiver contains a PIC microcontroller which is a very powerful device that has many useful built in modules e.g. EEPROM, Timers, Analogue comparators, UART. The MOC 3020 is the Opto-isolators or Opto-couplers, that are made up of a light emitting device, and a light sensitive device, all wrapped up in one package, but with no electrical connection between the two, just a beam of light. The BT 136 device is used for general purpose switching circuit. The planar passivity sensitive gate four quadrant triac is intended to be used in general purpose bidirectional switching and phase control applications. It has a high capability to block voltage. It is planar passivated for voltage ruggedness and reliability. It has lower holding current for low current loads. The ZB RF Modules support the unique and most important needs of low-cost and low-power wireless sensor networks and are designed to operate within the ZigBee protocol. The modules provide reliable delivery of data between remote devices and require minimal power. The XBee RF Modules interface through a logic-level asynchronous serial port to a host device. Through its serial port, the module can communicate with any logic and voltage compatible UART; or through a level translator to any serial device (for example: through a RS-232 in proposed system). Data enters UART through the DIN (pin 3) as an asynchronous serial signal. The

signal should be ideally high when no data is being transmitted.

8. RESULTS

The background subtraction algorithm gives the exact estimation of angle of the motion. The angle estimation for experimental motion is shown below.

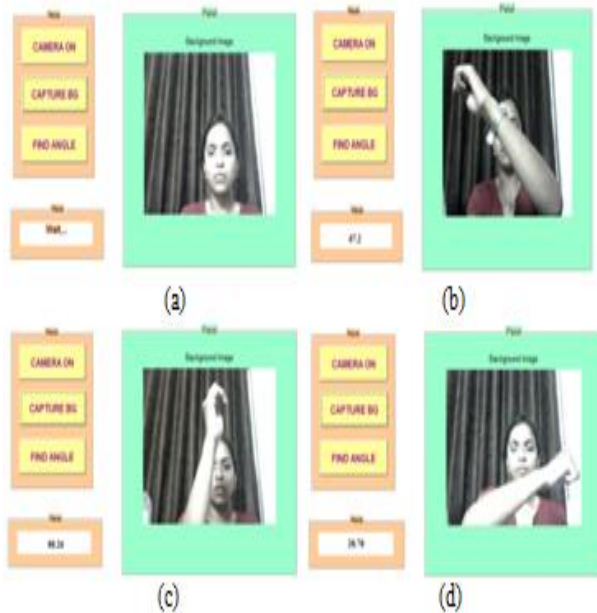


Fig 4: The estimation of angle according to the motion

Further we compared the obtained results with other motion detection algorithm. According to the experiments and survey the result fluctuations are shown in the following table

Table 1. Comparison of result of proposed method with other motion detection method

Sr. no.	Name of Paper	Result accuracy	Technique used
1	A Real-time Upper Limb Motion Estimation System	80 %	Consecutive frame differencing method for detection of motion.
2	The proposed method	97 %	Background subtraction method for motion detection.

9. CONCLUSION

In this paper, we introduce a real-time human motion detection and estimation system, in which we used a background subtraction technique with regionprops that gave the accurate measurement of the properties of the motion region in the difference image. The experiments show that the proposed system can not only accurately recognize the motion type, but also perform processing in real-time to facilitate efficient switching of the room lights and controlling the speed of the dimmer.

The work can be extended in future to minimize the minor fluctuations in the angle estimation that were seen during the experiments. The coupling of hardware and software caused some challenges in interaction which can be handled in future.

10. ACKNOWLEDGMENTS

Sincere thanks to the anonymous reviewers for their comments and suggestions that helped improve the quality of this paper.

11. REFERENCES

- [1] Namrata Verma, Tejeshwari Sahu, Pallavi Sahu, "Efficient Motion Estimation by Fast Three Step Search Algorithms", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 1, Issue 5, November 2012
- [2] Thomas B. Moeslund, Adrian Hilton, Volker Kruger, "A survey of advances in vision-based human motion capture and analysis", Computer Vision and Image Understanding 104 (2006) 90–126
- [3] J. Yao and J. R. Cooperstock, "Arm Gesture Detection in a Classroom Environment," WACV '02 Proceedings of the 6th IEEE Workshop on Applications of Computer Vision.
- [4] Y. Azoz, L. Devi and R. Sharma, "Reliable Tracking of Human Arm Dynamics by Multiple Cue Integration and Constraint Fusion," Computer Vision and Pattern Recognition, Santa Barbara, CA, USA, pp. 905-910, June, 1998.
- [5] T. Schlomer, B. Poppinga, N. Henze and S. Boll, "Gesture Recognition with a Wii Controller," TEI '08 Proceedings of the 2nd international conference on Tangible and embedded interaction ACM New York, NY, USA, 2008.
- [6] V. Mantyla, J. Mantylarvi, T. Seppanen and E. Tuulari, "Hand Gesture Recognition of a Mobile Device User," Multimedia and Expo, 2000. ICME 2000 IEEE International Conference on, 2000.
- [7] Ronald Poppe, "Vision-based human motion analysis: An overview", Computer Vision and Image Understanding 108 (2007) 4–18
- [8] Paresh Rawat, Jyoti Singhai, "Review of Motion Estimation and Video Stabilization techniques For hand held mobile video", Signal & Image Processing : An International Journal (SIPIJ) Vol.2, No.2, June 2011
- [9] Matthew Padiaditis, Manolis Tsiknakis, Norbert Leitgeb, "Vision-based motion detection, analysis and recognition of epileptic seizures—A systematic review", computer methods and programs in biomedicine 108 (2012) 1133–1148
- [10] R.S.Rakibe, Prof.B.D.Patil, "Human Motion Detection using Background Subtraction Algorithm", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 2, February 2014
- [11] Sigal Berman, Member, IEEE, and Helman Stern, Member, IEEE. "Sensors for Gesture Recognition Systems", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART C: APPLICATIONS AND REVIEWS, VOL. 42, NO. 3, MAY 2012

- [12] Thomas B. Moeslund, Adrian Hilton, Volker Kruger, “A survey of advances in vision-based human motion capture and analysis”, *Computer Vision and Image Understanding* 104 (2006) 90–126
- [13] Lijing Zhang ,Yingli Liang “Motion human detection based on background Subtraction” *Second International Workshop on Education Technology and Computer Science,IEEE Computer Science,2010 IEEE.*
- [14] J. Shotton, A. Fitzgibbon M. Cook T. Sharp, M. Finocchio, R. Moore, A. Kipman and A. Blake, “Real-Time Human Pose Recognition in Parts from Single Depth Images,” *Computer Vision and Pattern Recognition, 2011 IEEE Conference*, pp1297-1304, doi:10.1109/CVPR.2011.5995316.
- [15] M. Lin, L. Peng and L. Xun, “A Motion Detection Algorithm Based on Background Subtraction and Three Frame Differencing”
- [16] Z. Hu, M. Chen, R. Chu and H. Lim, “Human Arm Estimation Using Convex Features in Depth Images,” *Proceedings of 2010 IEEE 17th International Conference on Image Processing*, Sep. 26-29, 2010, Hong Kong.
- [17] T. Schlomer, B. Poppinga, N. Henze and S. Boll, “Gesture Recognition with a Wii Controller,” *TEI '08 Proceedings of the 2nd international conference on Tangible and embedded interaction ACM New York, NY, USA, 2008.*