Human Action Recognition using Motion Based Features

Miss. Shikha. A. Biswas
M.E.Student,Branch-DigitalElectronics, Department of Electronics and Tele-communication,Shri Sant Gajanan Maharaja College of Engineering,Shegaon,S.G.B.Amravati University(Maharashtra State),India.

Prof. Vasant. N. Bhonge
*Associate Professor,
Department of Electronics and Tele-communication, Shri Sant Gajanan Maharaja College of Engineering, Shegaon , S.G.B. Amravati University (Maharashtra State), India.*

Abstract

Analyzing the actions of a person from a video by using computer is termed as Action Recognition. This is an active research topic in the area of computer vision. There are many applications of this research which include video surveillance systems, patient monitoring systems, human performance analysis, content-based image/video retrieval/storage, virtual reality. Although many applications are available of action recognition, the most common application in computer vision is to "look at people".

In this paper, motion-based feature is extracted because motion-based features can portray the direction and velocity of moving body and hence human action can be effectively characterized by motion rather than other features such as color, texture or shape. In the motion-based approach, the method that extracts motion of the human action such as motion blob, optical flow, FIR-filtering or watershed transform are used for recognizing action.

This paper presents a novel method of action recognition that analyzes human movements directly from video. The overall system consists of major three steps: blob extraction, feature extraction and action recognition. In the first step input video is preprocessed to extract the 2D blob. In the second step, motion feature is extracted using optical flow and at last action is recognized using Hidden Morkov Model (HMM).
The use of HMM make our system to deal with time-sequential data and also provide time-scale invariability as well as learning ability for recognition.

1. Introduction

Recognizing or understanding the actions of a person from a video is the objective of action recognition. The main objective of our method is to improve the accuracy of recognition. The concept of understanding in action recognition can be classified in to four types namely: Object-level, Tracking-level, Pose-level and Activity-level. Object-level recognize the location of object, Tracking-level recognize the Object trajectories, Pose-level recognize the pose of a person & Activity-level recognize the activity of person.
The major problems faced in action recognition are: view-point variations: movement of camera, human variation: Humans are of different sizes/shapes, spatial/action variation: Different people perform different actions in different ways with different physics, temporal variation: variations in duration and shift, Occlusions: Action may not be fully visible, Background: Other objects present in the video frame.

In our paper we present a method that eliminates all the above problems by using the concept of optical flow and HMM.

2. Methodology

Human action Recognition can be done in two ways: template matching and state-space model. The template matching technique [12, 13] convert an image sequence into a static shape pattern and then compares it with the prestored prototypes during the recognition process. The advantage of using the template matching technique is that its computational cost is low but it is sensitive to the temporal (duration) variation.

On the other hand, approaches based on state-space model define each static posture as a state. These states are connected by certain probabilities. Any motion sequence is considered as a tour going through these states. Joint probabilities are computed through these tours & the maximum value is selected as the criterion for classifying activities. In such a scenario, duration of motion is no longer an issue because each state repeatedly visits itself. Hence this method of state-space model is robust against temporal variation. And hence for human action recognition, most efforts used state-space model. [3,5-8,11,16]

The three steps of the proposed method are shown below:

Blob Extraction
Feature Extraction
Action Recognition

Figure 1.Stages of Action Recognition

2.1. Blob Extraction

The most commonly used low level feature for identifying human action is 2D blob. Hence the first stage is called blob extraction or segmentation or pre-processing stage.

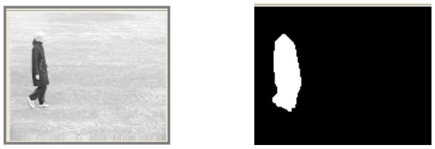
In this stage the color video is first converted from RGB to gray and then finally to binary by thresholding. Thus this stage divide the input video into two classes i.e. foreground (activity of the human) called the 2D blob, and background (empty frame). To remove the salt and pepper noise, the binary video sequence is also median filtered. Median filter remove salt-and-pepper noise without reducing the sharpness of the image.

Then for enhancement dilation is done. In the dilation process the binary video is dilated by a structuring element (called mask or window) of size 3 x 3, 5 x 5 or 7 x 7.The structuring element of the proposed method is shown below:

$$\begin{bmatrix} 0&0&0&1&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&1&0&0&0 \\ 1&1&1&1&1&1&1 \\ 0&0&0&1&0&0&0 \\ 0&0&0&1&0&0&0 \\ 0&0&0&1&0&0&0 \end{bmatrix}$$

After dilation, enhanced 2D blob is obtained. For fast and easy extraction of the ‘motion’ feature enhancement is done. [2, 4, 11]

The blob extraction is shown below:



Input frame dilated frame

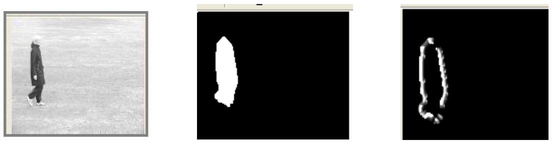
Figure 2. Blob extraction

2.2. Feature Extraction

After segmentation process the next stage is feature extraction. In this stage, mid level feature ‘motion’ is extracted from the blob. Since the human action can be effectively characterized by motion rather than other features such as color, texture or shape, ‘motion’ feature is extracted from the blob.

We use optical flow to estimate motion. Optical flow estimates the direction and speed of moving object from one video frame to another. There are two methods to find the Optical flow Horn-Schunck or Lucas-Kanade method. For floating point input Horn-Schunck method is used and for fixed point input Lucas-Kanade method is used. We use Lucas-Kanade method to find the optical flow. [1,8]

The optical flow of input video frame is shown below:



Input frame dilated frame optical flow

Figure 3.Feature Extraction

2.3. Action Recognition

After the motion is extracted using the optical flow, the last stage is action recognition. For action recognition we use Hidden Morkov Model (HMM). [11,16]

2.3.1. Outline of HMM

HMM is a stochastic state-space transit model which is robust against temporal, spatial & view-point variations. Moreover HMM can deal with time-sequential data and also provide time-scale invariability in recognition. Hence HMM is mostly used as classifier for action recognition. [3, 5-8, 11, 16]

A HMM consists of a number of states each of which is assigned a probability of transition from one state to another state. With time, state transitions occur stochastically. Like Morkov model, state at any time depends only on the state at the preceding time. Symbols are emitted from the states according to the emission probabilities assigned to them. HMM states are not directly observable and can be observed only through a sequence of observed symbols. To describe a discrete HMM following notations are defined:

T = length of the observation sequence.
$Q = \{q_1, q_2, \ldots, q_N\}$ – set of N states of model.
N = number of states in the model.
$V = \{v_1, v_2, \ldots, v_M\}$ – set of M output symbols.
$A = \{a_{ij}\}$, is the N×N transition matrix whose elements $a_{ij} = P\ (q_{t+1} = S_j \mid q_t = S_i)$ are transition probabilities.
$B = \{b_j(O_k)\}$ is the N×M emission matrix of emitting symbol, where $\{b_j(O_k) = P(O_k = v_k \mid q_t = S_j)\}$ is the probability of emitting v_k at time t by state S_j.
$\pi = \{\pi_i \mid \pi_i = P\ (S_1 = q_i\}$, Initial state probability.
$\lambda = \{A, B, \pi\}$ complete parameter set of the model.

Using this model, transitions are described as follows:
$S = \{S_t\}$, t = 1, 2,…, T: State S_t is the t th state (unobservable).
$O = \{O_1, O_2, \ldots, O_T\}$: Observed symbol sequence.

The concept of HMM is shown in Figure 4.There are three states in this example. Each state stochastically outputs a symbol v_k with a probability of $b_j(k)$.If there are M symbols, $b_j(k)$ becomes a matrix of N×M. The HMM outputs the symbol sequence $O = \{O_1, O_2, .., O_T\}$from time 1 to T. We can observe the symbol sequence but cannot the HMM states. The initial state of HMM is also determined stochastically by the initial state probability π. A complete HMM is defined by $\lambda = \{A, B, \pi\}$.

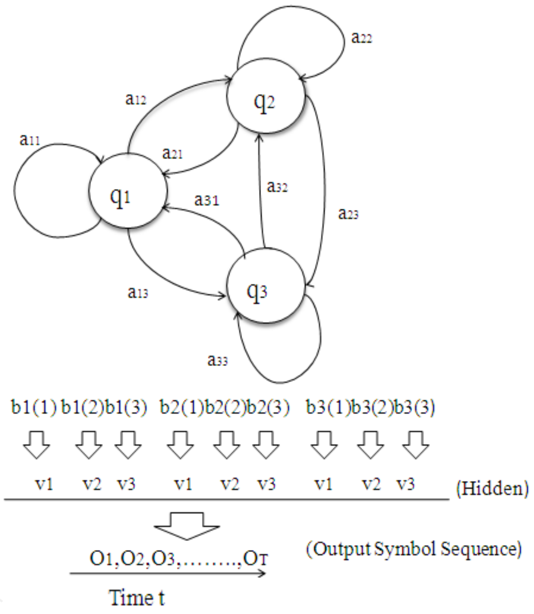


Figure 4.HMM Concept

There are two steps of action recognition using HMM: Learning and training.

2.3.2. Learning of HMM

The input to the feature extraction stage are the time-sequential video frames $I = \{I_1, I_2, \ldots, I_T\}$.The feature extraction stage extract feature vector f_i from each input frame I_i where $f_i \in R^n$, (i = 1, 2,…, T & n is the dimension of the feature space R^n).In the learning phase, HMM is generated which transforms each feature vector f_i into a symbol O_i. Thus a morkov chain of symbol sequence $O = \{O_1, O_2, \ldots, O_T\}$ are generated from the model.

2.3.3. Training of HMM

Once the HMM learning phase is completed it is trained to recognize the human action. In the training phase, the model parameters (A, B, π) are optimized to maximize the probability of observation sequence P(O/λ).The Baum-Welch also called forward-backward or viterbi algorithm is used to find the P(O/ λ).

$$P(O/\lambda) = \arg\max_{i} (P(\lambda_i/O)) \qquad (\text{for } i = 1 \text{ to } T)$$

where P(O/ λ) is called the maximum likelihood estimation of the model parameters. This maximum likelihood is selected as the recognition result.

3. Proposed Method

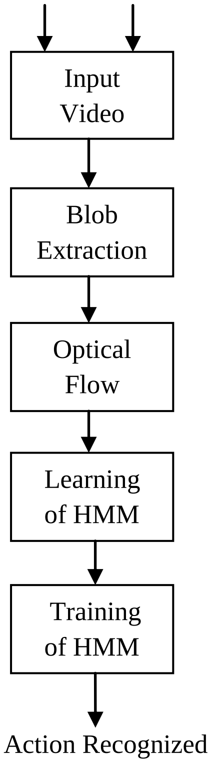


Figure 5.Flowchart of the proposed method

4. Datasets

The datasets used for Action Recognition are KTH and Weizmann.[9,17]

<u>Weizmann</u>

It contains 10 types of actions performed by 9 subjects. Thus it contains total 90 AVI videos, taken with a static camera and static background and with a frame rate of 25 fps. Actions in this dataset include: bend, jack, run, side, skip, wave1, wave2, jump, P-jump walk.

<u>KTH</u>

It contains 6 types of actions performed by 25 subjects in 4 homogenous backgrounds. Thus it contains total 600 AVI videos, taken with a static camera over homogenous backgrounds & with a frame rate of 25 fps. Actions in this dataset include**:** walking, jogging, running, boxing, hand waving, and hand clapping.

5. Results

By using the proposed method the recognition results are shown in Table1and Table 2.

Table 1.Action Recognition using KTH Dataset

Type of sequence	Total No	Correctly Recognized	% of success
walking	10	10	100
running	10	10	100
handwaving	10	10	100
handclapping	10	10	100
Boxing	10	10	100
jogging	10	10	100
	Σ = 60	Σ = 60	Avg =100

Average % of success using KTH dataset is 100%

Table 2. Action Recognition using Weizmann Dataset

Type of sequence	Total No	Correctly Recognized	% of success
walk	9	9	100
run	9	9	100
jack	9	9	100
skip	9	9	100
side	9	9	100
bend	9	9	100
jump	9	9	100
pjump	9	9	100
wave 1	9	9	100
wave 2	9	9	100
	Σ = 90	Σ = 90	Avg =100

Average % of success using Weizmann dataset is 100%

6. Conclusion

This paper has presented a Hidden Morkov Model based approach for action recognition. It has used 2D blob as low level feature and extracts motion feature from the blob using Lucas Kanade method of optical flow. The motion vectors so obtained are transformed into symbol sequence using HMM.

The HMM is then trained to get the maximum likelihood estimation of the model. This maximum likelihood is selected as the recognition result. The average % of success using the proposed method is 100% on Weizmann and KTH datasets.

7. References

[1] S. Hari Kumar, P.Sivaprakash, " New Approach for Action Recognition Using Motion based Features", Proceedings of 2013 IEEE Conference on Information and Communication Technologies (ICT 2013), pp.1247-1252.

[2] Hetal Shah, N. C. Chauhan, "Recognition of Human Actions in Video", International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC) May 2013, ISSN 2321- 8169,Volume 1,Issue 5, pp.489 – 493.

[3] Vasant Kumar, A.D. Saravanan, S.R. Vinotha, "Human Activity Recognition Based on Spatial Transform in Video Surveillance", International Conference on Computational Techniques and Artificial Intelligence (ICCTAI'2011), pp. 19-23.

[4] Muhammad Hameed Siddiqi, Muhammad Fahim, Sungyoung Lee,Young-Koo Lee, " Human Activity Recognition Based on Morphological Dilation followed by Watershed Transformation Method ", 2010 International Conference on Electronics and Information Engineering (ICEIE 2010),Volume 2, 2010 IEEE,V2 433-V2 437.

[5] Ronald Poppe, "A survey on vision-based human action recognition", Science Direct Image and Vision Computing 28 (2010) 976–990.

[6] M.Z. Uddin, J.J. Lee, and T.S. Kim, "Shape-Based Human Activity Recognition Using Independent Component Analysis and Hidden Markov Model," Proc. of 21Ist International Conference on Industrial Engineering and other Applications of Applied Intelligent Systems, 2008, Springer-Verlag Berlin Heidelberg, pp. 245-254.

[7] Mohiuddin Ahmad, Seong-Whan Lee, "Human action recognition using shape and CLG-(Combined local-global) motion flow from multi-view image sequences", Science Direct Pattern Recognition 41 (2008), 2237 – 2252.

[8] Mohiuddin Ahmad and Seong-Whan Lee, "HMM-based Human Action Recognition Using Multiview Image Sequences", Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06), 2006 IEEE.

[9] Moshe Blank, Lena Gorelick, Eli Shechtman, Michal Irani, Ronen Basri, " Actions as space–time shapes", in Proceedings of the International Conference On Computer Vision (ICCV'05), vol. 2, Beijing, China, October 2005, pp. 1395– 1402.

[10] O. Masoud and N. Papanikolopoulos, " A method for human action recognition," IVC, Vol. 21, 2003, pp.729-743.

[11] J. K. Aggarwal and Q. Cai, "Human Motion Analysis: A Review", idealibrary: Computer Vision and Image Understanding,Vol. 73, No. 3, March 1999, pp. 428–440.

[12] J. E. Boyd and J. J. Little, "Global versus structured interpretation of motion: Moving light displays", in Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects, Puerto Rico, 1997, pp. 18–25.

[13] R. Polana and R. Nelson, "Low level recognition of human motion", in Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects, Austin,TX, 1994, pp. 77–82.

[16] Junji Yamato, jun Ohya, Kenichiro Ishii, "Recognizing Human Action in Time-Sequential Images using Hidden Morkov Model",1992 IEEE, pp. 379-385.

[17] http://www.nada.kth.se/cvap/actions/