A REVIEW ON IMAGE BASED PERSON AND OBJECT RECOGNITION USING SIFT AND HMM

Aanchal Verma[1], Ratnesh Srivastava[2], H.L Mandoria[3] Binay Pandey[4]

[1] M.Tech Student, Department of Information Technology, GBPUAT Pantnagar, Uttarakhand, India
[2 4]Asst. Professor, Department of Information Technology, GBPUAT Pantnagar, Uttarakhand, India
[3]Professor and Head, Department of Information Technology, GBPUAT Pantnagar, Uttarakhand, India

---***---

Abstract - *This paper proposes a new method i.e. SIFT(scale-invariant feature transform) and HMM(hidden markov modeal) to develop a program capable of performing recognition of individuals from an image sequence of a person and object. The program should be able to store the derived object for comparison at a later stage. Automatic extraction of relevant object feature points should be available from an image sequence in order to automate the classification process. This process will significantly reduce the test-time computational complexity and also retain or even improve the recognition accuracy. When compared with other existing methods in the literature, the proposed method is shown to have the promising performance for the case of object and person.*

Key Words: SIFT , HMM, FAR, FRR, EER, DoG etc

1.INTRODUCTION

As of late, Activity recognition has picked up ubiquity and has been demonstrated powerful for some application situations and a critical zone of PC vision research and application. Activity recognition is one of the vital research fields to acknowledge psychological capacity of the PCs, and is relied upon to be connected to Robot eyes or head mounted show. As of late, manual recovery and arrangement of the picture end up troublesome as volume of information ends up immense. Here we are endeavoring to perceive action over a grouping of pictures. Consequently we required a mechanized investigation instead of human administrators checked it. Electronic Activity recognition framework moves toward becoming unmistakable quality in such situation. The issue in the Activity recognition is to manage the pivots of the action, scale changes, and light changes. Also, there is the issue of impediment that makes the Activity recognition troublesome.

We have proposed an effective calculation for Activity recognition utilizing SIFT and HMM. In the robotized investigation we will separate element first utilizing filter calculation and will apply HMM approach edge to casing to perceive movement. The separated highlights are followed casing to outline. The followed highlights can be investigated to perceive action.

2. RELATED WORK

D. Zhang, et al. [01], This paper focuses on the detection of the abnormal motion behaviour recognition of the crowd, and proposes an innovation method which is consist of three steps, i.e. SIFT flow + weighted orientation histogram + Hidden Markov Model(HMM). Analogous to optical flow, which is used to get the motion information of the pixels from two adjacent frames, SIFT flow is of higher precision. Next, we build up a weighted orientation histogram as a statistical measurement for the SIFT flow features from the first step. Finally, the derived histogram is taken as the input for HMM in preparation for the detection of abnormal crowd motion. Experimental results show that compared to the existing method, our proposed one can detect the abnormal motion behaviour more effectively.

Shivakanth, et al. [02], In this paper, the object recognition system which can resolve the difficulty of rotations of object, scale changes and illumination are resolved with the help of “SIFT algorithm”. It is implemented with different phases such as scale space extreme detection, key point localization. The SIFT algorithm implemented with MATLAB, which is one of efficient tool to perform image processing.

J. L. MazherIqbal, et al. [03],The paper discusses abnormalities in the human activity and provides efficient solution to detect abnormality. The first step in the proposed work is to capture the video using webcam and then to detect abnormal behaviour. The captured video is divided into frames and extracts the features such as edges and boundaries using scale invariant feature transform (SIFT). Feature vectors are developed from the extracted features. These feature vectors are compared using Hidden Markov Model with the data set developed to recognize abnormal behavior. If there is a match between feature vectors and available data set, then abnormal human activity is detected and simultaneously an alarm is given for medical assistance. The proposed algorithm is tested on six different activities. The proposed methods achieve accurate recognition.

N. Shukla et al. [04], In Handwritten signatures analyzed for forgery have to undergo feature extraction process, due to varied samples in size rotation and intra-domain changes, invariance has to be achieved during feature extraction process; circular Hidden Markov Model with discrete radon transform approach of feature extraction provides invariance. On other hand Scale Invariant Feature Transform (SIFT) has inherent invariant feature extraction approach. This paper compares both approaches on common signature databases for False acceptance rate(FAR),False Rejection Rate(FRR) and Equal Error Rate(EER).

S. Pandita, et al. [05], This paper presents a method for recognizing hand gestures by extracting distinctive invariant features from images that can be used to perform efficient matching between different views of a hand gesture. The features are invariant to image scale and rotation, and provide robust matching across a considerable range of affine distortion, change in 3D viewpoint, addition of noise, and change in illumination.

G. K. Yadav, et al. [06], Human action recognition has been a challenging task in computer vision because of intra-class variability. State-of-the-art methods have shown good performance for constrained videos but have failed to achieve good results for complex scenes. Reasons for their failing include treating spatial and temporal dimensions without distinction as well as not capturing temporal information in video representation. To address these problems we propose principled changes to an action recognition framework that is based on video interest points (IP) detection with capturing differential motion as the central theme. First, we propose to detect points with high curl of optical flow, which captures relative motion boundaries in a frame. We track these points to form dense trajectories. Second, we discard points on the trajectories that do not represent change in motion of the same object, yielding temporally localized IPs. Third, we propose a video representation based on spatio-temporal arrangement of IPs with respect to their neighboring IPs. The proposed approach yields a compact and information-dense representation without using any local descriptor around the detected IPs.

Thanh Phuong Nguyen, Antoine Manzanera [07], A new spatio temporal descriptor is proposed for action recognition. The action is modelled from a beam of trajectories obtained using semi dense point tracking on the video sequence. We detect the dominant points of these trajectories as points of local extremum curvature and extract their corresponding feature vectors, to form a dictionary of atomic action elements. The high density of these informative and invariant elements allows effective statistical action description. Then, human action recognition is performed using a bag of feature model with SVM classifier. Experimentations show promising results on several well-known datasets.

3. IMPORTANT ALGORITHMS

3.1 SIFT

A SIFT feature is a selected image region (also called keypoint) with an associated descriptor. Keypoints are extracted by the SIFT detector and their descriptors are computed by the SIFT descriptor. It is also common to use independently the SIFT detector (i.e. computing the A SIFT feature is a selected image region (also called keypoint) with an associated descriptor. Keypoints are extracted by the SIFT detector and their descriptors are computed by the SIFT descriptor. It is also common to use independently the SIFT detector (i.e. computing the keypoints without descriptors) or the SIFT descriptor (i.e. computing descriptors of custom keypoints). keypoints without descriptors) or the SIFT descriptor (i.e. computing descriptors of custom keypoints). To recognize and classify objects efficiently, feature points from objects can be extracted to make a robust feature descriptor or representation of the objects. David Lowe has introduced a technique called Scale Invariant Feature Transform (SIFT) to extract features from images. These features are invariant to scale, rotation, partial illumination and 3D projective transform and they are shown to provide robust matching across a substantial range of affine distortion, change in 3D viewpoint, addition of noise and change in illumination. SIFT features provide a set of features of an object that are not affected by occlusion, clutter and unwanted noise in the image. In addition, SIFT features are highly distinctive in nature which have accomplished correct matching on several pair of feature points with high probability between a large database and a test sample. Following are the four major filtering steps of computation used to generate the set of image feature based on SIFT.

3.1.1 Scale-Space Extrema Detection:

The original SIFT descriptor (Lowe 1999, 2004) was computed from the image intensities around interesting locations in the image domain which can be referred to as interest points, alternatively key points. These interest points are obtained from scale-space extrema of differences-of-Gaussians (DoG) within a difference-of-Gaussians pyramid. The concept of difference-of-Gaussian bandpass pyramids was originally proposed by Burt and Adelson (1983) and by Crowley and Stern (1984). A Gaussian pyramid is constructed from the input image by repeated smoothing and subsampling, and a difference-of-Gaussians pyramid is computed from the differences between the adjacent levels in the Gaussian pyramid. Then, interest points are obtained from the points at which the difference-of-Gaussians values assume extrema with respect to both the spatial coordinates in the image domain and the scale level in the pyramid.

This filtering approach attempts to identify image locations and scales that are identifiable from different views. Scale space and Difference of Gaussian (DoG) functions are used to detect stable keypoints. Difference of Gaussian is used for identifying key-points in scale-space and locating scale space extrema by taking difference between two images, one with scaled by some constant time of the other. To detect the local maxima and minima, each feature point is compared with its 8 neighbors at the same scale and in accordance with its 9 neighbors up and down by one scale. If this value is the minimum or maximum of all these points then this point is an extrema.

3.1.2. Keypoints Localization in Laplacian Space:

To localize keypoints, few points after detection of stable keypoint locations that have low contrast or are poorly localized on an edge are eliminated. This can be achieved by calculating the Laplacian space. After computing the location of extremum value, if the value of difference of Gaussian pyramids is less than a threshold value, the point is excluded. If there is a case of large principle curvature across the edge but a small curvature in the perpendicular direction in the difference of Gaussian function, the poor extrema is localized and eliminated.

3.1.3. Assignment of Orientation:

A peak in the DoG scale space fixes 2 parameters of the keypoint: the position and scale. It remains to choose an orientation. In order to do this, SIFT computes a histogram of the gradient orientations in a Gaussian window with a standard deviation which is 1.5 times bigger than the scale σ of the keypoint.

This step aims to assign consistent orientation to the key-points based on local image characteristics. From the gradient orientations of sample points, an orientation histogram is formed within a region around the key-point. Orientation assignment is followed by key-point descriptor which can be represented relative to this orientation. A 16x16 window is chosen to generate histogram. The orientation histogram has 36 bins covering 360 degree range of orientations. The gradient magnitude and the orientation are pre-computed using pixel differences. Each sample is weighted by its gradient magnitude and by a Gaussian-weighted circular window.

3.1.4. Keypoint Descriptor:

A SIFT descriptor of a local region (keypoint) is a 3-D spatial histogram of the image gradients. The gradient at each pixel is regarded as a sample of a three-dimensional elementary feature vector, formed by the pixel location and the gradient orientation. Samples are weighed by the gradient norm and accumulated in a 3-D histogram, which (up to normalization and clamping) forms the SIFT descriptor of the region. An additional Gaussian weighting function is applied to give less importance to gradients farther away from the keypoint center.

In the last step, the feature descriptors which represent local shape distortions and illumination changes are computed. After candidate locations have been found, a detailed fitting is performed to the nearby data for the location, edge response and peak magnitude. To achieve invariance to image rotation, a consistent orientation is assigned to each feature point based on local image properties. The histogram of orientations is formed from the gradient orientation at all sample points within a circular window of a feature point. Peaks in this histogram correspond to the dominant directions of each feature point. For illumination invariance, 8 orientation planes are defined. Finally, the gradient magnitude and the orientation are smoothened by applying a Gaussian filter and then are sampled over a 4 x 4 grid with 8 orientation planes.

3.2 Hidden Markov Model

The Hidden Markov Model (HMM) is a powerful statistical tool for modeling generative sequences that can be characterised by an underlying process generating an observable sequence. HMMs have found application in many areas interested in signal processing, and in particular speech processing, but have also been applied with success to low level NLP tasks such as part-of-speech tagging, phrase chunking, and extracting target information from documents. Andrei Markov gave his name to the mathematical theory of Markov processes in the early twentieth century, but it was Baum and his colleagues that developed the theory of HMMs in the 1960s.

Hidden Markov models (HMMs) are a formal foundation for making probabilistic models of linear sequence 'labeling' problems. They provide a conceptual toolkit for building complex models just by drawing an intuitive picture. They are at the heart of a diverse range of programs, including gene finding, profile searches, multiple sequence alignment and regulatory site identification. HMMs are the Legos of computational sequence analysis. N HMM is a doubly stochastic process with an underlying stochastic process that is not observable (it is hidden), but can only be observed through another set of stochastic processes that produce the sequence of observed symbols. A hidden Markov model (HMM) is one in which you observe a sequence of emissions, but do not know the sequence of states the model went through to generate the emissions. Analyses of hidden Markov models seek to recover the sequence of states from the observed data.

3.2.1 HMM Applications

Stock market: bull/bear market hidden Markov chain, stock daily up/down observed, depends on big market trend.

- Speech recognition: sentences & words hidden Markov chain, spoken sound observed (heard), depends on the words
- Digital signal processing: source signal (0/1) hidden Markov chain, arrival signal fluctuation observed, depends on source
- Bioinformatics: sequence motify finding, gene prediction, genome copy number change, protein structure prediction, protein-DNA interaction prediction

We seek a way to build a representation for the gait of every individual in the database. We opt for the stochastic approach of using HMMs. In this case, training involves learning the HMM parameters Y = (A; B; II). Here A denotes the transition probability matrix, B is the observation probability and II is the initial probability vector. In order to capture the gait of an individual, we train the HMM using the width vectors derived from the silhouette for several gait cycles of the person. We express the pdf of the observation as

$$b_j(\mathbf{o}) = \mathcal{N}(\mathbf{o}; \mu_\mathbf{j}, \mathbf{U_j}), \quad 1 \leq j \leq N \qquad (1)$$

Where o is the observation vector and $u_{j,}$ N is the number of states in the HMM and U_j are the mean and covariance, respectively. The reliability of estimates of B depends on the number of training samples available and the dimension of the observation vector.

In a practical situation, only a finite amount of training data is available. Since the means and covariance in equation (1) have to be learnt from the training samples, the dimension of the observation vector becomes critical. The required number of training samples increases with the dimensionality of the observation vector. To be precise, assume for the moment that the data can be modeled by a single Gaussian distribution. Then, for a d-dimensional observation vector, we need at least d training samples to estimate the centroid and d* (d+1)/2 training samples in order that the covariance matrix would have a well-defined inverse. In our experiments, the smallest dimension of the width vector of the silhouette is approximately 100. This implies that we require at least 100 observations to learn the mean value. To learn the covariance, we would need as many as 5150 vectors! For a mixture of *m*-Gaussian model, there would be a further *m*-fold increase in the number of training vectors. Clearly, the possibility of using the width vector directly is ruled out. A more compact way of encoding the observation, while retaining all relevant information, is needed. We propose the following methodology to tackle the dimensionality issue in the gait problem. To decide on the number of stances we plot the average rate distortion curve of the quantization error as a function of number of stances. We observe that the quantization error does not decrease appreciably beyond 5 stances. Let us denote the width vectors corresponding to the five stances for the j^{th} person as S_1^j.... S_5^j. These stances are the ones that result from application of the Active Shape Model procedure to the training data available for that individual.

4 CONCLUSIONS

This paper has proposed a new method of activity recognition which includes SIFT (Scale Invariant Feature Transform) and HMM (Hidden Markov Model). SIFT used Feature extraction set of key points make a Feature set. Hidden Markov Model has been successfully applied in Object recognition. They assume that the current state is only influenced by the previous state and is independent of the history state. The postures of the subjects are regarded as the states of HMM and HMM parameters are trained by binaries Image feature vectors. After analyzing the working the final conclusion computed was the performance was better than the previous work.

REFERENCES:

[1] Dongping Zhang, KaihangXu, HuailiangPeng, Ye Shen, "Abnormal Crowd Motion Behaviour Detection based on SIFT Flow", International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol.9, No.1 (2016), ISSN: 2005-4254,

http://dx.doi.org/10.14257/ijsip.2016.9.1.28,

[2] Shivakanth, Archana Mane, "Object Recognition using SIFT", IJISET - International Journal of Innovative Science, Engineering & Technology, Vol. 1 Issue 4, June 2014, ISSN 2348 – 7968,

[3] J. L. MazherIqbal, J. Lavanya and S. Arun, "Abnormal Human Activity Recognition using Scale Invariant Feature Transform", International Journal of Current Engineering and Technology, Vol.5, No.6 (Dec 2015), E-ISSN 2277 – 4106, P-ISSN 2347 – 5161,

[4] NeerajShukla Dr. MadhuShandilya, "Invariant Features Comparison in Hidden Markov Model and SIFT for Offline Handwritten Signature Database", International Journal of Computer Applications (0975 – 8887) Volume 2 – No.7, June 2010,

[5] S. Pandita, S. P. Narote, "Hand Gesture Recognition using SIFT", International Journal of Engineering Research & Technology (IJERT), Vol. 2 Issue 1, January- 2013, ISSN: 2278-0181,

[6] Gaurav Kumar Yadav, PrakharShukla, AmitSethi, "ACTION RECOGNITION USING INTEREST POINTS CAPTURING DIFFERENTIAL MOTION INFORMATION", Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference, DOI: 10.1109/ICASSP.2016.7472003, 19 May 2016, ISSN: 2379-190X,

[7] AshwanAnwerAbdulmunem, "Human Action Recognition using Saliency-based Global and Local Features", December 2017, Cardiff University, department of Computer Science & Informatics, http://orca.cf.ac.uk/id/eprint/107750,

[8] Thanh Phuong Nguyen, Antoine Manzanera, "ACTION RECOGNITION USING BAG OF FEATURES EXTRACTED FROM A BEAM OF TRAJECTORIES", Image Processing (ICIP), 2013 20th IEEE International Conference, ISSN: 2381-8549, 13 February 2014, DOI: 10.1109/ICIP.2013.6738897,

[9] MohibUllah, HabibUllah and Ibrahim M. Alseadoon, "HUMAN ACTION RECOGNITION IN VIDEOS USING STABLE FEATURES", Signal & Image Processing : An International Journal (SIPIJ) Vol.8, No.6, December 2017, DOI: 10.5121/sipij.2017.8601

[10] Shugang Zhang, Zhiqiang Wei, JieNie, Lei Huang, Shuang Wang, "A Review on Human Activity Recognition Using Vision-Based Method", Journal of Healthcare Engineering, Volume 2017, Article ID 3090343, https://doi.org/10.1155/2017/3090343

[11] Guangyu Zhu, Ming Yang, Kai Yu, Wei Xu, Yihong Gong, "Detecting Video Events Based on Action Recognition in Complex Scenes Using Spatio-Temporal Descriptor", Proceeding MM '09 Proceedings of the 17th ACM international conference on Multimedia, Pages 165-174, October 19 - 24, 2009, ACM New York, ISBN: 978-1-60558-608-3 doi>10.1145/1631272.1631297,

[12] Jianxin Wu, AdebolaOsuntogun, TanzeemChoudhury, "A Scalable Approach to Activity Recognition based on Object Use", Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference, 26 December 2007, ISBN: 978-1-4244-1631-8,DOI: 10.1109/ICCV.2007.4408865