International Journal of Emerging Technology in Computer Science & Electronics (IJETCSE)
ISSN: 0976-1353 Volume 24 Issue 13 – OCTOBER 2017.

Human Pose Estimation in Video using K-NN and LBP

Vandana V Bhat [#1] and Kavitha G [*2]

[#] *M.Tech(4th Sem), CSE, UBDTCE, Davanagere*

[*] *Assistant Professor, CSE, UBDTCE, Davanagere*

***Abstract*— Human action recognition is the process of labeling image sequences with action labels. Robust solutions to this problem have applications in domains such as visual surveillance, video retrieval and human–computer interaction. The task is challenging due to variations in motion performance, recording settings and inter-personal differences. Recognizing basic human actions from a monocular view is an important task for many applications such as video surveillance, human computer interaction and video content retrieval. Automatic recognition of human activities in video would be useful for surveillance, content-based summarization, and human-computer interaction applications, yet it remains a challenging problem. Some approaches seek ways to measure directly how humans are moving in the scene, using techniques for tracking, body pose estimation, or space-time shape templates while others aim to categorize activities based on the video's over- all pattern of appearance and motion. LBP is used for feature extraction. KNN Classifier is used for classification.**

***Index Terms*—About four key words or phrases in alphabetical order, separated by commas.**

I. INTRODUCTION

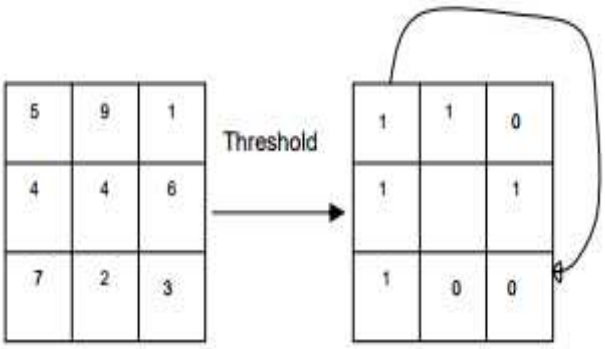
Human action recognition is the process of labeling image sequences with action labels. Robust solutions to this problem have applications in domains such as visual surveillance, video retrieval and human–computer interaction. The task is challenging due to variations in motion performance, recording settings and inter-personal differences. Recognizing basic human actions from a monocular view is an important task for many applications such as video surveillance, human computer interaction and video content retrieval. Automatic recognition of human activities in video would be useful for surveillance, content-based summarization, and human-computer interaction applications, yet it remains a challenging problem. Some approaches seek ways to measure directly how humans are moving in the scene, using techniques for tracking, body pose estimation, or space-time shape templates while others aim to categorize activities based on the video's over- all pattern of appearance and motion, often using spatio-temporal interest operators and local descriptors to build the representation.

II. PREVIOUS WORK

A review of the literature on people tracking is well beyond the scope of this paper. We focus our attention here on the work most similar in spirit to ours. Many early approaches [12]–[18] were based on simple appearance models (e.g., silhouettes) and performed tracking using stochastic search with kinematic constraints. However, silhouette extraction becomes unreliable because of complex backgrounds, occlusions, and moving cameras. Moreover, stochastic search in these high-dimensional paces is notoriously difficult. Facilitated by the advances in human detection methods [4]–[7], [19], tracking by detection has been a focus of recent work. For instance, Andriluka et al. [20], [21] combined the initial estimate of the human pose across frames in a tracking-by-detection framework. Sapp et al. [22] coupled locations of body joints within and across frames from an ensemble of tractable sub-models. Wu and Nevatia [23] propose an approach for detecting and tracking partially occluded people using an assembly of body parts. Such tracking-by-detection approaches are attractive because they can avoid drift and recover from errors. The most similar work to ours are the recent fusion method by stitching together N-best hypotheses from frames of a video. Burgos et al. [24] merged multiple independent pose estimates across space and time using a non-maximum suppression. Park and Ramanan [25] generated multiple diverse high-scoring pose proposals from a tree-structured model and used a chain CRF to track the pose through the sequence. Inspired by the recent success on using convolutional neural network (CNN) [26] for the task of human body pose detection, Jain et al.. [27] proposed MoDeep for articulated human pose estimation in videos using a CNN architecture, which incorporates both color and motion features. Compared to these methods, our work enforces temporal consistency by matching video trajectories to a spatio-temporal 3D model, and provide robustness to view-point changes.

III. LBP

The original LBP operator was introduced by Ojala. The operator labels the pixels of an image by thresholding the 3 X 3 neighbourhood of each pixel with the center value and considering the result as a binary number. Then the histogram of the labels can be used as a texture descriptor.



Binary: 11010011

The limitation of the basic LBP operator is its small 3 X 3 neighbourhood cannot capture dominant features with large scale structures. Hence the operator was extended to use neighbourhood of different sizes [9]. Using circular neighbourhoods and bilinearly interpolating the pixel values allow any radius and number of pixels in the neighbourhood. At a center pixel tc, each neighboring pixel is assigned with a binary label, which can be either "0" or "1," depending on whether the center pixel has higher intensity value than the neighboring pixel (see Fig. 1 for an illustration). The neighboring pixels are the angularly evenly distributed sample points over a circle with radius R centered at the center pixel.

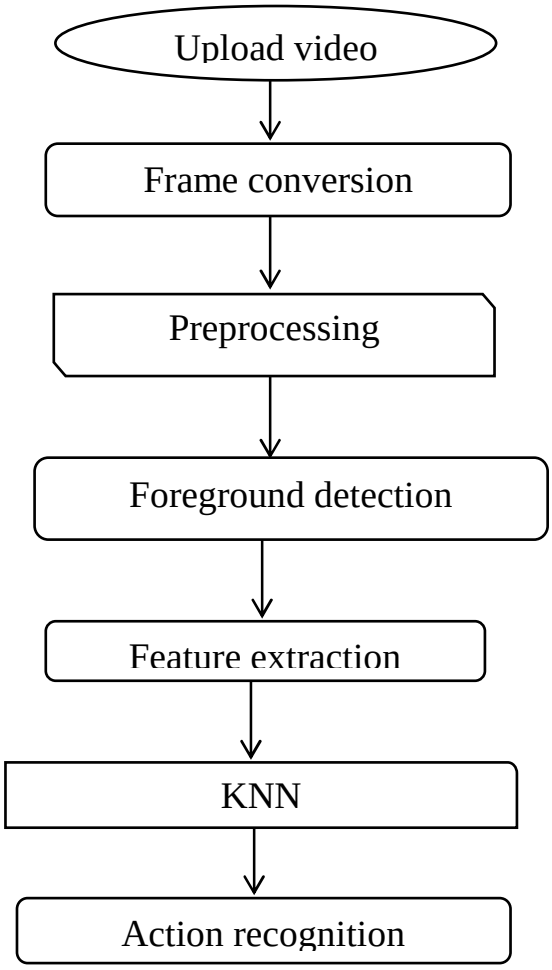
IV. K-Nearest Neighborhood (KNN) Classifier

K-nearest neighbor algorithm is a technique for classifying data based on the closest training examples in the feature space. Before using the KNN the protocol should be followed, i.e. given as: First the dataset is divided into a testing and training set. For each row in the testing set, the „K" nearest training set objects is found, and the classification of test data is determined by majority vote with ties are broken at random. If there are ties for the Kth nearest vector then all the instances are included in the vote. The way the KNN classifier works is, first by calculating the distances between the testing data vector and all of the training vectors using a particular distance calculation methodology which is given as follows: Considering the case of two input variable; the Euclidean distance between two input vectors p and q is computed as the magnitude of difference in vectors i.e. p - q , Where both the data are having „m" dimensions i.e. p= (p1, p2… pm) and q= (q1, q2,…,qm). The Euclidean distance between „p" and „q" is found to be

$$D(p,q) = |p - q|$$
$$= \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_m - q_m)^2}$$

The KNN classifier takes the test instance „x" and finds the Knearest neighbors in the training data and assigns „x" into the class occurring most among the K neighbors.

Methodology



Input video is acquired and frames are extracted. Then frames will be resized. Resized frame is converted to grayscale and video is reconstructed. Then filter is applied to obtain clear frames. Image is converted to binary and edge detection is done. LBP features will be extracted and trained using knn algorithm. Knn classifier is used to detect the pose in the video.

V. CONCLUSION

In this paper, the system has proposed a view based algorithm. It has used *K- nearest neighbours* (KNN) that inherently provides slight invariance to translational and rotational shifts, partial occlusions as well as background noise. Recognition of group activities is fundamentally different from single, or multi-user activity recognition in that the goal is to recognize the behavior of the group as an entity, rather than the activities of the individual members within it. Group behavior is emergent in nature, meaning that the properties of the behavior of the group are fundamentally different then the properties of the behavior of the individuals within it, or any sum of that behavior. The main challenges are in modeling the behavior of the individual group members, as well as the roles of the individual within the group dynamic and their relationship to emergent behavior of the group in parallel. Challenges which must still be addressed include quantification of the behavior and roles of individuals who join the group, integration of explicit models for role description into inference algorithms, and scalability evaluations for very large groups and crowds. This system can accurately perform the human activity recognition.

REFERENCES

[1] Y. Tian, R. Sukthankar, and M. Shah, “Spatiotemporal deformable part models for action detection,” in Pro c . IEEE Conf . Comput . Vis . Pattern Recog ., Jun. 2013, pp. 2642–2649.
[2] Y. Hu, L. Cao, F. Lv, S. Yan, Y. Gong, and T. S. Huang, “Action detection in complex scenes with spatial and temporal ambiguities,” in P ro c . I E E E C o n f . C o m p u t . Vi s . P a t t e rn R e c o g ., Sep.–Oct. 2009, pp. 128–135.
[3] S. Ma, J. Zhang, N. Ikizler-Cinbis, and S. Sclaroff, “Action recognition and localization by hierarchical space-time segments,” in Proc . IEEE Int . Conf . Comput . Vis ., Dec. 2013, pp. 2744–2751.
[4] T. Lan, Y. Wang, and G. Mori, “Discriminative figure-centric models for joint action localization and recognition,” in Proc . IEEE Int . Conf .Comput . Vis ., Nov. 2011, pp. 2003–210.
[5] M. Raptis, I. Kokkinos, and S. Soatto, “Discovering discriminative action parts from mid-level video representations,” in Proc . IEEE Conf . C o mput . Vis . Pattern Recog ., Jun. 2012, pp. 1242–1249.
[6] F. Shi, Z. Zhou, J. Xiao, and W. Wu, “Robust trajectory clustering for motion segmentation,” in Proc . IEEE I n t . Conf . Comput . Vis., Dec. 2013, pp. 3088–3095.
[7] A. Gaidon, Z. Harchaoui, and C. Schmid, “Recognizing activities with cluster-trees of tracklets,” in Proc . Brit . Mach . Vis . Conf ., Sep. 2012, pp. 30.1–30.13.
[8] Y.-G. Jiang, Q. Dai, X. Xue, W. Liu, and C.-W. Ngo,“Trajectory-based modeling of human actions with motion reference points,” in Proc . Eur . Conf . Comput . Vis ., Oct. 2012, vol. 7576, pp. 425–438.
[9] M. D. Rodriguez, J. Ahmed, and M. Shah, “Action MACH: A spatiotemporal maximum average correlation height filter for action recognition,” in Pro c . IEEE Con f . Compu t . Vi s . Pattern Recog ., Jun. 2008, pp. 1–8.
[10] I. Laptev, M. Marszalek, C. Schmid, and B. Rozenfeld, “Learning realistic human actions from movies,” in P ro c . I E E E C o n f . C o m p u t . Vi s . P a t t e r n R e c o g ., Jun. 2008, pp. 1242–1249.