MOTION BASED ACTION RECOGNITION USING k-NEAREST NEIGHBOR

Shikha.A.Biswas¹, Vasant.N.Bhonge²

¹ M.E. Student, Branch Digital Electronics, Department of Electronics and Tele-communication, Shri Sant Gajanan Maharaja College of Engineering, Shegaon, Sant Gadgje Baba Amravati University, Maharashtra State, India.

² Associate Professor, Branch Digital Electronics, Department of Electronics and Tele-communication, Shri Sant Gajanan Maharaja College of Engineering, Shegaon, Sant Gadgje Baba 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 surveillance systems, patient monitoring systems, human performance analysis, content-based image/video retrieval/storage, virtual reality. Although many efficient applications are available of action recognition, the most active application domain in the area of computer vision is to “look at people”. In this paper, motion feature is extracted because motion features can portray the moving direction of human body and hence human actions can be effectively recognized 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, the 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 classifier k-Nearest Neighbor (kNN).

Keywords: Action Recognition, 2Dblob, Optical Flow, kNN

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. Action recognition is classified into four types: Object-level, Tracking-level, Pose-level & Activity-level. Object-level recognize the locations 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 variation: movement of camera, temporal variation: Variation in duration and shift, spatial variation: Different people perform the same action in different ways. In our paper we present a method that eliminates all the above problems by using the concept of optical flow and kNN.

2. TOOL
The tool used for recognition is MATLAB, version 7.10.0(R2010a)

3. DATASETS
The datasets used for Action Recognition are KTH and Weizmann. [9], [12]

Fig-1: Stages of Action Recognition
4.1 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. To remove the salt and pepper noise, the gray scale video is first median filtered and then is converted into binary using autothresholding.

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). 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:

\[
\begin{bmatrix}
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 \\
1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0
\end{bmatrix}
\]

After dilation enhanced 2D blob is obtained. For fast and extraction of the ‘motion’ feature, enhancement is done. [2], [5], [11]

The blob extraction is shown below:

![Input frame](image1) ![dilated frame](image2)

**Fig-2: Blob extraction**

4.2 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], [7], [8]

The optical flow of input video frame is shown below:

4.3 Action Recognition

This is the last stage of action recognition. For recognition, we use k Nearest Neighbor (kNN). [1], [3], [4], [6]

kNN is a standard classifier which is mostly used for action recognition because it does not require any learning process and also it is invariant against view-point, spatial and temporal variations.

Before classification using kNN, the proposed method computes the following steps:

i) First the centroid of the connected region in the optical flow is computed. This is called Blob analysis.

ii) The dimension of Blob analysis is very large; hence to reduce its dimension Principal Component Analysis (PCA) is done. PCA is a technique which reduces the dimension of large data sets.

iii) Then covariance matrix of PCA is found.

iv) Then eigenvalues (EVA) of the covariance matrix are found. EVA measure the magnitude of the corresponding relative motion. This is called the training data for classification.

v) Then the nearest neighbor in the training data is searched using the distance metric ‘Euclidean distance’ of kNN.

vi) After k-Nearest Neighbor (kNN) search, they are classified using kNN classifier. If the 1-NN is obtained for each action in the dataset then it means that the action is recognized.

vii) The result is also plotted using function ‘gscatter’ to observe the classification.
5. FLOWCHART OF THE PROPOSED METHOD

6. RESULTS

Fig – 4: Recognition Result
7 DISCUSSION

By using the proposed method the accuracy of recognition are shown below:

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Recognition using KTH Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of sequence</td>
<td>Total seq</td>
</tr>
<tr>
<td>walking</td>
<td>10</td>
</tr>
<tr>
<td>running</td>
<td>10</td>
</tr>
<tr>
<td>hand waving</td>
<td>10</td>
</tr>
<tr>
<td>handclapping</td>
<td>10</td>
</tr>
<tr>
<td>Boxing</td>
<td>10</td>
</tr>
<tr>
<td>jogging</td>
<td>10</td>
</tr>
<tr>
<td>Σ = 60</td>
<td>Σ = 60</td>
</tr>
</tbody>
</table>

Average % of accuracy using KTH dataset is 100%.

<table>
<thead>
<tr>
<th>Table 2</th>
<th>Recognition using Weizmann Dataset</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type of sequence</td>
<td>Total seq</td>
</tr>
<tr>
<td>walk</td>
<td>9</td>
</tr>
<tr>
<td>run</td>
<td>9</td>
</tr>
<tr>
<td>jack</td>
<td>9</td>
</tr>
<tr>
<td>skip</td>
<td>9</td>
</tr>
<tr>
<td>side</td>
<td>9</td>
</tr>
<tr>
<td>bend</td>
<td>9</td>
</tr>
<tr>
<td>jump</td>
<td>9</td>
</tr>
<tr>
<td>pjump</td>
<td>9</td>
</tr>
<tr>
<td>wave 1</td>
<td>9</td>
</tr>
<tr>
<td>wave 2</td>
<td>9</td>
</tr>
<tr>
<td>Σ = 90</td>
<td>Σ = 90</td>
</tr>
</tbody>
</table>

Average % of accuracy using Weizmann dataset is 100%.

8. CONCLUSION

This paper has presented a motion-based approach for action recognition. It has used 2D blob as low-level feature and extracts mid-level feature ‘motion’ from the blob using the method Lucas Kanade of optical flow. The motion features so obtained are classified using kNN classifier. The advantage of using kNN is that it does not require any learning process and also it is invariant against view-point, temporal and spatial variations; hence its accuracy is good. The average % of accuracy using the proposed method is 100% on Weizmann and KTH datasets.

REFERENCES