iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 5 Issue: VI Month of publication: June 2017

DOI:

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

Human Pose Estimation in Videos Using K-NN and LBP

Rashmi. M. S[1], Dr. M. B. Meenavathi[2], Dr. H. N. Suresh[3]
[1] *M. Tech Student, Department of Electronics and Instrumentation, Bangalore Institute of Technology, VTU*
[2] *Professor, Department of Electronics and Instrumentation, Bangalore Institute of Technology, VTU*
[3] *Professor, Department of Electronics and Instrumentation, Bangalore Institute of Technology, VTU*

Abstract: The main aim of this paper is to estimate the pose of a human in a video. The input videos are preprocessed and converted into frames. At the recognition stage, segmentation of the scene is performed and the objects and their poses are recognized by searching the k-nearest neighbors in the database. Spatio-temporal local binary patterns are used as features. The distance of the nearest neighbors in the k-nn classifier is calculated by the Euclidean distance. During tracking process, human regions are detected in each frame by foreground detection and are used to track the person in the video.
Keywords – Spatio-temporal, local binary patterns, k-nn classifier, Euclidean distance

I. INTRODUCTION

The ultimate aim of image processing is to extract important features from the image or video data. Recognizing human pose in videos has been an important area of research in the computer vision community for long. It is required in applications like automatic video retrieval and indexing, video surveillance, suspicious activity detection, sports video analysis, personal gaming etc. Automatic analysis of human movements and recognizing the actions is one of the interesting domain but it is challenging in image processing and computer vision. In present, researches on recognition of actions performed by humans have gained.

The existing methods for pose estimation from still images adopt probabilistic and compositional graphical models where nodes represent part appearance and edges represent geometrical deformation. Errors mainly arise from small parts, like forearms, wrists due to large variation. Video pose estimation methods capture the information by adding many pairwise terms among parts at subsequent frames to the graphical model. However, these models are loopy and require approximate inference.

The proposed method integrates the testing and training of pose estimation of video and action recognition. In the training process, information from both tasks is used to minimize the model variables or parameters and in the testing process, the labels and part positions are combined jointly. We start building a histogram of Spatio-temporal LBP model to show the poses and actions together. An ordered form of this model can portray overall geometric design in a single frame and the temporal pose relation in succeeding frames. The action information which is of low resolution is captured and this action is decomposed into poses at each frame. Each of the poses is broken up into ST parts that covers most of the portion of the human body and LBP feature extraction rugged to picture contrast. The LBP with spatio-temporal extracts textural features element or component can be viewed as a pose, it will have a small difference and distortion and each of the components is shown features in the mid-level and fine level part features from a single pose estimation of an image. In order to capture the specific motion information of each action, histogram of spatio temporal LBP parts at adjacent frames are connected to represent temporal co-occurance and deformation. The parameters of the model at different levels are separately trained.

In the present work, the main goal is to design a human pose estimation tool using spatio-temporal and LBP histogram features matching in a video. The pose descriptor is based on the idea that a human pose descriptor can be represented as a distribution of the local texture patterns extracted from a spatio-temporal template. This project proposes a method of designing a spatio-temporal template.

II. DATA ACQUISITION

The video samples of different poses used in this research work are collected. We have collected videos of different human poses, from which the database is created. The pose videos are represented in RGB format.

Volume 5 Issue V, May 2017
ISSN: 2321-9653
International Journal for Research in Applied Science & Engineering Technology (IJRASET)

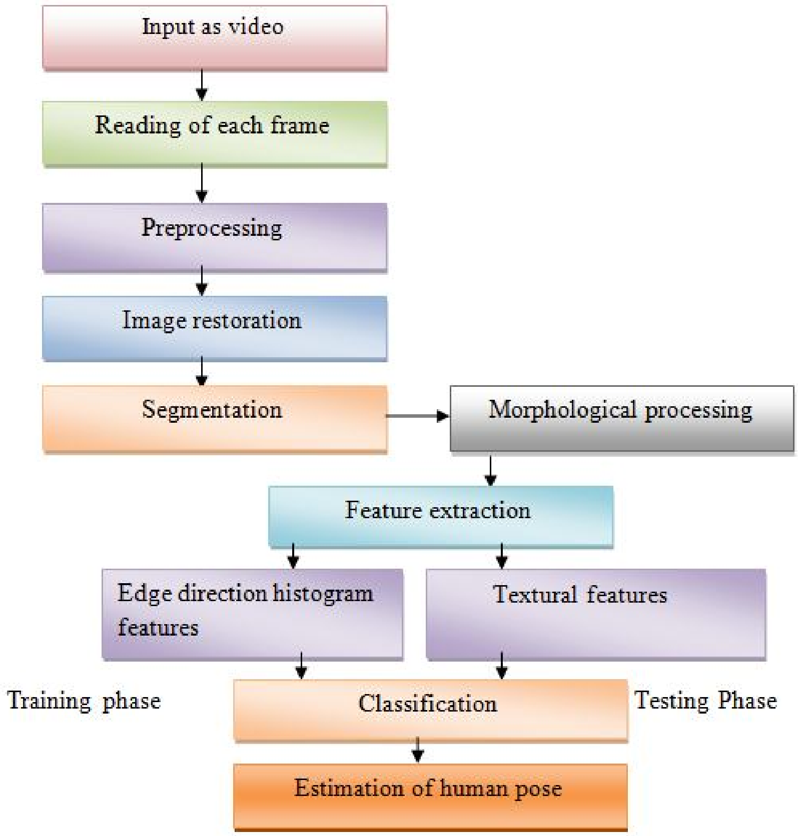


Fig 1 : Estimation of human pose

The development of automated tool for the estimation of human pose of 2D videos follows the flow of system architecture as shown in above Fig 1. This shows the different stages of an image processing techniques being carried out to design a tool.

III. IMPLEMENTATION

This section describes the various steps in the proposed methodology. A typical estimation of matching different human poses in video comprises of pre-processing, segmentation and morphological processing and feature extraction phases. The input video is selected from human pose video database. These videos are trained and tested for pose estimation. Pre-processing incorporates the conversion of the input image into gray scale image. Then the frames are subjected to background subtraction and is given by the equation.

$$B(x,y,t) = I(x,y,t-1) \quad (1)$$

$$\Downarrow$$

$$| I(x,y,t) - I(x,y,t-1) | > th \quad (2)$$

Where th = threshold.

Frame difference

$$| frame_i - frame_{i-1} | > th \quad (3)$$

Volume 5 Issue V, May 2017
ISSN: 2321-9653

Median filtering is used to remove noise from foreground frames images, and is given by

$$\text{Median } [A(x) + B(x)] \neq \text{Median } [A(x)] + \text{Median } [B(x)]$$

The frame processed is subjected to segmentation. In the segmentation process we implement the morphological operations. We apply a structuring element to an input image, creating an output image of the same size. In the tracking process, bounding box is applied to track the person in the video. One of the prerequisites of identification and recognition of human poses is feature extraction. The features will be extracted image processing techniques and an analysis based on visual perception will be carried out with spatio temporal histogram LBP templates. Here we implement Local Binary Patterns (LBP) for feature extraction. Mathematically LBP operator can be written as:

$$\text{LBP}(g_c) = \sum_{i=0}^{p-1} B(gi - gc) * 2i \tag{4}$$

$$B(x) = \begin{cases} 1\,; & \text{if } x \geq \text{threshold} \\ 0\,; & \text{otherwise} \end{cases} \tag{5}$$

For the classification of poses k-nearest neighbor classifier to train and test the features. The input consists of the k-closest training examples in the feature vectors and the output depends on whether k-NN is used for classification.

IV. PERFORMANCE EVALUATION

A. Sensitivity

Sensitivity [12] or recall is defined as

$$\text{Recall} = \frac{\text{Number of relevant Images retrieved}}{\text{Total number of relevant images}}$$

B. Specificity

Specificity [12] or precision is defined as

$$\text{Precision} = \frac{\text{Number of relevant Images retrieved}}{\text{Total number of relevant images}}$$

C. F score

The F measure [16] is the harmonic mean of Precision and Recall. It is defined as

$$\text{F measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

V. EXPERIMENTAL RESULT AND DISCUSSION

This section provides the detailed information of the experimental observations. The images used as training samples and testing samples, results obtained using from algorithms, used for feature extraction and the outcomes obtained after the feature extraction and identification, are as given below:

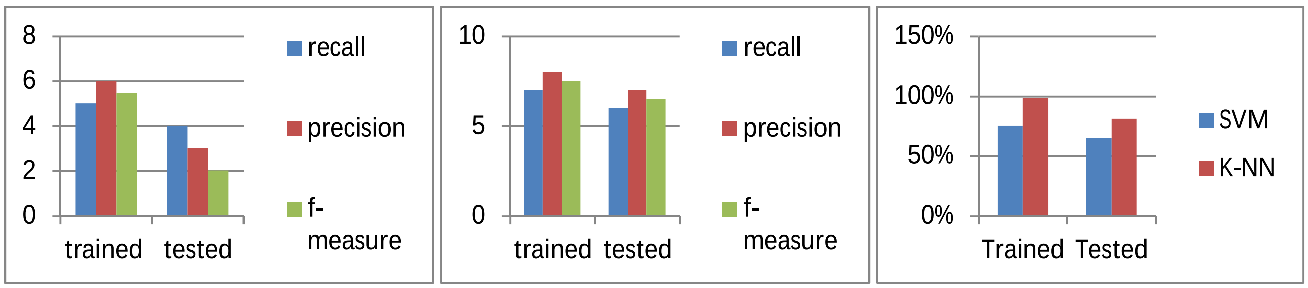
The performance values of F-measure, precision and recall, precision are determined and equated for the both the features with spatio temporal features using SVM classifier [1] and Histogram spatio temporal with LBP features, the classification using K-NN classifier. The dataset is divided in to testing and training videos of which we have trained boxing, jogging, hand waving, hand clapping, running, cycling, and surfing using K-NN classifier of different human poses. The average values of recall, precision, and F-measure of the classifiers which are nearly new to estimate the human pose are calculated.

Human Poses	Recall	Precision	F-measure	Recall	Precision	F-measure
	Trained	Trained	Trained	Tested	Tested	Tested
Poses identified of 8 poses	5	6	5.45	4	3	2

Table I : The performance evaluation of trained and tested videos using SVM classifier.

Human Poses	Recall	Precision	F-measure	Recall	Precision	F-measure
	Trained	Trained	Trained	Tested	Tested	Tested
Poses identified of 8 poses	7	8	7.5	6	7	6.5

Table II**:** The performance evaluation of trained and tested videos using k-nn classifier.



(a) (b) (c)

Fig 2: (a) Chart of trained and tested videos using SVM classifier. (b) Chart of trained and tested videos using K-NN classifier. (c) Average percentage calculated from the performance evaluation of recall, precision, f- Table I and Table II.

For trained video samples, the percentage of human pose estimation using spatio tempo features with SVM classifier is 75%, for the tested video the percentage is 65%.

For trained video samples, the percentage of human pose estimation using histogram spatio tempo LBP features with K-NN classifier is 98%, for the tested video the percentage is 85%

VI. CONCLUSION

In this work, we proposed a human pose estimation and recognition method with edge direction histogram and histogram of spatio tempo LBP texture feature extraction algorithm. K-nearest neighbor classifies the human poses with spatio tempo LBP textural features more accurately in videos than SVM classifier. The recognition rate estimation of human poses by K-NN classifier is 95%. It has been shown that without constructing any complex model, the proposed simple and compact descriptor performs well on different actions and the recognition rate is promising enough for practical use compared to the state of the art methods.

VII.ACKNOWLEDGEMENT

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without the mention of the people who made it possible, whose constant guidance and encouragement crown all the efforts with success. I am deeply indebted to all people who have guided, inspired and helped me in the successful completion of this project. I would like to thank the faculty and staff members for their support. I wish to thank all who directly or indirectly extended their help and cooperation during the completion of the project.

REFERENCES

[1] HanguenKim, SangwonLee, DongsungLee, SoonminChoi, JinsunJu and HyumMyung, Real time human pose estimation and gesture recognition from depth

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

images using superpixels and SVM classifier, Sensors 2015.

[2] Zhang, Z.Microsoft Kinect sensor and its effect. IEEEMultimed.2012, 19, 4-10.

[3] Arieli.Y, Freedman.B, Machline.M; Shpunt.A, Depth Mapping Using Projected Patterns. U.S. Patent 8, 150, 142, 3 April 2012

[4] Zhang.Z, SoonSeah.H, KwangQuah.C, Sun.J.GPU-acceleratedreal-timetracking of full body motion with mlti-layer search. IEEE Trans. Multimed. 2013, 15, 106-119.

[5] Shotton.J, Girshick.R, Fitzgibbon.A, Sharp.T, Cook.M, Finocchio.M, Moore.R, Kohli.P, Criminisi.A, Kipman.A, et al. Efficient human pose estimation from single depth images. IEEE Trans, Pattern Anal. Mach. Intell. (PAMI) 2013, 35, 2821-2840.

[6] Hemandez-Vela, A.Zlateva, N.Marinov, A.Rayees, M.Radeva, P.Dimov, D.Escalera, S.Graph Cuts optimization for Multi-limb Human segmentation in Depth Maps. In Processing of IEEE conference on computer vision and paper(CVPR), St.Paul, 14 – 18th may 2012,

[7] Gananpathi.V, Plegamann,C, Koller.D, Tharun.S.Real time motion capture using a single time-of-flight camera., In Proceedings of IEEE conference on computer vision and pattern recognition (CVPR), San Francisco, CA, USA, 13 – 18 June 2010.

[8] Ye.M, Wang.X, Yang.R, Ren.L, Pollefeys.M, Accurate 3D pose estimation from a single depth image. In Proceedings of IEEE International conference on computer vision (ICCV), Bercelona, sain, 6-13 November 2011

[9] Siddiqui.M, Medioni.G, Human pose estimation from a single view point, real-time range sensor. In proceedings of IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW), San Francisco, CA, USA, 13-18 June 2010.

[10] Wu.J, Konrad.J, Ishwar P, Dymanic time warping for gesture–based user Identification and authentication with kinect. In proceedings of IEEE international conference on Aucostics, speech and signal processing (ICASSP), Vancouver, BC, Canada, 26-31 May 2013.

[11] Kim.H, Hong.S.H, Myung.H, Gesture recognition algorithm for moving kinect sensor. In proceedings of IEEE international workshop on robots and human interactive communications (RO-MAN), Gyeongju, Korea, 26-29 august 2013.

[12] Xia.L, Chen.C.C, Aggarwal.J, ViewInvariant Human action Recognition Using Histograms of 3D Jionts. In Proceedings of IEEE computer society conference on computer vision and pattern recognition workshops (CVPRW), Providence, RI, USA, 16-21 June 2012.

[13] Sigalas.M, Baltzakis.H, Trahanias.P, Gesture recognition based on Arm tracking for human robot Interaction. In the proceedings of IEEE/RSJ International conference on Intelligent Robots anfd systems (IROS), Taipei, Taiwan, 18-22 october 2010.

[14] Achanta.R; Shaji.A, Smith.K, Lucchi.K, Lucchi.A, Fua.P, Susstrunk.S, SLIC superpixels compared to state of art superpixels methods. IEEE Trans. Pattern Anal.Maech. Intell. (PAMI) 2012.

[15] ASUS Xtion pro. Available online: http://www.asus.com/Multimedia/Xtion_PRO/ (accessed on 26 February 2015). Available online: http://www.openni.ru/files/nite (accessed on 26 February 2015).

[16] IPi Mocap Studio. Available online: http://ipisoft.com/ (accessed on February 2015

[17] R.Poppe, A survey on vision-based human action recognition, Imange and vision computing, vol 28 no 6, pp. 676-990, 2010.

[18] S.M.M.Ahsan, J.K.Tan, H.Kim and S.Ishikawa, Histogram of Spatio-temporal binary patterns for human action recognition, Joint Int.Conf.on soft computing Intelligent systems and Int. Symposium on advanced Intelligent systems, 1007-10011,2014.