Audio-Visual Fusion and Tracking with Multi camera of Articulated human motion using shape and motion cues

MEGHALATHA C.K[1] and I.KULLAYAMMA[2]

[1]M.TECH,COMMUNICATION SYSTEMS,S.V.UNIVERSITY,TIRUPATI,AP.
[2]ASSISTANT PROFESSOR,DEPT.of ECE,S.V.UNIVERSITY,TIRUPATI,AP.

Abstract

We present a completely automatic algorithm for initializing and tracking the articulated motion of humans using image sequences obtained from multiple cameras. A detailed articulated human body model composed of sixteen rigid segments that allows both translation and rotation at joints is used. Voxel data of the subject obtained from the images is segmented into the different articulated chains using LaplacianEigenmaps. The segmented chains are registered in a subset of the frames using a single-frame registration technique and subsequently used to initialize the pose in the sequence. A temporal registration method is proposed to identify the partially segmented or unregistered articulated chains in the remaining frames in the sequence. The proposed tracker uses motion cues such as pixel displacement as well as 2-D and 3-D shape cues such as silhouettes, motion residue, and skeleton curves.

Keywords:
Dynamicprogramming,Hidden MarkovModels,Humanposeestimation, Motion tracking, silhouettes.

1.INTRODUCTION

Speech is an important modality in human–human and human–computer interactions. Speech signals provide valuable information required to understand human activities and interactions. They are also a natural mode of communication for humans. Thus, speech-based interfaces are of primary importance in intelligent spaces. Similarly, visual cues such as gestures, eye gaze, and affect are also natural modes of expression for humans. The articulated structure of the human body which is composed of a number of segments, each with its associated shape and pose, makes human pose estimation a challenging task. The complexity of the human body and the range of poses it can assume necessitate the use of a detailed model in order to represent its pose and to guide pose estimation. Body models typically incorporate both the shape of individual body parts and structural aspects such as the articulated connectivity and joint locations of the human body monocular techniques suffer from the problems of self-occlusion and "kinematic singularities" and multi cameras are required to estimate pose in a robust and accurate manner.

2.BASIC CONCEPT OF HUMAN TRACKING

This human tracking consists of two parts. The first part describes the partial solutions developed by the members of WP-11 on human activity detection and understanding systems in video in the last eighteen months. In these systems features extracted from video and associated audio are processed by Hidden Markov Models (HMMs), Neural Networks (NN), Dynamic Programming (DP), and eigenspace methods to extract meaningful information. In this report, the following methods are described

- Human face detection in video using DP,
- Human body detection using classification of moving regions in video,
- Moving object classification by eigenanalysis of periodic motion,
- Detection of falling persons in video using both audio and video track data.
- Detection of fire in video, and
- Detection of moving tree leaves in video.

In the second part of the human tracking , is on multimedia databases and the partial solutions to the following problems are described:

- Natural language interfaces for interaction with multimedia database

- Gesture-based interaction with multimedia databases, and

- Text recognition in video.

Finally gaps in know-how and recommended future directions are discussed.

2.1 Human Face Detection

The first step of the human face detection is to find possible face regions in a typical image or video. This can be determined by detecting regions with possible human skin colors. In these regions edges are estimated by using a standard edge detector. Edges can be estimated by summing the absolute values of high-low, low-high and high-high wavelet sub-images of the two-dimensional wavelet transform. Wavelet coefficients of the low-high (high-low) sub-image correspond to horizontal (vertical) edges of the image region.

Horizontal and vertical projections are used as features in region classification which can be carried out using dynamic programming or a standard neural network or support vector machines.

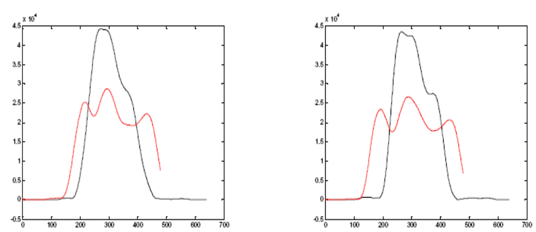

Figure 1: Face edge projections of two face images. Black (red) coloured plots are the horizontal (vertical) projections. In red plots, the first peak corresponds to the left

2.2 Human body detection

Object classification based on contours of moving regions were studied by many authors, see e.g., the report by [Collins1999, Collins2000]. In this approach the moving regions in video is extracted by background subtraction and moving blobs in video are estimated. Once the object contour is determined a one-dimensional function is estimated from the boundary as shown in Figure 2. The distance of the contour from the center of mass of the object idea is computed in a clockwise direction. This distance function is used in object classification.

An example silhoutte and its distance function is shown in Figures 2 respectively.

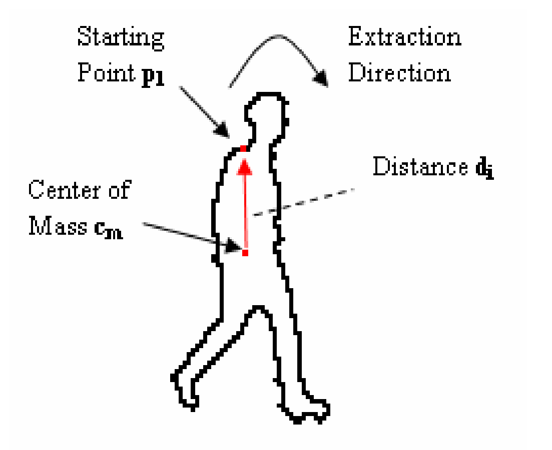

Figure 2: A moving blob in video and the construction of a one-dimensional function to represent the moving blob.

2.3 Moving object classification

In Moving object classification by eigenanalysis of periodic motion, the boundary contour of the moving object is first computed efficiently and accurately. After normalization, the implicit representation of a sequence of silhouette contours given by their corresponding binary images, is used for generating eigenshapes for the given motion. Singular value decomposition produces these eigenshapes that are then used to analyze the sequence. It is also shown that the method can be used not only for object classification but also behavior classification based on the eigen-decomposition of the binary silhouette sequence.

The problem of detection and characterization of periodic activities was addressed by several research groups and the prevailing technique for periodicity detection and measurements is the analysis of the changing 1-D intensity signals along spatio-temporal curves associated with a moving object or the curvature analysis of feature point trajectories, see e.g., [Tsai 1994]. Technion group estimated global characteristics of motion such as moving object contour deformations and the trajectory of the center of mass.

In Figure 3, the motion characteristics of a walking person is shown. It is experimentally observed that a walking cat or a dog exhibit entirely different periodic motion in video and this fact can be used in object classification

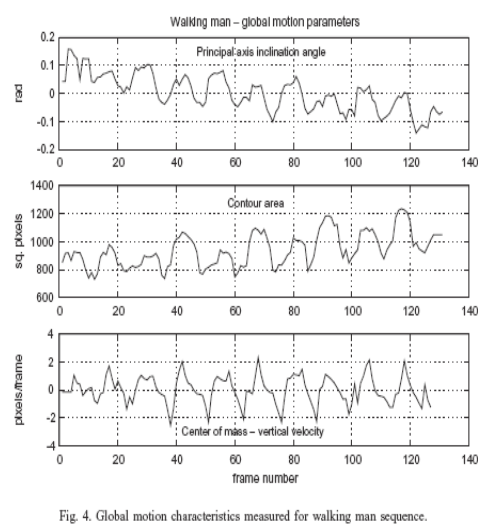

Figure 3: The motion characteristics of a walking person

3.PROPOSED WORK AND RESULTS

3.1 Description of Proposed work

We present an algorithm that fuses together information from these two kinds of cues. Since we use motion and spatial cues in our tracking algorithm, we are able to better deal with cases where the body segments are close to each other, such as when the arms are by the side of the body. Our algorithm consists of a predictor and corrector. We use eight cameras that are placed around the subject. We use parametric shape models connected in an articulated tree to represent the human body as described.

The motion and spatial cues are complementary in nature. We present a framework for unifying different spatial cues into a single energy image. The energy of a pose can be described in terms of this energy image. We can then obtain the pose that possesses the least energy using optimization techniques.

3.2 ALGORITHM

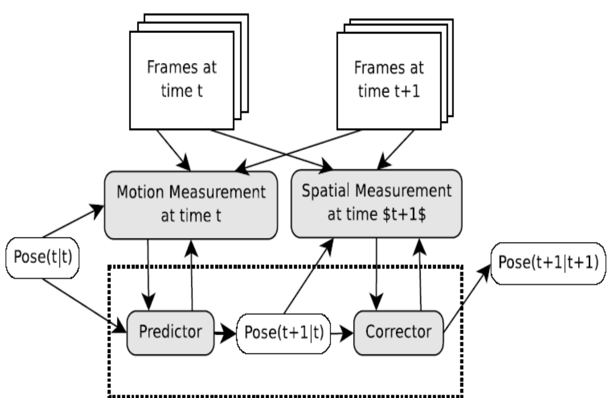

FIGURE4: Overview of the algorithm

FUNCTIONS OF DIFFERENT BLOCKS USED IN PROPOSED WORK:

1)3D shape cues

- spatially registered splines
- temporally registered splines

2)2D motion cues

- pixel displacement

3)2D shape cues

- motion residues
- silhouettes

4)laplacian eigenspace

- pose initialization
- model estimation

5)current work

- pose tracking using prediction and correction.

Steps of algorithm

Step1: 3D pose from pixel displacement

Here it recognize pose at time t and pixel displacement between t and t+1.

Step2: Skeleton curve registration

Here it recognize registered and unregistered skeleton curve at time t0.

Step3: Tracking

Here at time t=t0 2D pixel displacement between frames t and t+1, correct pose using 2D &3D shape cues.

3.3 RESULTS

- When the system is executed using malab RUN command it displays **MOTION TRACKING BLOCK** which is shown below snapshot.

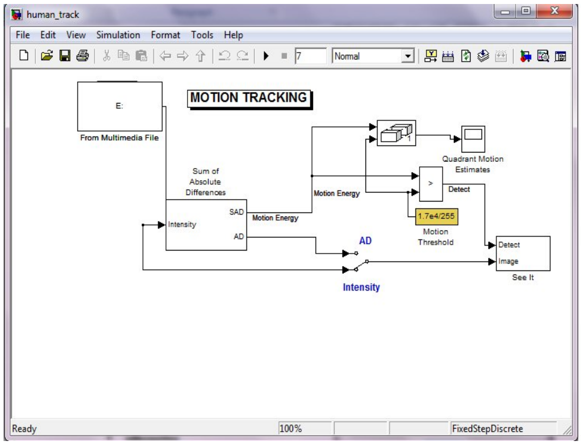

FIGURE 5:Snapshot of human track

- By using **MULTIMEDIA FILE** we will select the input video using browse option which is shown below snapshot.

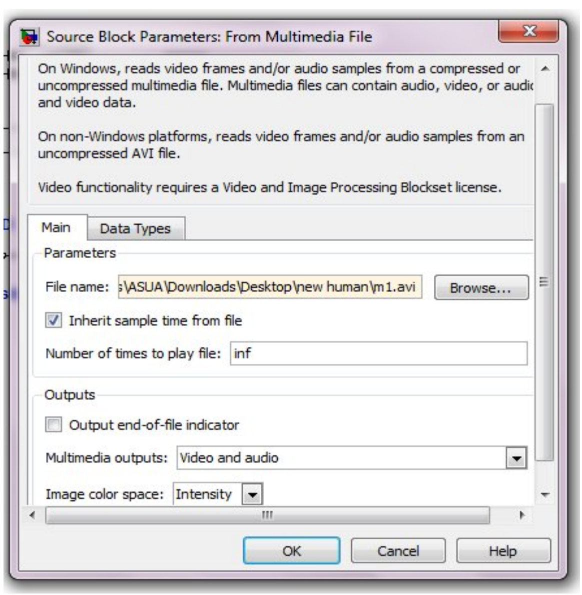

FIGURE 6: Snapshot of multimedia file

- After selecting input video then go to **simulation menu** which is at motion tracking block and then click start then it will show the output video.

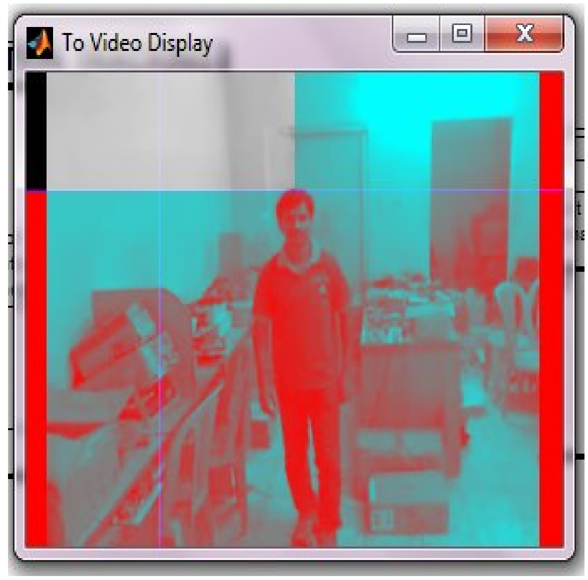

FIGURE 7: Snapshot of output video

4.CONCLUSION

The paper describes a complete pose initialization and tracking algorithm using flexible and full human body model that allows translation at complex joints such as the shoulder joint . An algorithm to perform temporal registration of partially segmented vowels for tracking was also suggested. I used both motion cues and shape cues such as skeleton curves obtained from bottom up voxel segmentation as well as silhouettes and motion residues to perform the tracking .The complete motion capture system has been written in mat lab and I note that currently the computational due to the number of cameras used in the processing and the inefficiency of the mat lab platform.

ACKNOWLEDGEMENT

I would like to expess my gratitude towards my guide, parents and friends whose support and consideration has been invaluable asset during the course of this work.

REFERENCES

[1] *D.Ramana and D.A.Forsyth, "Finding and tracking people from the bottom up" in Proc.IEEE conf.computer vision and pattern Recognition,Madison,WI,Jun.2003,vol. 2,pp.467-474.*

[2]*I.Mikic,M.Trivedi,E.Hunter, and p.cosman "Human body model acquisition and tracking using voxel data,"Int.J.comput.vis.,vol.53,no.3,2003.*

[3] *s.wachter and H-H.nager," tracking person in monocular image sequences,"comput.vis.image understand.,vol.74,no.3,pp.174-192, Jun.1999.*

[4]*Valera,M.,Velastin,S.A,"Intelligent distributed surveillance system a review",IEEE proc.vis.image signal process.,2005 vol.152,pp.192-204.*

[5] *S. T. Shivappa, B. D. Rao, and M. M. Trivedi, "Role of head pose estimation in speech acquisition from distant microphones,"inProc.IEEEInt.Conf.Acoust., Speech,SignalProcess.(ICASSP), 2009, pp.3557–3560.*

[6] H. K. Maganti, D. Gatica-Perez, and I. McCowan, "Speech enhancement and recognition in meetings with and Audio-visualsensorarray,"IEEETrans.Audio,Speech,Lang.Process., vol. 15, no. 8, pp.2257–2269, Nov. 2007.