Effective Human skeleton extraction from single monocular video

A.MADESWARAN[1], and Dr. K.SAKTHIVEL[2]
[1]PG Scholar, K.S.Rangasamy College of Technology, Tiruchengode, India.
Email: mailtomades@gmail.com, Mobile No: +91 9750880489.
[2]Professor, K S Rangasamy College of Technology, Tiruchengode, India.

Abstract

Human motion has been a topic of great interest for a long time now. The study of human motion in videos sequences and the analysis of this motion in a virtual environment using a various model can be contribute to a great extent to different fields like surveillance, gaming, animation etc. In this paper human motion is tracked from monocular videos. After the silhouette is extracted finding of skeleton extraction of the video by using thinning Algorithm. Here human motion has been tracked from monocular video sequences by using segmentation and thinning.

***Keywords*: Thinning, Normalization, Skeletonization, silhouette.**

I.INTRODUCTION

In this paper, skeleton of the human is extracted from monocular videos; analysed and compared some of the existing techniques which can be used and the setup of the system is marker-less system. The method does not use markers on the characters and is useful in some Medical, Surveillance and Movie-making applications. It is based on normalization, and skeletonization. Here, the work attempts to extract human motion from monocular video sequences by following procedures. Initial step is to capture the motion by using high resolution cameras. Followed by motion capture, input the video file to the system and *system* will convert the video sequences into the frames. After the frame conversion, foreground is subtracted from background using various background subtraction algorithms. After the foreground is subtracted, silhouette of the human is thinned by using thinning algorithm called hilditch algorithm. Overall implementation of this system by using MATLAB v7.11.0 (R2010b). In this paper, tells both the merits and demerits of the system setup and that are explained in each steps.

Types of Methods and Systems

(i). Non-optical systems

a). Inertial systems

Inertial Motion Capture technology is based on miniature inertial sensors, biomechanical models and sensor fusion algorithms. The motion data of the inertial sensors (inertial guidance system) is often transmitted wirelessly to a computer, where the motion is recorded or viewed.

b). Mechanical Motion

Mechanical motion capture systems directly track body joint angles and are often referred to as exo-skeleton motion capture systems, due to the way the sensors are attached to the body. Performers attach the skeletal-like structure to their body and as they move so do the articulated mechanical parts, measuring the performer's relative motion. Mechanical motion capture systems are real-time, relatively low-cost, free-of-occlusion, and wireless systems that have unlimited capture volume

(ii). Optical systems

Optical systems utilize data captured from image sensors to triangulate the 3D position of a subject between one or more cameras calibrated to provide overlapping projections. Data acquisition is traditionally implemented using special markers attached to an actor; however, more recent systems are able to generate accurate data by tracking surface features identified dynamically for each particular subject.

a). Passive markers

Passive optical system use markers coated with a retro-reflective material to reflect light back that is generated near the cameras lens. The camera's threshold can be adjusted so only the bright reflective markers will be sampled, ignoring skin and fabric. An object with markers attached at known positions is used to calibrate the cameras and obtain their positions and the lens distortion of each camera is measured. Providing two calibrated cameras see a marker, a 3 dimensional fix can be obtained.

b). Active marker

Active optical systems triangulate positions by illuminating one LED at a time very quickly or multiple LEDs with software to identify them by

their relative positions, somewhat akin to celestial navigation. Rather than reflecting light back that is generated externally, the markers themselves are powered to emit their own light.

c). Marker-less

Emerging techniques and research in computer vision are leading to the rapid development of the marker-less approach to motion capture. Marker-less systems such as those developed at Stanford, University of Maryland, MIT, and Max Planck Institute; do not require subjects to wear special equipment for tracking. Marker-less Motion capture can be classified into two: Stereo/multiple camera tracking and Monocular camera tracking. Special computer algorithms are designed to allow the system to analyze multiple streams of optical input and identify human forms, breaking them down into constituent parts for tracking. Applications of this technology extend deeply into popular imagination about the future of computing technology. Several commercial solutions for marker-less motion capture have also been introduced, including systems by Organic Motion and Xsens. Microsoft's Kinect system, released for the XBOX 360, is capable of Marker-less motion capture.

II.RELATED WORK

Motion Capture

Motion capture, motion tracking, is the process of recording movement and translating that movement on to a digital model. In filmmaking it refers to recording actions of human actors, and using that information to animate digital character models in 2D or 3D computer animation. The motion capture is the primary task to be done before the video sequence sends to the system. Here some constraints are applied to the system for work properly. There are 1. Video file format should be wmv or AVI with 30 frames per second and also 2. Background should be static with first few frames should be background alone i.e. no human intrusion for first few frames. 3. And Camera should be high resolution.

Frame conversion

In the initial stage, video sequence is converted into frames. The following method is performed under the assumption that the camera is stationary and the background is constant. It is also a requirement that there should only be one human in the foreground. An average of the background images are taken for use in the background subtraction process. The subtraction of background itself was done using various methods such as standard Background subtraction, approximate median method, Running average method and Mixture of Gaussian.

VARIOUS TYPES OF ALGORITHMS:

VIDEO TO FRAME CONVERSION ALGORITHM

Step 0: Acquisition of video sequence from the Video camera to MATLAB environment.

Step 1: Read the video file using 'aviread' function and store it in a variable name.

Step 2: Assign the required frame as 'jpg'.

Step 3: Determine the size of video file and number it.

Step 4: Then, For i=1: fnum,

```
strtemp=strcat(int2str(i),'.',pickind);
imwrite mov(i).cdata(:,:,:),strtemp);
end
```

SIMPLE BACKGROUND SUBTRACTION ALGORITHM

Step0: Read the Video data and convert it into video frames.

Step1: Set the background image using the first frames which does not have the human figure.

Step2: Read the current frame from the video sequence.

Step3: Separate R, G, B components individually for the computation.

Step4: Subtract the R, G, B components of the current frame from the R, G, and B components of background frame.

Step5: Check the threshold values of the difference.

Step6: If value greater than threshold, pixel part of foreground.

Step7: Convert frames to output video

RUNNING AVERAGE ALGORITHM

Step 0: Read the Video data and convert it into video frames.

Step 1: Set the background by taking average of a window of frames before current frame

Step 2: Read the current frame from the video sequence.

Step 3: Separate R, G, B components individually for the computation.

Step 4: Subtract the R, G, B components of the current frame from the R, G, B components of background frame.

Step 5: Check the threshold values of the difference.

Step 6: If value greater than threshold, pixel part of foreground.

Step 7: Convert frames to output video

Approximate median algorithm

Step0: Read the Video data and convert it into Video frames.
Step1: Set the background by taking the first frame
Step2: Read the current frame from the video sequence.
Step3: Separate R, G, B components individually for the computation.
Step4: Subtract the R, G, B components of the current frame from the R, G, B components of background frame.
Step5: Adapt model by adding a value into if difference is positive and subtracting a value when the difference is negative
Step6: Check the threshold values of the difference.
Step7: If value greater than threshold, pixel part of foreground.
Step8: Convert frames to output video.

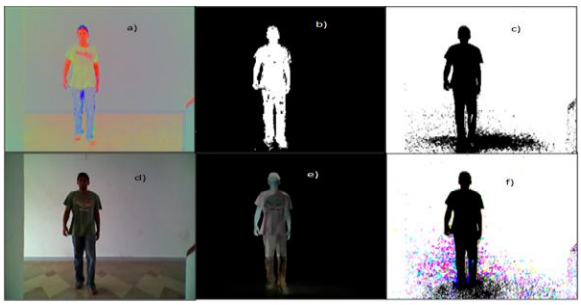

Fig: 1 The result of various background subtraction methods tried. 1) Sample image, 2) Normalized background subtraction, 3) Running average method, 4) Approximate Median method, 5) Normalized sample image, 6) background subtraction after normalising

Normalized Background Subtraction

The standard background subtraction produced a good result provided the background did not change over a period of time. Running average method had the advantage of adapting its background at the risk of the human figure itself being absorbed into it. The mixture of Gaussian also produced good output. But in all these methods the effect of shadows and the presence of other unwanted noises in the data due to lighting effects were present. This problem was solved by performing normalization operation on all the frames before the background subtraction was performed. Normalization process can be defined as extraction of the colour intensities of each pixel, thereby removing the effects of brightness or contrast in the image. This process is done using the following method.

$$Nr = \frac{Ir}{\sqrt{Ir^2 + Ig^2 + Ib^2}},$$

$$Ng = \frac{Ig}{\sqrt{Ir^2 + Ig^2 + Ib^2}},$$

$$Nb = \frac{Ib}{\sqrt{Ir^2 + Ig^2 + Ib^2}}$$

N_r – Normalized red, N_g – Normalized green,

N_b – Normalized blue, I_r – Intensity of red,

I_g – Intensity of green, I_b – Intensity of blue.

Morphological Operations

After the foreground is obtained from the frame after removing the background pixels, morphological operations are performed to process the foreground into a more usable form. Different operations like dilate, erode, clean, fill and bridge are used in this process. The effectiveness of these morphological operations depends on the dimensions of the structuring element selected. Also important is the order in which the morphological operations are done. Various combinations and permutations of these processes give varied result and an optimum solution was obtained after trying various combinations by trial and error. Clean removes isolated pixels (individual 1s that are surrounded by 0s) such as the centre pixel in this pattern. Bridge connects unconnected pixels, that is, sets 0-valued pixels to 1 if they have two nonzero neighbours that are not connected. Fill fills isolated interior pixels (individual 0s that are surrounded by 1s), such as the centre pixel in this pattern. After performing these operations, within the silhouette is obtained.

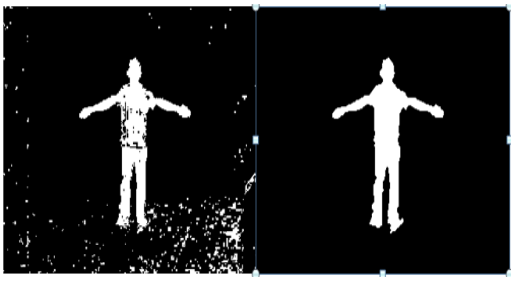

Fig: 2 1) Frame before morphological operations, 2) after morphological operations

Thinning

The next step in the process after morphological operations is process of thinning.

Thinning is done to get the skeleton of the foreground image. This is done as a pre-process to calculate the feature points in the foreground image. But during certain scenarios, due to the problem of occlusion, all the body parts could not be tracked properly in every from. As a solution to this, the initial normalized image was split into its RGB components and each component was thinned separately. This helped in attaining a more effective result. Since thinned images of individual RGB components showed different individual body parts more clearly, the same colour based approach was pursued during segmentation too. Here the different human body was segmented using colours and classified based on their properties. And afterwards thinning was applied to obtain the various feature points. The thinning process itself was performed using 2 different methods: Hit-or-miss and Hilditch. Hilditch though represented a faster means to obtain output, often produced skeletons which were smaller than the silhouette itself. Further variations were made to the thinning algorithm where the very small branches of the thinned image were removed to obtain a better thinned image which represented a better skeleton to the silhouette.

Fig: 3 Result of the different thinning algorithms applied to the a) given silhouette, b) Hilditch Algorithm, c) Hit-or-miss transform, d) the combined thinned image of the RGB components, e) thinning the R-component, f) thinning the G-component, g) thinning the B-component, The combined forms gives a better output as it removes the loop from the skeleton and finds partially-occluded parts.

Two types of thinning algorithm: hit - miss transform

The Morphological thinning is used for skeletonization by structuring elements. There are two structure elements given below. For each iteration, the original image is first thinned by the first structuring element and the result image is thinned with the second structuring element and then thinned with the remaining six 90 degree rotation of the two structuring elements. Repeat this process cycle until none of the thinning produces any future change. Always, the origin of the structuring element is at the centre

0	0	0
	1	
1	1	1

	0	0
1	1	0
	1	

Hilditch algorithm

Step 1: Consider the following 8-neighborhood of a pixel p1.

Step 2: To decide whether to peel off p1 or keep it as part of the resulting skeleton. For we arrange the 8 neighbors of p1 in a clock-wise order and we define the two functions:

(i).B(p1) = number of non-zero neighbors of p1.

(ii).A(p1) = number of 0,1 patterns in the sequence p2,p3,p4,p5,p6,p7,p8,p9,p2

Step 3: For each pixel, check if

(i). B(p1) is greater than or equal to 2 and less than equal to 6.

(ii).A(p1) is equal to 1.

(iii).product of p2, p4, p8 are equal to zero or A(p2) is not equal to 1.

(iv).product of p2,p4,p6 are equal to zero or A(p4) is not equal to 1.

III. KEY OBSERVATIONS AND APPROACH OVERVIEW

The system was tested successfully on a video of length 21 seconds (30fps) of AVI format with a stable background. The human in the image did not perform complex tasks. The accuracy of the skeleton of the image and extraction of foreground from background are almost perfect. During the course of the project, the following the observations were made (i). Normalized background subtraction gives a better foreground. (ii). Thinning of RGB components will give better skeleton when self occlusion occurs. (iii). Shadow of the human cannot affect the skeletionization process. (iv). Thinning using Hilditch is faster but produces a smaller skeleton. (v). Illumination of light cannot the skeletionization process.

Conclusion

The system was tested successfully on a video of length 21 seconds (30fps) of AVI format with a stable background. This system has both merits and demerits. The human in the image did not perform complex tasks. Normalized background subtraction gives a better foreground. Thinning of RGB components will give better skeleton when self occlusion occurs. Thinning using Hilditch is faster but produces a smaller skeleton. Because of Normalization, shadows in the video sequence also not affect the skeleton of the human motion. Dynamic thresholding to make the code work for large input range of videos. The skeleton of the human motion is more effective if there are only one human in the foreground.

REFERENCES

[1] K. Srinivasan, K.Porkumaran, G.Sainarayanan , "Marker-less 3D Human Body Modelling using Thinning algorithm in Monocular Video". *(IJCSIS) International Journal of Computer Science and Information Security, Vol. 8, No.2 May 2010.*

[2] Pedram Azad, Tamim Asfour, RUdiger Dillmann, "Robust Real-time Stereo-based Marker-less Human Motion Capture". 2008 8 IEEE-RAS International Conference on Humanoid Robots.

[3] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review".

[4] J. Saboune, F. Charpillet , "MARKER-LESS HUMAN MOTION CAPTURE FOR GAIT ANALYSIS". INRIA-LORIA, B.P.239, 54506 Vandoeuvre-lès-Nancy, France.

[5] Thomas B. Moeslund, Adrian Hilton, Volker Kru¨ger, "A survey of advances in vision-based human motion capture and analysis". Science Direct, Computer Vision and Image Understanding 104 (2006) 90–126.

[6] S.Veni, K.A.Narayanankutty, M.Kiran Kumar, "Design of Architecture for Skeletonization on Hexagonal Sampled Image Grid". ICGST-GVIP Journal, ISSN 1687-398X,

[7] N.Jin, F. Mokhtarian, "Human motion recognition based on statistical shape analysis," Proceedings of AVSS, pp. 4-9, 2005.

[8] Wei Niu, Jiao Long, Dan Han, and Yuan-Fang Wang, "Human activity detection and recognition for video surveillance," Proceedings of ICME, Vol. 1, pp. 719-722, 2004.

[9] H.Su, F. Huang, "Human gait recognition based on motion analysis," Proceedings of MLC, pp. 4464-4468, 2005.

[10] Tao Zhao, Ram Nevatia and Bo Wu, "Segmentation and Tracking of multiple humans in crowded environments," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 30, No. 7, pp.1198-1211, July 2008.

[11] Mun Wai Lee, and Ramakant Nevatia, "Human Pose Tracking in monocular sequence using multilevel structured models," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 31, No.1, pp.27-38, 2009.