Grenze International Journal of Engineering and Technology, Special Issue

Occupancy Tracking and Direction Detection using Machine Learning and Image Processing

Dr.S.Kalarani[1], Krithika M.S[2] and Sneha[3]
[1-3] St Joseph's Institute of Technology /Information Technology, OMR, Chennai, India
Email: kala_subramanium@yahoo.co.in

***Abstract*— Facing direction detection plays a critical role in human computer interaction (HCI), and has received significant attention in recent research, such as surveillance systems, computer games, driving awareness recognition, assistive technologies and suspicious behavior detection. Currently the most popular approach for facing direction detection is optical-camera based facial feature extraction, such as such as skin color or eye gaze. The associated identification algorithms include a static head-pose estimation algorithm and a visual 3-D tracking algorithm for driver awareness monitoring. However, optical-camera based approaches raise concerns on privacy invasion and these devices do not usually work in a dark environment. And thus they are not an ideal option for many applications, such as rapidly increased smart home applications. With the popularity of smart home devices, unobtrusive HCI solutions are becoming more and more desirable, such as the use of passive infrared (PIR), and thermopile array sensors. PIR sensor is widely used as motion detector for automatic lighting and HVAC control, it has also been extended for occupancy moving direction detection.**

***Index Terms*— human computer interaction; visual 3-D tracking; style; passive infrared.**

I. INTRODUCTION

Facing direction detection plays a critical role in human computer interaction and has a wide application in surveillance systems, driving awareness recognition, smart home appliances, computer games etc. Current detection methods are mainly focused on extracting specific patterns from user's optical images, which raises concerns on privacy invasion and these detection techniques do not usually work in a dark environment. Also two feature extraction methods are compared. One is manually-defined and the other is based on a pre-trained artificial neural network (ANN) model. The facing direction detection accuracy resulting from manually-defined features reaches 85.3%, 90.6% and 85.2% at detection distance of 0.6m, 1.2m and 1.8 m, respectively. The level of accuracy resulted from using pre-trained ANN features demonstrates a much more reliable performance (89.1%, 95.3% and 95.1% at distance of 0.6m, 1.2m and 1.8m, respectively). An artificial neuron network (ANN) is a computational model based on the structure and functions of biological neural networks. Information that flows through the network affects the structure of the ANN because a neural network changes - or learns, in a sense - based on that input and output.

The face and non face image is identified using Viola jones framework. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. The problem to be

Grenze ID: 01.GIJET.6.2.3

solved is detection of faces in an image. A human can do this easily, but a computer needs precise instructions and constraints. To make the task more manageable, Viola–Jones requires full view frontal upright faces. Thus in order to be detected, the entire face must point towards the camera and should not be tilted to either side. The features sought by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection.
The original goal of the ANN approach was to solve problems in the same way that a human brain would. However, over time, attention moved to performing specific tasks, leading to deviations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis. Once the classification is done using ANN next we have to enter the testing phase which involves Gabor feature extraction. The optical-camera based approaches raise concerns on privacy invasion and these devices do not usually work in a dark environment. And thus they are not an ideal option for many applications, such as rapidly increased smart home applications.

II. Motivation for the Work

- To develop a system that can identify the direction of the person from the thermal images even in the dark environment.
- It may seem very easy but in reality we have to consider many constraints like single face or multiple faces, image rotation, pose etc.
- To identify the facing direction without invading the privacy of the person.
- To develop a system that helps in the mining process to help with the lead.

III. Related Works

A method for detecting direction when interfacing with a computer program is provided. The method includes capturing an image presented in front of an image capture device. The image capture device has a capture location in a coordinate space. When a person is captured in the image, the method includes identifying a human head in the image and assigning the human head a head location in the coordinate space. The method also includes identifying an object held by the person in the image and assigning the object an object location in coordinate space. The method further includes identifying a relative position in coordinate space between the head location and the object location when viewed from the capture location. The relative position defines a pointing direction of the object when viewed by the image capture device. The method may be practiced on a computer system, such as one used in the gaming field.
The driver's face is key to a less intrusive method to monitoring the driver to derive information such as distraction, drowsiness, intent, and where they are looking. A vital step in extracting these higher level information is to find the driver's face and individual components such as eyes, nose and mouth, along with the direction they are facing towards. In the context of safety critical situation like driving, it is important that this vital step be robust otherwise the higher level information is not available or unreliable. The lighting condition of a driver's cabin varies greatly from dark due to driving under a bridge or in a parking structure to extremely bright on a sunny day. Various occlusions on the face may also occur due to hand activities such as drinking water or different gestures. This work introduces a system based on existing CNN structures with slight modifications to implement face detection, landmark localization, and landmark-based head pose estimation method which addresses the challenges found in the driver cabin.
Then a multi-class classification technique like support vector machine (SVM) is used to train the machine learning (ML) model in order to identify the best threshold level that separates trust- worthy interactions from others. In this research, our main objective is to differentiate malicious interactions from trustworthy interactions with maximum boundary separation and minimum outliers rather than classification itself. Therefore, it is not necessary to go for other algorithms like Random Forest, especially with a low dimensional dataset compared to its sample size used in this work. However, depending on the data set, dimensionality, number of classifications required and noise levels of the samples, a model comparison can be performed to find out the best possible algorithm for each individual case. A well-trained model like this can differentiate an incoming interaction between two or more objects much more efficiently than linear weightage methods and is much more beneficial in the decision making process.

Pedestrian Detection is a critical technique for avoiding the collision between the vehicle and people, and it can be used in the advanced driver assistance system. Most research of the pedestrian detection areas are focused on the standing or walking people at the training process. INRIA's pedestrian dataset is composed of persons standing and facing the front, however another datasets comprise various types of pedestrian without classification for direction. In other words, movement directions of the pedestrian are not considered on creating detectors. In this paper, we propose a pedestrian detection method using pedestrian data classified into four by moving directions such as front, back, left and right.

Each of detectors created by categorized data are integrated, which are used for pedestrian detection. For the training, we use histograms of oriented gradients using the direction distribution of the edges. In the experiments, we use the pedestrian datasets obtained by moving vehicle in order to enhance public confidence. Our result shows the improved detection ratio in comparison to existing methods under utilised the moving direction.

Safety has, for a long time, been one big thing everyone is concerned about. Security breach of private locations has become a threat that everyone intends to eliminate. The traditional security systems trigger alarms when they detect a security breach. However, the usage of image processing coupled with deep learning using convolution neural networks for image identification and classification helps in identifying a breach in an enhanced fashion thereby increasing security furthermore to a great extent. This is due to its capability to extract complex features from the images using accurate and advanced face and body detection algorithms.

The rate at which machine learning, especially deep learning, is transitioning is very high. The use of such technology in taking the existing systems and models to the next level would be a great step towards advancements in every field of science and technology.

The same goes with computer vision. These two coupled and brought together to be used in the field of security results in achieving a lot more than what is imagined to be possible and this paper aims to do the same.

IV. System Architecture

A. Design Methods

There are two techniques involved. One is manually defined and the other is the Artificial neural network and the accuracy varies according to that. Artificial neural networks (ANN) or connectionist systems are computing systems inspired by the biological neural networks that constitute animal brains. The neural network itself is not an algorithm, but rather a framework for many different machine learning algorithms to work together and process complex data inputs. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain.

Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between artificial neurons are a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs. The connections between artificial neurons are called 'edges'. Artificial neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or decreases specific tasks, leading to deviations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis. Once the classification is done using ANN next we have to enter the testing phase which involves Gabor feature extraction.

In image processing, a Gabor filter, named after Dennis Gabor, is a linear filter used for texture analysis, which means that it basically analyses whether there are any specific frequency content in the image in specific directions in a localized region around the point or region of analysis. Frequency and orientation representations of Gabor filters are claimed by many contemporary vision scientists to be similar to those of the human visual system, though there is no empirical evidence and no functional rationale to support the idea. They have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

Disadvantages

- Hardware dependence: Artificial neural networks require processors with parallel processing power, in accordance with their structure. For this reason, the realization of the equipment is dependent.
- Unexplained behavior of the network: This is the most important problem of ANN. When ANN produces a probing solution, it does not give a clue as to why and how. This reduces trust in the network.
- Determination of proper network structure: There is no specific rule for determining the structure of artificial neural networks. Appropriate network structure is achieved through experience and trial and error.
- Directly influence the performance of the network. This depends on the user's ability.

V. ARCHITECTURAL DESIGN

A. Direction Detection

The first step is to create the database. The database includes both face and non-face images. To identify the difference between them Viola-jones object framework is used.

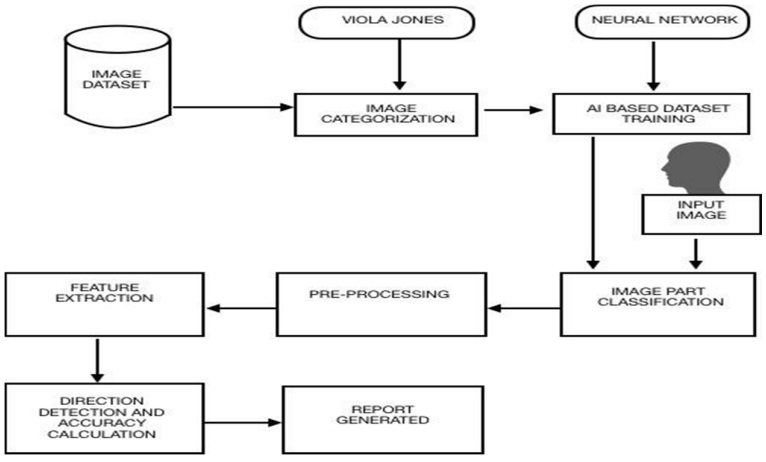
- Difficulty of showing the problem to the network: ANNs can work with numerical information. Problems have to be translated into numerical values before being introduced to ANN. The display mechanism to be determined here will directly influence the performance of the network. This depends on the user's ability.

The Viola–Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed in 2001 by Paul Viola and Michael Jones. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. The problem to be solved is detection of faces in an image. A human can do this easily, but a computer needs precise instructions and constraints. To make the task more manageable, Viola–Jones requires full view frontal upright faces. Thus in order to be detected, the entire face must point towards the camera and should not be tilted to either side. While it seems these constraints could diminish the algorithm's utility somewhat, because the detection step is most often followed by a recognition step, in practice these limits on pose are quite acceptable.

The algorithm has four stages:

1. Haar Feature Selection
2. Creating an Integral Image
3. Adaboost Training
4. Cascading Classifiers

The features sought by the detection framework universally involve the sums of image pixels within rectangular areas. As such, they bear some resemblance to Haar basis functions, which have been used previously in the realm of image-based object detection. However, since the features used by Viola and Jones all rely on more than one rectangular area, they are generally more complex.



System Architecture

The above figure is the system Architecture for identifying the facing direction of the person using Machine learning and image processing. This is the detailed explanation of the series of steps performed to identify the direction.

VI. System Modules

A. Categorizing face and non-face

The database includes all the images both face and non-face. The difference between them is identified using Viola jones object framework. The Viola–Jones object detection framework is the first object detection framework to provide competitive object detection rates in real-time proposed by Paul Viola and Michael Jones. Although it can be trained to detect a variety of object classes, it was motivated primarily by the problem of face detection. The problem to be solved is detection of faces in an image. A human can do this easily, but a computer needs precise instructions and constraints. This framework use Haar features to identify face and non-face. Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector.

Historically, working with only image intensities (i.e., the RGB pixel values at each and every pixel of image) made the task of feature calculation computationally expensive. A publication discussed working with an alternate feature set based on Haar wavelets instead of the usual image intensities. Viola and Jones adapted the idea of using Haar wavelets and developed the so-called Haar-like features. A Haar-like feature considers adjacent rectangular regions at a specific location in a detection window, sums up the pixel intensities in each region and calculates the difference between these sums. This difference is then used to categorize subsections of an image. For example, let us say we have an image database with human faces. It is a common observation that among all faces the region of the eyes is darker than the region of the cheeks. Therefore a common Haar feature for face detection is a set of two adjacent rectangles that lie above the eye and the cheek region. The position of these rectangles is defined relative to a detection window that acts like a bounding box to the target object (the face in this case).

In the detection phase of the Viola–Jones object detection framework, a window of the target size is moved over the input image, and for each subsection of the image the Haar-like feature is calculated. This difference is then compared to a learned threshold that separates non-objects from objects. Because such a Haar-like feature is only a weak learner or classifier (its detection quality is slightly better than random guessing) a large number of Haar-like features are necessary to describe an object with sufficient accuracy. In the Viola–Jones object detection framework, the Haar-like features are therefore organized in something called a classifier cascade to form a strong learner or classifier.

The key advantage of a Haar-like feature over most other features is its calculation speed. Due to the use of integral images, a Haar-like feature of any size can be calculated in constant time (approximately 60 microprocessor instructions for a 2-rectangle feature).

B. Training the database using ANN

After loading the face and non-face images, we need to train the database. There are two methods, one is manually defined and the other one is ANN. Better performance is achieved by using ANN. An ANN is based on a collection of connected units or nodes called artificial neurons, which loosely model the neurons in a biological brain.

Each connection, like the synapses in a biological brain, can transmit a signal from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it.

The learning algorithm can be divided into two phases: propagation and weight update.

Phase 1: propagation

Each propagation involves the following steps:

1. Propagation forward through the network to generate the output value(s)
2. Calculation of the cost (error term)
3. Propagation of the output activations back through the network using the training pattern target to generate the deltas (the difference between the targeted and actual output values) of all output and hidden neurons.

Phase 2: weight update

For each weight, the following steps must be followed:

1. The weight's output delta and input activation are multiplied to find the gradient of the weight.

2. A ratio (percentage) of the weight's gradient is subtracted from the weight.

This ratio (percentage) influences the speed and quality of learning; it is called the learning rate. The greater the ratio, the faster the neuron trains, but the lower the ratio, the more accurate the training is. The sign of the gradient of a weight indicates whether the error varies directly with, or inversely to, the weight. Therefore, the weight must be updated in the opposite direction, "descending" the gradient. Learning is repeated (on new batches) until the network performs adequately.

C. Face Detection

Detection Preprocessing

Pre processing functions involve those operations that are normally required prior to the main data analysis and extraction of information, and are generally grouped as radiometric or geometric corrections. Some standard correction procedures may be carried out in the ground station before the data is delivered to the user. These procedures include radiometric correction to correct for uneven sensor response over the whole image and geometric correction to correct for geometric distortion due to Earth's rotation and other imaging conditions (such as oblique viewing).

Image Enhancement

Image enhancement is conversion of the original imagery to a better understandable level in spectral quality for feature extraction or image interpretation. It is useful to examine the image Histograms before performing any image enhancement.

The x-axis of the histogram is the range of the available digital numbers, i.e. 0 to 255. The y-axis is the number of pixels in the image having a given digital number. Examples of enhancement functions include:

- Contrast stretching to increase the tonal distinction between various features in a scene. The most common types of enhancement are: a linear contrast stretch, a linear contrast stretch with saturation a histogram-equalized stretch.
- Filtering is commonly used to restore imagery by avoiding noises to enhance the imagery for better interpretation and to extract features such as edges and lineaments. The most common types of filters: mean, median, low-, high pass.

Haar Feature Extraction

Haar-like features are digital image features used in object recognition. They owe their name to their intuitive similarity with Haar wavelets and were used in the first real-time face detector.

A simple rectangular Haar-like feature can be defined as the difference of the sum of pixels of areas inside the rectangle, which can be at any position and scale within the original image. This modified feature set is called 2-rectangle feature. Viola and Jones also defined 3-rectangle features and 4-rectangle features. The values indicate certain characteristics of a particular area of the image. Each feature type can indicate the existence (or absence) of certain characteristics in the image, such as edges or changes in texture. For example, a 2-rectangle feature can indicate where the border lies between a dark region and a light region.

Histogram Equalisation

Histogram equalization is used to enhance contrast. It is not necessary that contrast will always be increase in this. There may be some cases were histogram equalization can be worse. In that cases the contrast is decreased. Finally they obtained image is mapped with the training data and then the direction is identified.

First we have to calculate the PMF (probability mass function) of all the pixels in this image. If you do not know how to calculate PMF, please visit our tutorial of PMF calculation.

Our next step involves calculation of CDF (cumulative distributive function). Again if you do not know how to calculate CDF, please visit our tutorial of CDF calculation.

There is also one important thing to be note here that during histogram equalization the overall shape of the histogram changes, where as in histogram stretching the overall shape of histogram remains same.

Performance Evaluation

Accuracy is determined using the concept of confusion matrix. In the field of machine learning and specifically the problem of statistical classification, a confusion matrix, also known as an error matrix, is a specific table layout that allows visualization of the performance of an algorithm, typically a supervised learning one (in unsupervised learning it is usually called a matching matrix).

Each row of the matrix represents the instances in a predicted class while each column represents the instances in an actual class (or vice versa). The name stems from the fact that it makes it easy to see if the system is confusing two classes (i.e. commonly mislabelling one as another).
It is a special kind of contingency table, with two dimensions ("actual" and "predicted"), and identical sets of "classes" in both dimensions (each combination of dimension and class is a variable in the contingency table).A confusion matrix is a summary of prediction results on a classification problem. The number of correct and incorrect predictions are summarized with count values and broken down by each class. This is the key to the confusion matrix.
The confusion matrix shows the ways in which your classification model is confused when it makes predictions. It gives us insight not only into the errors being made by a classifier but more importantly the types of errors that are being made.
Below is the process for calculating a confusion Matrix.

1. You need a test dataset or a validation dataset with expected outcome values.
2. Make a prediction for each row in your test dataset.

From the expected outcomes and predictions count: The number of correct predictions for each class. The number of incorrect predictions for each class, organized by the class that was predicted.

VII. Conclusion

This paper proposes a way to identify the direction of the person from the thermal images. The main advantage is that works even under the dark environment and it doesn't invades the privacy of an individual. Since the analysis is done using the machine learning languages accurate values can be obtained using the artificial neural network (ANN) algorithm. Our main contribution includes an algorithm/technique for abstracting the features and mapping with the training data and identifying the direction and the accuracy is achieved using confusion matrix

Future Enhancement

In the future work, plans are made to further improve the analyzing methods and to make a real time project instead of loading from the database with less processing time and to reduce space and time complexity. Based on the complete survey and analysis some of the techniques work better while some have issues, concerns and problems, according to that system can be designed.

References

[1] Weiss, Gary M., and Foster Provost. "Learning when training data are costly: the effect of class distribution on tree induction." Journal of Artificial Intelligence Research (2003): 315–354.
[2] Turney, Peter. "Types of cost in inductive concept learning." (2000).
[3] Abney, Steven. Semisupervised learning for computational linguistics. CRC Press, 2007.
[4] Žliobaitė, Indrė, et al. "Active learning with evolving streaming data." Machine Learning and Knowledge Discovery in Databases. Springer Berlin Heidelberg, 2011. 597–612.
[5] S. S. Mukherjee and N. M. Robertson, "Deep head pose: Gaze-direction estimation in multimodal video," IEEE rans. Multimedia, vol. 17, no. 11, pp. 2094-2107, Nov. 2015.
[6] R. Arroyo, J. J. Yebes, L. M. Bergasa, I. G. Daza, and J. Almazán, "Expert video-surveillance system for real-time detection of suspicious behaviors in shopping malls," Expert syst. App., vol. 42, no. 21, pp. 7991-8005, Nov. 2015.
[7] Chamveha, Y. Sugano, D. Sugimura, T. Siriteerakul, T. Okabe, Y. Sato, and A. Sugimoto, "Head direction estimation from low resolution images with scene adaptation," Comput. Vis. Image. Underst. vol. 117, no. 10, pp. 1502-1511, Oct. 2013.
[8] R. Dhumane, J. Ling, V. Aute, and R. Radermacher, "Portable personal conditioning systems: Transient modeling and system analysis," Applied Energy, vol. 208, pp. 390-401, Dec. 2017.
[9] S. Parnin and M. M. Rahman, "Human Location Detection System Using Micro-Electromechanical Sensor for Intelligent Fan", IOP Conf. Series:Materials Sci. Eng., vol. 184, no. 1, p. 012042, 2017.
[10] J. Lu, T. Zhang, F. Hu, and Q. Hao, "Preprocessing Design in Pyroelectric Infrared Sensor-Based Human-Tracking System: On Sensor Selection and Calibration," IEEE Trans. on Syst., Man, and Cybern. Syst,, vol. 47, pp. 263-275, Feb. 2017.