HUMAN BODY SKELETON DETECTION AND TRACKING

Rucha Damle1, Aditya Gurjar[2], Anuja Joshi[3], Kartik Nagre[4]
Pune Institute of Computer Technology, Pune 411043, Maharashtra.
[1]rucha_dam@yahoo.co.in
[2]aditya.smarttalk99@gmail.com
[3]joshi_anuja@rocketmail.com
[4]kartik.nagre@gmail.com

Abstract— **Nowadays, Motion Sensing technology is increasingly deployed in many applications where human detection is required such as gaming, security and military. Motion sensing input device such as Microsoft's Xbox 360 Kinect provides this applicability using infrared sensors. IR sensors are much more expensive compared to optical cameras with comparable resolutions and additional installation overheads, making it inconvenient to be widely used. Also inability to be used in mobile devices further limits its applicability. We create a system using OpenCV to be used as a library which will be used for human body skeleton detection and tracking. The proposed system uses video stream input through an integrated webcam and processes it to obtain human skeleton. The input video stream is processed frame by frame .All the major body joints are identified and tracked by the system using a combination of approaches like haar-training, blob detection, edge detection, optical flow tracking, and background subtraction. Skeleton model is developed using the obtained joints. Obtained skeleton is properly tracked to obtain real time results. Application developer will be able to include this library for use in his application and the system can be thus deployed without any sensors. Thus, the system provides the utility of Kinect-like devices easily with requirement of only normal camera in the device and thus will be easily made available. The system is developed with final aim of taking it on handheld devices like Smartphone, tablets etc.**

Index terms- **Object detection, Object tracking, Optical flow.,**

I. INTRODUCTION

Advancement of technology in field of motion sensing has led to development of products such as Microsoft's Xbox Kinect, PS4 eye etc. Human body modeling plays an important role in computer vision. The problem to recognize the human body and detect its joints is arguably one of the difficult problems in computer vision. This is because there can be n number of poses to detect a   human body and also the appearance and geometry varies from people to people. These devices enable the users to communicate with their console or computer without need of any natural interface. They use Depth sensor which comprises of an infrared projector and a monochrome CMOS (complimentary metal-oxide semiconductor) sensor which work together which sense the 3D features of the body regardless of the lighting conditions.

The use of high quality infrared depth-sensors limits the applicability of Kinect-like devices to be used with hand held devices like mobiles, tablets, etc. Furthermore cost of setting up of these devices is another limiting factor for its widespread use.

We create a system which uses normal camera for human body detection and tracking. The system can be used as library by application developers to be included in their application so that the human body can be detected and tracked by the application like any motion sensing game. The aim of the work is to take the video as an input from a single calibrated camera, process one of the frames to detect and localize the human body joints. The system does not use any depth sensors or stereo sensor cameras. The system eliminates use of any sensors and uses normal camera input and processes it to identify all the major joints and generates a skeleton.

The system provides the utility of Kinect-like devices easily with requirement of only normal camera in the device and thus will be available at a very low comparative cost.

II. RELATED WORK

Human body detection has received considerable amount of interest and several approaches have been proposed for the given problem. One such approach is described by J.C.S. Jacques et al [1]. This approach relies on the initial color model for skin and chest/shirt and anthropometrical measures to identify the chest and upper limbs.

Jitendra Malik et al [2] explain the method to use low level segmentation to drive recognition of joints. Segments and Superpixels are generated by normalized cuts algorithm which are used to propose candidates for limbs and torsos. Finding consistent body configuration becomes a Constraint Satisfaction Problem. Another approach was described by Richard Yi Da Xu, Michael Kemp [3] which describes a method for detecting human body parts from 2D image silhouette using curvature analysis and ellipse fitting. The human upper body is modeled using multiple connected ellipses. The body is extracted by segmented features and ellipses are fitted to each segment. Lastly a non-linear least square method is applied to minimize the differences between connected ellipse model and the edge of the silhouette.

III. OUR APPROACH

Our System consists of two major parts-human body detection and human body joints tracking. Human body detection is done using Viola-Jones algorithm and tracking is done using Lucas-Kanade optical flow algorithm.

A. Haarcascade – Viola Jones

We use the algorithm proposed by Paul Viola and Michael Jones [4] for detection of body joints. It is machine learning approach for visual object detection which is capable of processing images extremely rapidly and achieving high detection rates. The algorithm consists of three parts namely integral image which aids for very rapid feature value calculation. It uses haar like features which are rectangular box internally consisting of white and black rectangles. Value of each feature is calculated which is difference between the sum of all pixels inside white region and black region. It will be very inefficient in terms of time and space to calculate the pixel value of each pixel. Therefore, integral image is used which calculates the pixel value of only the corner pixel to compute the feature value thereby considerably reducing the time and space overhead.
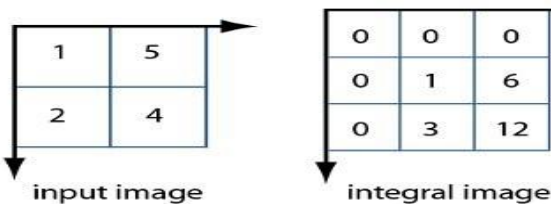

1. Figure 1: Haar features mapped on a face

Figure 2: Input image matrix and the

B. Integral image corresponding

The second part of Viola Jones method is a method
for constructing a classifier by selecting a small number Of important features using Adaboost [5]. Adaboost is used to select a small set of features and train the classifier. In its original form, it is used to boost the performance of simple learning algorithm. In a normal 24x24 window 1, 60,000 features are deployed which are far larger than the number of pixels. Adaboost algorithm does the work of identifying only

the necessary features and discards the unnecessary ones. For this purpose a weak learning algorithm is designed to select the single rectangle feature which best separates the positive and the negative samples. The weak learner defines the threshold value. Further a linear combination of weal classifiers gives a single strong classifier.

The third part is the construction of a cascade classifier. The stages in this cascade classifier are constructed by trainingclassifiers using AdaBoost. Each stage contains a set of features which an image/frame is passed through. Each image/frame having at least 50 % of the matching features is passed to the next stage. The structure of the cascade classifier is essentially that of a degenerate decision tree.

In our work, we use the face detection, upper body detection and palm/fist detection haarcascades for the location of human upper body joints. The detected upper body joints are tracked using Lucas Kanade method [6] which is a differential method for optical flow estimation

C. Our Detection Approach:
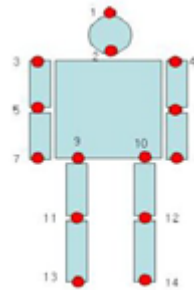
1. Figure 3.Human Body Model using joints

The aim of the work is to detect and track above 14 joints of human body and obtain a skeleton from them.

i) In above figure, 1st point is obtained using haarcascade for frontal face.
ii) 3rd and 4th points for right and left shoulder respectively is obtained using haarcascade for upper body.
iii) 2nd point for the neck joint is located by obtaining centroid of 1st, 3rd and 4th points.
iv) 7th point which represents wrist joint is located using haarcascade for fist/palm. Also, additionally one extra point denoting the index fingertip is located .Using the concept of Golden Ratio, ratio of distance between this extra point to right shoulder joint (3rd point) to the distance between 3rd point and 7th point is equal to 1.618.This ratio is then used to initially locate the elbow joint (5th point).Thus, all the joints on right arm are successfully detected.
Similar approach is used for detecting left arm joints.

D. Optical Flow – Lucas Kanade

Once all the joints are located, Lucas-Kanade optical flow algorithm is applied to all the detected points for tracking them. Optical flow is the pattern of apparent motion of image objects between two consecutive frames caused by the movement of object or camera. It is 2D vector field where each vector is a displacement vector showing the movement of points from first frame to second.

Optical flow works on several assumptions:
The pixel intensities of an object do not change between consecutive frames.

Neighboring pixels have similar motion.

The algorithm does not use color information in an explicit way. It does not scan the second image looking for a match for a given pixel. It works by trying to guess in which direction an object has moved so that local changes in intensity can be explained. The optic flow computation from Lucas and Kanade is based on the image brightness constancy assumption which states that for a motion (u, v) of a point in an image I the brightness of the point does not change:

$$I(x, y, t) = I(x+u, y+v, t+1)$$

Using first order Taylor expansion leads to the gradient constraint equation

$$Ix \ u + Iy \ v + It = 0$$

This underdetermined system is solved by a least squares estimate over an image patch given by

$$(u \ v)T = [\Sigma \ (Ix \ Iy)T \ (Ix \ Iy) \ ]\text{-}1 \ \Sigma \ It(Ix \ Iy)T$$

Lucas-Kanade works well for slow object motions. However with large object motion tracker loses points and tracking in inaccurate. In order to improve the method's ability to track fast moving features which have large displacements, the algorithm is run in a course-fine manner using Gaussian pyramids. This means that, at first, the two frames in

consideration are smoothed and downsampled, effectively reducing their resolution, so that a general estimate for the displacement vectors can be calculated. The images are then upsampled and the process is repeated at different levels of detail until the images have returned to their original resolutions and the most precise estimate for the feature's displacement is derived.

2. Figure 4.Lucas Kanade using Gaussian Pyramids

The Lucas-Kanade algorithm makes a "best guess" of the displacement of a neighbourhood by looking at changes in pixel intensity which can be explained from the known intensity gradients of the image in that neighborhood and is an efficient method for obtaining optical flow information at interesting points in an image (i.e. those exhibiting enough intensity gradient information). It works well for moderate object speeds.

IV. RESULTS AND CONCLUSION:

In this work we have demonstrated how to use haarcascades using Viola Jones algorithm which is used to drive detection of human body joints. It internally uses AdaBoost algorithm which decreases the number of features required and constructs the stages for cascade classifier. The tracking part of the project is done using Optical flow Lucas Kanade function. The problem of recovering human body configurations in a general setting is arguably the most difficult recognition problem in computer vision. We do not claim to have completely solved the problem. But using an approach like Lucas Kanade optical flow to track the detected points which is extremely fast tracking algorithm gives us real time results.

V. V. REFERENCES

[1] J. C. S. Jacques Jr., L. Dihl, C. R. Jung, M. R. Thielo R. Keshet, S. R. Musse, " HUMAN UPPER BODY IDENTIFICATION FROM IMAGES".
Hewlett-Packard Brasil Ltda. using incentives of Brazilian Informatics Law (Law n 8.2.48 of 1991)

[2] Greg Mori, Xiaofeng Ren, Alexei A. Efros and Jitendra Malik, " Recovering Human Body Configurations: Combining Segmentation and Recognition".

[3] Richard Yi Da Xu, Michael Kemp, " MULTIPLE CURVATURE BASED APPROACH TO HUMAN UPPER BODY PARTS DETECTION WITH CONNECTED ELLIPSE MODEL FINE-TUNING"

[4] Paul Viola, Michael Jones , " Rapid Object Detection using a Boosted Cascade of Simple Features"

[5] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt '95, pages 23–37.
Springer-Verlag, 1995.

[6] Jean-Yves Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm".

[3] Richard Yi Da Xu, Michael Kemp, " MULTIPLE CURVATURE BASED APPROACH TO HUMAN UPPER BODY PARTS

DETECTION WITH CONNECTED ELLIPSE MODEL FINE-TUNING"

[4] Paul Viola, Michael Jones , " Rapid Object Detection using a Boosted Cascade of Simple Features"

[5] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. In Computational Learning Theory: Eurocolt '95, pages 23–37. Springer-Verlag, 1995.

[6] Jean-Yves Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm".