Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 6, June -2017

Image Classification using Naive Bayes Model for Deep Head Pose Estimation

Priyanka Jadhav.[1], Mrs. Shanthi K. Guru[2]

[1]ME Student, Department of Computer Engineering, DYPCOE, Akurdi, SPPU, Pune, India[1]
[2]Assistant Professor, Department of Computer Engineering, DPCOE, Akurdi, SPPU, Pune, India[2]

Abstract — *In visiual survillence and Human Computer Interactions (HCI), human head pose is an important feature in estimating human focus-of-attention where close level iris/eye tracking is not possible. Head-pose estimation suffers from two major problems, one is to localize the head in low resolution images and second, different poses of the same person may appear more similar compared to the same head-pose of different persons. To overcome these problems, a Naive Bayes classification model is introduced for human head pose estimation in low resolution using RGB or RGB-D data and pose the problem as one of classification of human gazing direction. The proposed system works more efficient and helps to improve classification accuracy and gives accurate results with less number of training samples with simpler structure and requires less time for execution. Using this probabilistic model, many higher level scene understanding like human-human/scene interaction detection can be achieved.*

Keywords *- Head pose estimation, RGB-D data, Naive Bayes classifier, Gaze direction.*

I. INTRODUCTION

In computer vision and signal processing, modelling human head pose is not an easy task and also determining human head position and it's orientation is difficult. Head pose estimation is the process of determining the orientation of human head in digital imagery. Head pose estimation and gaze direction are important in many applications like visual surveillance, human computer interaction for the analysis of human behaviour. The gaze is nothing but the combination of head pose and eye location. Head pose estimator can demonstrate invariance in images or image sequences. Head pose estimation can be applied on different algorithms for identification of frontal and non-frontal faces[1].

Head pose signals contains meta information about communication gesture. This information is useful in various application such as anomaly detection, group detection and crowd behavioural dynamics and tracking[2]. Head pose estimation is used in such domains where close level iris or eye tracking is not possible. In visual surveillance and Human Computer Interactions (HCI) two separate and distinct regions are present with different methodologies required due to the difference in the quality of the input.

In higher resolution of HCI domain, facial landmark detec-tion approaches are employed for better accuracy. In surveillance videos for head-pose and body-pose estimation a multi-level Histogram of Oriented Gradients (HOG)[3] used for the head and body pose features for extracting a feature vector for an adaptive classification using high dimensional kernel space methods[1].

Deep learning CNNs are used to learn robust non-linear representations from input data. CNN are supervised, discriminative and have mostly surpassed the Deep Belief Newtworks(DBFs) in terms of accuracy on large labelled datasets like the Imagenet[4]. In recent years, head pose estimation can be done using RGB or RGB depth data. The depth data can be used to enhance RGB-based head pose estimation and human tracking. In low resolution, human gazing direction problem can be solved by using RGB-D data classification[5].

To estimate the head pose and body pose estimation there are different approaches like Non-linear regression approaches like Artificial Neural Networks and high-dimensional manifold based approaches, etc. The Human Computer Interaction, the formulation of head pose is limited to 2 meter distance from the sensor along with near-frontal head-poses. In previous works, low resolution head pose estimation was used a detector based on template training to classify head poses in eight directional bins. Head-pose estimation faces two major problems:

1) It is difficult to localize the head poses in low resolution,

2) Different poses of the same person may appear more similar compared to the non-frontal head-poses of different persons.

To overcome these problems, a Naive Bayes classifier is used for the classification of frontal and near-frontal images. Naive Bayes classifier is used due to it's specific features like simpler structure and requires less time for execution than neural network. Also useful with less number of training samples. Section I defines introduction about Head pose

estimation for gaze direction, section II includes literature survey and section III includes system architecture, section IV describes system analysis, section V describes results and section VI includes conclusion.

II. RELATED WORK

Head pose estimation, Head Pose Classification are the main threads :

A. HeadPose Estimation

Robertson and Reid [6] proposed a feature based vectors model on skin detection to classify head poses in 8 different orientations in low resolutions. This technique was extended by Benfold et al.[7]. They proposed algorithm for classification of head poses in low resolution and mapping between colours and labels. But all template based methods faces problems such as localize head poses in low resolution is difficult, and non-frontal head poses may appear more similar. To avoid such problems some researchers proposed different feature space for representing head images.

A Neural Network based approach was proposed by Stiefelhagen [8] to estimate horizontal and vertical head orientation of a person from facial expressions. Non-linear regression like high-dimensional manifold based approach was proposed to determine head pose and face images in various pose angles [9]. Chen and Odobez [3] proposed multi-level Histogram of Oriented Gradients (HOG) based method for head-pose and body-pose estimation in surveillance videos.

On the other hand, in HCI domain problem solutions are limited to 2 meter distance from the sensor along with near-frontal head-poses. An iterative closest point (ICP) based mesh fitting method has been proposed for head pose detection[10]. Work on head pose regression has been introduced for scene and human interaction understanding[3]. This work focuses on head-pose regression and interaction detection in 2D movies/tv-series scenes.

Recently, manifold based metric learning methods have been applied to head pose estimation[11]. In other approaches, the spherical nature of the view manifold of objects is used as a strong prior for manifold learning[12]. Faces are detected using chrominance-based features, Grey-level normalized face imagettes serve as input for linear auto-associative memory over a wide range of angles from low-resolution images[13]. Head pose estimation can be useful in many applications like anomaly detection, crowd behavioral dynamics. Head pose estimation provides a interface for computing. Some existing examples includes control to computer mouse using head pose movements, respond to pop-up dialog boxes with head nods and shakes or use head gestures to interact with embodied agents. Murphy-Chutorian et. al. [14] proposed system to estimate a drivers head pose. The system is fully autonomous and operates online in daytime and nighttime driving conditions, using a monocular video camera sensitive to visible and near-infrared light. A new approach is proposed to estimate the head pose from monocular images in three stages. First, a face detector roughly classifies the pose as frontal, left, or right profile. Then, classifiers trained to detect distinctive facial features such as the nose tip and the eyes. Based on the positions of these features, a neural network finally estimates the three continuous rotation angles to model the head pose[15].

B. HeadPose Classification Methods

Deep learning, especially convolutional neural networks (CNNs) are used to learn non-linear representations from input data and have been especially successful on images[4] and audio[16]. But, this is in contrast to traditional computer vision pipelines like HOG[3]. These features would be used as input to machine learning framework like support vector machines (SVM) for achieving classification or regression. In [17], trained a large, deep convolutional neural network is used to classify the 1.2 million high-resolution images in the ImageNet in 1000 different classes[18].

On the other hand, CNNs are supervised, discriminative and have mostly surpassed the Deep Belief Net-works(DBNs) in terms of accuracy on large labelled datasets like the Imagenet[19]. CNNs are deep models which belongs to fully connected networks. CNNs are also applied in multi-modal RGB-D domains. In [20], author introduced a fusion of RGB-D channels and transfer learning for initialisation of the weights of the green, blue and depth channels with filters learned from the depth channel. But this form of early fusion is not very helpful. RGB-D networks are generally trained with late fusion[21] [22].

A Dynamic Bayesian Mixture Model (DBMM) is pro-posed for combine multiple classifier likelihoods into a single form, assigning weights and likelihoods are considered as posterior probability[23]. A computer assisted diagnosis method is proposed by Zhou, Xingxing, et al. based on a Wavelet entropy of the feature space approach and a Nave Bayes classification method is used for improving the brain diagnosis accuracy by means of Nuclear magnetic resonance images[24]. CNNs are used to train large scale labelled training data. The number of parameters in the convolution layers are orders of magnitude lower than the fully connected layer. Separate CNN will be trained on RGB and depth modularities based on the architecture[2].

CNN framework is more complex in structure and takes more execution time. CNN have problem with small number of training samples. To overcome these problems a Naive Bayes Classification method is proposed to

improve accuracy, reduce execution time. The proposed method will give accurate results with less number of training samples.

III. PROPOSED SYSTEM

The main objective of proposed system is to classify RGB or RGB-Depth images using Naive Bayes classifier. Depth features are extracted from input images and given to the Naïve Bayes classifier. By using such inputs the expected outputs are class label for gaze direction[25].

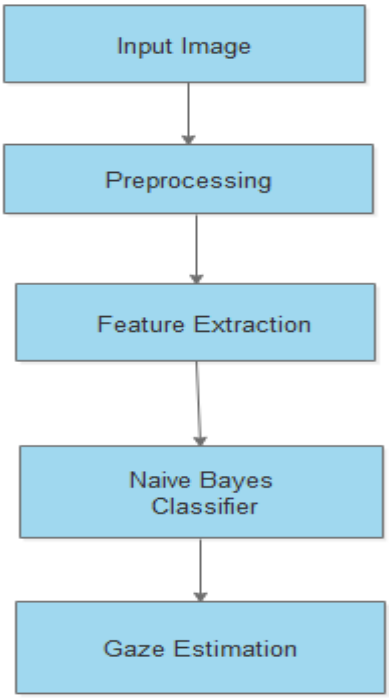

Fig.1. System Architecture

The Fig. 1 shows architecture of proposed system, which takes dataset of RGB or RGB-Depth images as input. The proposed system consist of Preprocessing phase, feature ex-traction, Naive Bayes classifier.

Preprocessing : Only head part of an image is considered for future processing and remaining part of image and noise will be removed.

Feature Extraction : In feature extraction, mean depth and standard deviation are calculated for feature extraction of RGB or RGB Depth images.

Naive Bayes Classifier : All the images are given to the Naive Bayes Classifier for classification and it's output is in the form of Class labels used for gaze estimation.

IV. ALGORITHM

The main purpose of the proposed algorithm is to reduce the execution time and to improve the classification accuracy.

Step 1 : Preprocessing :

In the preprocessing step, Mask image is applied over the input image for getting only head part from image and remaining part of image will be ignored.

Step 2 : Depth Encoding :

1. Encode the depth image using the depth modality with mean depth and standard deviation with the depth data[5]. Depth can be calculated by using pitch, raw and roll angles in spherical co-ordinate system.

2. Mean depth can be calculated by using formula :

$$Mean(\mu) = \frac{1}{N}\sum_{N=0}^{N-1} x_i$$

where, x_i is the depth value, N is the total number of depth values.

3. Standard Deviation can be calculated on three dimensional co-ordinate system by using formula :

$$Standard\ Deviation\ (\sigma^2) = \frac{1}{N-1}\sum_{N=0}^{N-1}(x_i - \mu)^2$$

where, x_i is the depth value, N is the total number of depth value, μ is the Mean of depth values.

Step 3 : Naïve Bayes Classifier

Bayes rule is :

$$P(Y|X_1, X_2, \ldots . X_n) = \frac{P(X_1, X_2, \ldots X_n|Y)P(Y)}{P(X_1, X_2, \ldots X_n)}$$

where, $P(Y|X_1, X_2, \ldots . X_n)$ represents likelihood, P(Y) represents prior and $P(X_1, X_2, \ldots X_n)$ represents normalization constant, Y is the output in the form of class labels as estimated gaze direction.

V. RESULTS AND DISCUSSION

A. DataSet

Experiments mainly conducted on head pose datasets. Here, Biwi Kinect Head Pose Database is used[5]. It is available freely on ("https://data.vision.ee.ethz.ch/cvl/gfanelli/headpose/ head-forest.html")

The database contains 24 sequences acquired with a Kinect sensor. 20 people (some were recorded twice - 6 women and 14 men) were recorded while turning their heads, sitting in front of the sensor, at roughly one meter of distance. For each sequence, the corresponding .obj file represents a head template of the neutral face of that specific person. For each frame, a rgb.png and a depth.bin files are provided, containing color and depth data. The depth is already segmented (the background is removed using a threshold on the distance) and the binary files compressed (an example c code is provided to show how to read and write the depth data into memory). The pose.txt and pose.bin files contain the ground truth information, i.e., the location of the center of the head in 3D and the head rotation. The .txt files encode the rotation as a matrix, while the .bin file contain 6 floats representing the head center coordinates followed by pith, yaw, and roll angles[5]. For the experiment, total 90 RGB images of 2 Females are used and depth, pose, bin files of the respective RGB images are used.

B. Results

The RGB input image is given as a input. RGB mask image and Depth image of respective RGB image is used. Depth features i. e. Mean and Standard Deviation are extracted in feature extraction. The extracted DAE features are given to the Naive Bayes Classifier for the classification. Here, Class 01 to Class 04 are created for training and each class contains 10 images for training and 10 images for testing. As the number of class labels will be increased, the accuracy will be increased. The systems F-score for CNN is 0.80 and Naive Bayes is 0.95.

The number of class labels will be increased in proposed system so that the accuracy will be increased.

TABLE : I
Comparison of F-measure

Algorithms	F-measure
CNN	80
Naïve Bayes	90

The metrics used to evaluate gaze estimation are accuracy, precision, recall and F-score or F-measure for deep head pose classification. Larger the F-measure values are better for gaze estimation that means it can give better classification result.

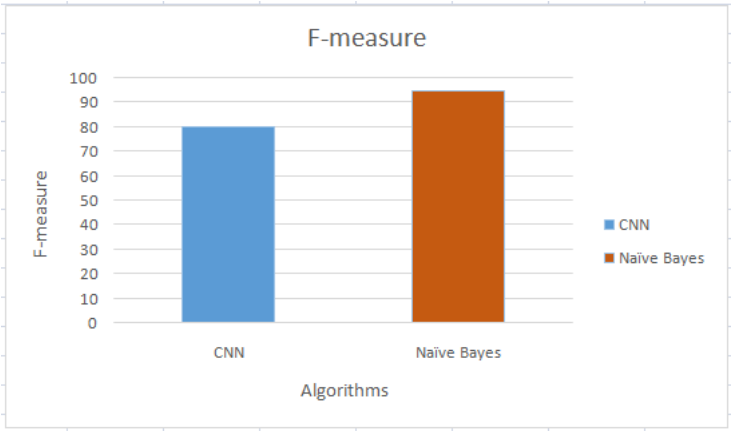

Fig. 2 : Comparision of F-measure of CNN and Naïve Bayes.

The Fig.2 shows the graph of F-measure comparison of CNN and Naive Bayes algorithm, where x-axis shows algorithms and y-axis shows F-measure ratio or F-score. Using Naive Bayes classifier for classification will help to get more accurate correctly classified head pose images. Also accuracy and performance will be parameters on which further results can be examined and improved.

VI. CONCLUSION AND FUTURE WORK

Human head pose is an important feature in estimating human focus-of-attention where close level iris/eye tracking is not possible. In head pose estimation, features are extracted from RGB or RGB-Depth images. Naive Bayes classifier is used to classify images using extracted features and gives output as gaze direction. A novel approach is used towards human attention modeling via head-pose estimation from low to high resolution. Naive Bayes clssification improves the classification accuracy and performance than Convolutional neural network. By using this model human-human/scene interaction detection can be achieved and head pose estimation is also useful in crowd behavioral dynamics and tracking and anomaly detection. In future, to achieve better performance hybrid approach can be build.

REFERENCES

1. Mukherjee, Sankha S., and Neil Martin Robertson. "Deep Head Pose: Gaze-Direction Estimation in Multimodal Video." IEEE Transactions on Multimedia 17.11 (2015): 2094-2107.
2. Baxter, Rolf H., et al. "An adaptive motion model for person tracking with instantaneous head-pose features." IEEE Signal Processing Letters 22.5 (2015): 578-582
3. Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection."2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05). Vol. 1. IEEE, 2005.
4. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Adv. Neural Inf. rocess. Syst., pp. 10971105, 2012.
5. Fanelli, Gabriele, Juergen Gall, and Luc Van Gool. "Real time head pose estimation with random regression forests."Computer Vision and Pattern Recognition (CVPR), IEEE Conference on. IEEE, 2011.
6. Robertson, Neil, and Ian Reid. "Estimating gaze direction from lowresolution faces in video." European Conference on Computer vision. Springer Berlin Heidelberg, 2006.
7. Benfold, Ben, and Ian Reid. "Colour Invariant Head Pose Classification in Low Resolution Video." BMVC. 2008.
8. Stiefelhagen, Rainer. "Estimating head pose with neural networks-results on the pointing04 ICPR workshop evaluation data." Pointing04 ICPR Workshop of the Int. Conf. on Pattern Recognition. 2004.
9. Balasubramanian, Vineeth Nallure, Jieping Ye, and Sethuraman Panchanathan "Biased manifold embedding: A framework for person independent head pose estimation."2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007.
10. Cazzato, Dario, Marco Leo, and Cosimo Distante. "An investigation on the feasibility of uncalibrated and unconstrained gaze tracking for human assistive applications by using head pose estimation." Sensors 14.5 (2014):8363-8379.
11. Peng, Xi, et al. "From circle to 3-sphere: Head pose estimation by instance parameterization." Computer Vision and Image Understanding 136 (2015): 92-102.
12. Ma, Bingpeng, et al. "CovGa: a novel descriptor based on symmetry of regions for head pose estimation." Neurocomputing 143 (2014): 97-108.
13. Fanelli, Gabriele, et al. "Random forests for real time 3D face analysis." International Journal of Computer Vision 101.3 (2013): 437-458

14. .Murphy-Chutorian, Erik, Anup Doshi, and Mohan Manubhai Trivedi. "Head pose estimation for driver assistance systems: A robust algorithm and experimental evaluation." Intelligent Transportation Systems Conference(ITSC), IEEE, 2007.
15. Vatahska, Teodora, Maren Bennewitz, and Sven Behnke. "Feature-based head pose estimation from images." Humanoid Robots, 2007 7th IEEERAS International Conference. IEEE, 2007.
16. Hinton, Geoffrey, et al. "Deep neural networks for acoustic modelling in speech recognition: The shared views of four research groups." IEEE Signal Processing Magazine 29.6 (2012): 82-97.
17. Benfold, Ben, Ian Reid. "Unsupervised learning of a scene-specific coarse gaze estimator." International Conference on Computer Vision. IEEE, 2011.
18. Gourier, Nicolas, et al. "Head pose estimation on low resolution images." International Evaluation Workshop on Classification of Events, Activities and Relationships. Springer Berlin Heidelberg, 2006.
19. Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).
20. Alexandre, Lus A. "3D object recognition using convolutional neural networks with transfer learning between input channels." Intelligent Autonomous Systems 13. Springer International Publishing, 20, 889-898.
21. Gupta, Saurabh, et al. "Learning rich features from RGB-D images for object detection and segmentation." European Conference on Computer Vision. Springer International Publishing, 2014.
22. Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.
23. Faria, Diego R., Cristiano Premebida, and Urbano Nunes. "A probabilistic approach for human everyday activities recognition using body motion from RGB-D images." Robot and Human Interactive Communication, 2014 RO-MAN: The 23rd IEEE International Symposium on. IEEE, 2014.
24. Zhou, Xingxing, et al. "Detection of pathological brain in MRI scanning based on wavelet-entropy and naive Bayes classifier." International Conference on Bioinformatics and Biomedical Engineering. Springer International publishing, 2015.
25. Xhemali, Daniela, Chris J. Hinde, and Roger G. Stone. "Naive Bayes vs. decision trees vs. neural networks in the classification of training web pages." (2009).

