Journal of Network Communications and Emerging Technologies (JNCET) www.jncet.org
Volume 8, Issue 4, April (2018)

Driver Emotion Recognition and Assistance Based on Facial Analysis and Machine Learning Approach

Asha.R [1], MD Samse Alam [2], MD Tibul Rain [3], Pradeep Sati [4], Saikatt Ghose [5]
[1] Assistant Professor (O.G), Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.
[2, 3, 4, 5] Department of Computer Science and Engineering, SRM Institute of Science and Technology, Chennai, Tamil Nadu, India.

Abstract – Aspiration to provide better safety and comfort to car drivers has greatly influenced research in the field of Advanced Driver Assistance System. Though there are several systems which alert driver of a possible undesirable event (possibility of collision with another vehicle or persons, a steep turn ahead, exceeding speed limit etc.), fatal accidents are reported daily. Several intrusive and non-intrusive techniques have been proposed in the past to monitor car driver's emotions but very little light was shed on using cameras for such applications. This project details with one such system that uses a single camera. The paper also details face detection techniques prior to emotion recognition.

1. INTRODUCTION

Facial recognition (or face recognition) is a type of biometric software application that can identify a specific individual in a digital image by analysing and comparing patterns. The Kinetic motion gaming system, for example, uses facial recognition to differentiate among players. Most current facial recognition systems work with numeric codes called face prints. Such systems identify 80 nodal points on a human face. In this context, nodal points are end points used to measure variables of a person's face, such as the length or width of the nose, the depth of the eye sockets and the shape of the cheekbones. These systems work by capturing data for nodal points on a digital image of an individual's face and storing the resulting data as a face print. The face print can then be used as a basis for comparison with data captured from faces in an image or video. Facial recognition systems based on face prints can quickly and accurately identify target individuals when the conditions are favourable. However, if the subject's face is partially obscured or in profile rather than facing forward, or if the light is insufficient, the software is less reliable. Nevertheless, the technology is evolving quickly and there are several emerging approaches, such as 3D modelling, that may overcome current problems with the systems currently, a lot of facial recognition development is focused on Smartphone applications. Smartphone facial recognition capacities include image tagging and other social networking integration purposes as well as personalized marketing. A research team at Carnegie Mellon has developed a proof-of-concept iPhone app that can take a picture of an individual and -- within seconds -- return the individual's name, date of birth and social security number. Facebook uses facial recognition software to help automate user tagging in photographs. Here's how facial recognition works in Facebook: Each time an individual is tagged in a photograph, the software application stores information about that person's facial characteristics. When enough data has been collected about a person to identify them, the system uses that information to identify the same face in different photographs, and will subsequently suggest tagging those pictures with that person's name.

1.1 DIGITALIMAGE PROCESSING

The identification of objects in an image would probably start with image processing techniques such as noise removal, followed by (low-level) feature extraction to locate lines, regions and possibly areas with certain textures.

The clever bit is to interpret collections of these shapes as single objects, e.g. cars on a road, boxes on a conveyor belt or cancerous cells on a microscope slide. One reason this is an AI problem is that an object can appear very different when viewed from different angles or under different lighting. Another problem is deciding what features belong to what object and which are background or shadows etc. The human visual system performs these tasks mostly unconsciously but a computer requires skilful programming and lots of processing power to approach human performance. Manipulating data in the form of an image through several possible techniques. An image is usually interpreted as a two-dimensional array of brightness values, and is most familiarly represented by such patterns as those of a photographic print, slide, television screen, or movie screen. An image can be processed optically or digitally with computer. To digitally process an image, it is first necessary to reduce the image to a series of numbers that can be manipulated by the computer. Each number representing the brightness value of the image at a particular location is called a picture element, or pixel. A typical digitized image may have 512 × 512 or roughly 250,000 pixels, although much larger images are becoming common. Once the image has been digitized, there are three basic operations that can be performed on it in the computer. For a point operation, a pixel value in the output image depends on a single pixel value. problem is deciding what features belong to what object and which are background or shadows etc. The human visual system performs these tasks mostly unconsciously but a computer requires skilful programming and lots of processing power to approach human performance. Manipulating data in the form of an image through several possible techniques. An image is usually interpreted as a two-dimensional array of brightness values, and is most familiarly represented by such patterns as those of a photographic print, slide, television screen, or movie screen. An image can be processed optically or digitally with a computer. To digitally process an image, it is first necessary to reduce the image to a series of numbers that can be manipulated by the computer. Each number representing the brightness value of the image at a particular location is called a picture element, or pixel. A typical digitized image may have 512 × 512 or roughly 250,000 pixels, although much larger images are becoming common. Once the image has been digitized, there are three basic operations that can be performed on it in the computer. For a point operation, a pixel value in the output image depends on a single pixel value.

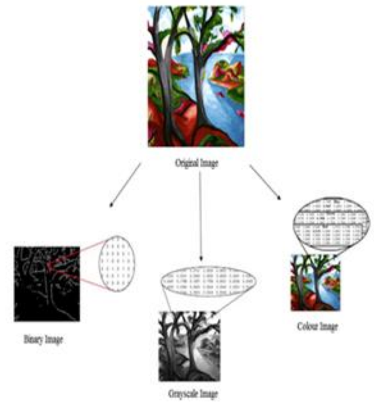
To digitally process an image, it is first necessary to reduce the image to a series of numbers that can be manipulated by the computer. Each number representing the brightness value of the image at a particular location is called a picture element, or pixel. A typical digitized image may have 512 × 512 or roughly 250,000 pixels, although much larger images are becoming common. Once the image has been digitized, there are three basic operations that can be performed on it in the computer. For a point operation, a pixel value in the output image depends on a single pixel value. problem is deciding what features belong to what object and which are background or shadows etc.

The human visual system performs these tasks mostly unconsciously but a computer requires skilful programming and lots of processing power to approach human performance. Manipulating data in the form of an image through several possible techniques. An image is usually interpreted as a two-dimensional array of brightness values, and is most familiarly represented by such patterns as those of a photographic print, slide, television screen, or movie screen. An image can be processed optically or digitally with a computer. To digitally process an image, it is first necessary to reduce the image to a series of numbers that can be manipulated by the computer. Each number representing the brightness value of the image at a particular location is called a picture element, or pixel. A typical digitized image may have 512 × 512 or roughly 250,000 pixels, although much larger images are becoming common. Once the image has been digitized, there are three basic operations that can be performed on it in the computer. For a point operation, a pixel value in the output image depends on a single pixel value.

1.2 CLASSIFICATIONOF IMAGES

There are 3 types of images used in Digital Image Processing. They are-

- Binary Image
- Gray Scale Image
- Color



BINARY IMAGE:

A binary image is a digital image that has only two possible values for each pixel. Typically the two colours used for a binary image are black and white though any two colours can be used. The colour used for the object(s) in the image is the foreground colour while the rest of the image is the background colour. Binary images are also called bi-level or two-level. This means that each pixel is stored as a single bit (0 or 1).This name black and white, monochrome or monochromatic are often used for this concept, but may also designate any images that have only one sample per pixel, such as rescale images. Binary images often arise in digital image processing as masks or as the result of Some input/output devices, such as laser printers, fax machines, and bi-level computer displays, can only handle bi-level images

GRAY SCALE IMAGE:

A gravy scale Image is digital image is an image in which the value of each pixel is a single sample. Images of this sort, also known as black-and-white, are composed exclusively of shades of gravy(0-255), varying from black (0) at the weakest intensity to white (255) at the strongest. Grayslake images are distinct from one-bit black-and-white images, which in the context of computer imaging are images with only the two colours, black, and white. Grayslake images have many shades of gravy in between. Grayslake images are also called monochromatic, denoting the absence of any chromatic variation. and in such cases they are monochromatic proper when only a given frequency is captured. But also they can be synthesized from a full colour image; see the section about converting to rescale.

COLOUR IMAGE:

A (digital) colour image is a digital image that includes colour information for each pixel. Each pixel has a particular value which determines it's appearing colour. This value is qualified by three numbers giving the decomposition of the colour in the three primary colours Red, Green and Blue. Any colour visible to human eye can be represented this way. The decomposition of a colour in the three primary colours is quantified by a number between 0 and 255. For example, white will be coded as R = 255, G = 255, B = 255; black will be known as (R,G,B) = (0,0,0); and say, bright pink will be : (255,0,255).

In other words, an image is an enormous two-dimensional array of colour values, pixels, each of them coded on 3 bytes, representing the three primary colours. This allows the image to contain a total of 256x256x256 = 16.8 million different colours. This technique is also known as RGB encoding, and is specifically adapted to human vision. It is observable that our behaviour and social interaction are greatly influenced by emotions of people whom we intend to interact with. Hence a successful emotion recognition system could have great impact in improving human computer interaction systems in such a way as to make them be more user-friendly and acting more human-like.

2. EXISTING APPROACH

GABOR FILTER APPROACH:

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.

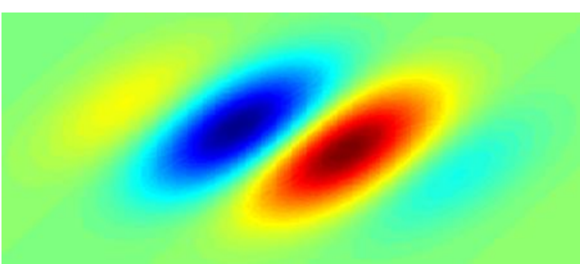


Fig: two-dimensional Gabor filter

They have been found to be particularly appropriate for texture representation and discrimination. In the spatial domain, a2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal wave. Some authors claim that simple cells in the visual cortex of mammalian brains can be modeled by Gabor functions. Thus, image analysis with Gabor filters is thought by some to be similar to perception in the human visual system.

Concept:

The low frequency sub bands of two source images are fused based on selection of appropriate coefficients using Gabor filtering. It is useful to discriminate and characterize the texture of an image through frequency and orientation representation. It uses the Gaussian kernel function modulated by sinusoidal wave to evaluate the filter coefficients for convolving with an image.The complex Gabor in space domain, here is the formula of a complex Gabor function in space domain g(x, y) = s(x, y) war(x, y) where s(x; y) is a complex sinusoidal, known as the carrier, and war(x; y) is a 2-D Gaussian-shaped function, known as the envelop. The complex sinusoidal is denotes as follows,

S(x, y) = expel (j (2*pi (u0 x + v0 y) + P))

Where (u0, v0) and P denotes the spatial frequency and the phase of the sinusoidal respectively.

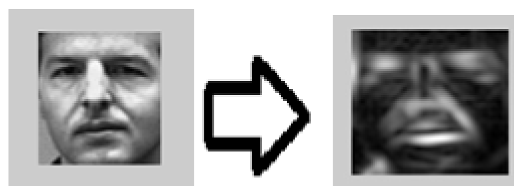


Fig: Gabor Filter Output

System Architecture:

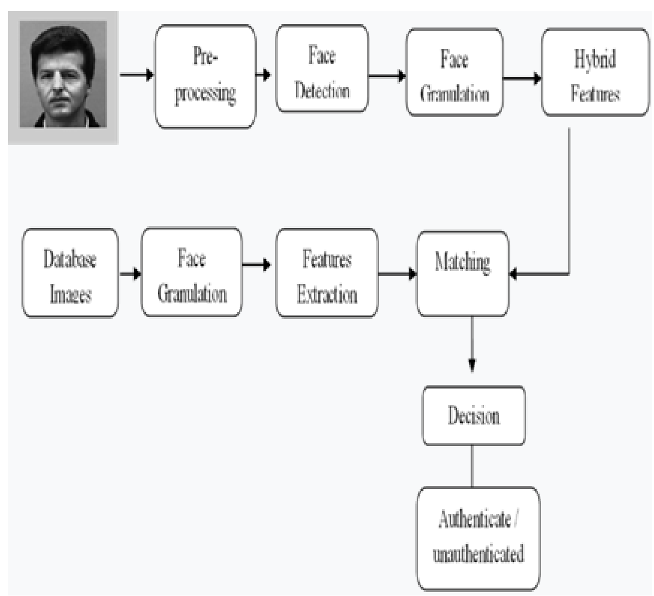


Fig 1.Block diagram of facial recognition based Person Re-Identification

3. FUNCTIONAL MODULES

Face Detection

It is process to extract face regions from input image which has normalized intensity and uniform in size. The appearance features are extracted from detected face part which describes changes of face such as furrows and wrinkles (skin texture). In this system model, an executable (.dell- dynamic link library) file is utilized to extract face region. It is used for face detection process is based on hair like features and adaptive boosting method

Weber local descriptor:

Weber local descriptor represents an image as a histogram of differential excitations and gradient orientations, and has several interesting properties like robustness to noise and illumination changes, elegant detection of edges and powerful image representation. WLD descriptor is based on Weber's Law. According to this law the ratio of the increment threshold to the background intensity is constant.

Weber Law:

According to Weber law, the ratio of the increment threshold to the background intensity is a constant. This relationship can be expressed as,

$$\frac{\Delta I}{I} = K$$

Where ∆I represents the increment threshold (just noticeable difference for discrimination); I represents the initial stimulus intensity and k signifies that the proportion on the left side of the equation remains constant despite variations in the Item. The fraction ∆I/I are known as the Weber fraction. Weber's Law, more simply stated, explains that the size of a just noticeable difference (i.e., ∆I) is a constant proportion of the original stimulus value. The process of extracting differential excitation and gradient.

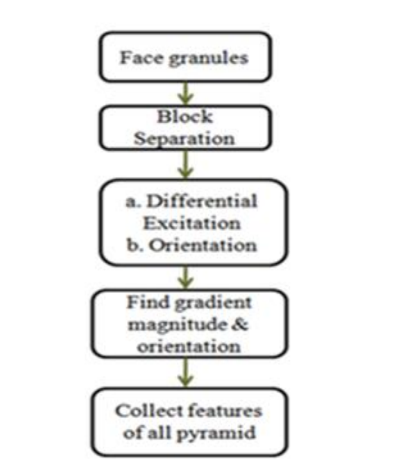


Fig: WLD Process Flow

4. STATISTICAL MEASUREMENTS

Texture is the innate property of all surfaces that describes visual patterns, each having properties of homogeneity. It also describes the relationship of the surface to the surrounding environment. In short, it is a feature that describes the distinctive physical composition of a surface.

Texture properties include:

- Coarseness
- Contrast
- Directionality
- Line-likeness
- Regularity
- Roughness

Texture is one of the most important defining features of an image. It is characterized by the spatial distribution of gravy levels in a neighbourhood. In order to capture the spatial dependence of gravy-level values, which contribute to the perception of texture, a two-dimensional dependence texture analysis matrix is taken into consideration. This two-dimensional matrix is obtained by decoding the image file.

Methods of Representation

There are three principal approaches used to describe texture; statistical, structural and spectral.

1. Statistical techniques characterize textures using the statistical properties of the grey levels of the points/pixels comprising a surface image. Typically, these properties are computed using: the grey level co-occurrence matrix of the surface, or the wavelet transformation of the surface.

2. Structural techniques characterize textures as being composed of simple primitive structures called "taxes" (or texture elements). These are arranged regularly on a surface according to some surface arrangement rules.

3. Spectral techniques are based on properties of the Fourier spectrum and describe global periodicity of the grey levels of a surface by identifying high-energy peaks in the Fourier spectrum.

At first the co-occurrence matrix is constructed, based on the orientation and distance between image pixels then meaningful statistics are extracted from the matrix as the texture representation. Texture features are:

1. Energy

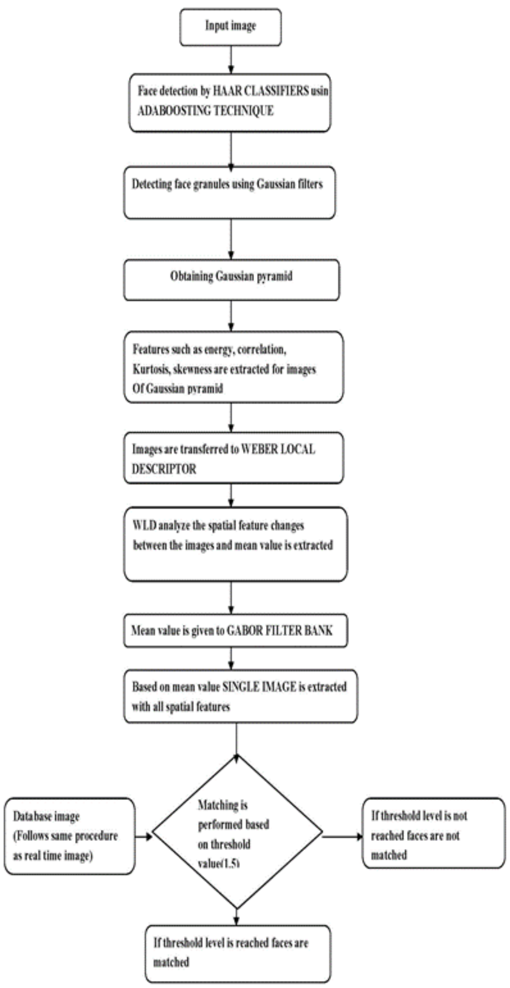
2. Contrast

3. Correlation

4. Homogeneity

5. Entrop

5. FLOW DIAGRAM OF THE ENTIRE PROCESS



The query image features will be matched with database image features for the person verification using Euclidean distance metric. It is defined by Ed= sort(sum(Q-Di).^2).

Where, Q- Input image features, D- Data feature base, i- No. of samples in database 1 to N

6. CONCLUSION

Face recognition systems used today work very well under constrained conditions, although all systems work much better with frontal mug-shot images and constant lighting. All current face recognition algorithms fail under the vastly varying conditions under which humans need to and are able to identify other people. Next generation person recognition systems will need to recognize people in real-time and in much less constrained situations. Identification systems are robust in natural environments, in the presence of noise and illumination changes, cannot rely on a single modality, so that fusion with other modalities is essential. Technology used in smart environments has to be unobtrusive and allow users to act freely. Wearable systems in particular require their sensing technology to be small, low powered and easily integral with the user's clothing.

REFERENCE

[1] JulienMeynet,(16th July 2003)'Fast Face Detection Using AdaBoost'.
[2] Moghaddam.B, Wahid.W and Pentland.A,(1998) 'Beyond eigenfaces:
[3] Probabilistic matching for face recognition', Proceeding of face and gesture recognition.
[4] Recognition as an Authentication Technique in Electronic Voting'.
[5] Swets.D.L andWeng.J.J,(1996) 'Using discriminate eigenfeatures for image retrieval', IEEE Trans. PAMI.
[6] TimoAhonen, AbdenourHadid, and MattiPietikainen, (2006)'Face Description with Local Binary Patterns: Application to Face Recognition.
[7] C. Busso, Z. Deng, S. Yildirim, M. Bulut, C. M. Lee, A.
[8] Kazemzadeh, S. Lee, U. Neumann, and S. Narayanan, "Analysis of emotion recognition using facial expressions, speech and multimodal information," Proc. 6th Int. Conf. Multimodal interfaces
[9] D. Singh, "Human emotion recognition system," Int. J. Image,Graph. Signal Process., vol. 4, no. 8, pp. 50–56, Aug. 2012.
[10] R. Yoshida, T. Nakayama, and T. Ogitsu, "Feasibility study on estimating visual attention using Electrodermal Activity," Proc. 8th Int. Conf. Sens. Technol., pp. 2–4, 2014.
[11] Y. Chen, "Stress state of driver: mobile phone use while driving,"Procedia - Soc. Behav. Sci., vol. 96, no. 86, pp. 12–16, Nov. 2013.
[12] S. Cohen and T. Wills, "Stress, social support, and the buffering hypothesis.," Psychol. Bull., 1985.
[13] B. Öz, T. Özkan, and T. Lajunen, "Professional and nonprofessional drivers' stress reactions and risky driving," Transp. Res. Part F Traffic Psychol. Behav., vol. 13, no. 1, pp. 32–40, Jan. 2010.