Enhancing Gray Scale Images for Face Detection under Unstable Lighting Condition

Mathias A. ONABID

Department of Mathematics and Computer Science,
Faculty of Sciences, PO Box 67 Dschang
University of Dschang, Cameroon

DJIMELI TSAMENE Charly

Department of Mathematics and Computer Science,
Faculty of Sciences, PO Box 67 Dschang,
University of Dschang, Cameroon

Abstract—**Facial expression plays a vital role in no verbal communication between human beings. The brain, in a quarter of second, can determine the state of mind and the behaviour of a person using different traits in a stable lighting environment. This is not the case in real applications such as online learning or driver monitoring system where lighting is not stable. It is therefore important to study and improve performance of some image enhancement techniques on face detection under varying lighting conditions in the spatial domain. The study is based on gray scale images. Nine gray scale standards based on colour space separating luminance to other colour components are used. The enhancement techniques compared are: the Global Histogram Equalisation (GHE), the Adaptive Histogram Equalisation (AHE) and Contrast Limited Adaptive Histogram Equalisation (CLAHE). Trials on the Labelled Face in the Wild (LFW) dataset using the Viola Jones Haar like features showed the CLAHE to outperform the GHE and AHE in face detection though the results appeared poor under low lighting condition. This motivated the need to stabilize lighting before applying Histogram Equalization techniques. The novelty in this research is that we have been able to apply the Gamma transform as a lighting stabiliser on the gray scale standard before enhancement. Comparing performance after lighting stabilisation showed AHE to be most appropriate for face detection, as it produced a detection rate of 99.31% and a relative high false positive rate (23.89 %).**

Keywords—Enhancement; AdaBoost; Haar like features; luma; peak signal to noise ratio (PSNR); Adaptive Histogram Equalisation (AHE); Contrast Limited Adaptive Histogram Equalisation (CLAHE); Global Histogram Equalisation (GHE); Gamma transform

I. INTRODUCTION

The face is one of the most important parts of the body, as it can be used to assess the state of mind, behaviour and even emotion of humans. Amongst the state of mind that can be assessed from facial expression are fatigue and somnolence. These aspects can be very useful when applied to driving control or online learning systems. The analysis of these aspects can only be possible after, face has been detected. Face detection under bright stable conditions possess no problem; but when lighting becomes unstable, the process of face detection becomes very challenging. This challenge is the motivation behind this research.

Lighting variation affects the perception of images and some of their features. The face in particular is highly affected since it has features which are light sensitive. It is therefore important to study image enhancement techniques and assess their impact on face detection using different features. This research is concerned with image enhancement in a spatial domain. Three enhancement techniques studied include, the Global Histogram equalisation (GHE), the Adaptive Histogram Equalisation (AHE) as developed by Pizer S M et al. [1] and the Contrast Limited Adaptive Histogram Equalisation (CLAHE) developed by Zuiderveld K. [2]. Most enhancement techniques are applied on gray scale images with different standards used to represent the images. To better assess their impact, nine standards which are mostly based on colour spaces which separate luminance to chrominance are used. In addition, many standards metrics used in assessing the impact of image enhancement use single images. This approach makes comparative study of dataset enhancement not feasible. In this research, an approach based on global peak signal to noise ratio (PSNR) is used.

A. Problem Statement

Most real time applications may require the detection of face in unstable or varying lighting conditions. This is a very crucial problem nowadays especially in the driving control and online learning systems where features of the face should be accurately detected in order to know the state of an individual. Commonly used enhancement techniques sometimes suffer from either over enhancement or under enhancement which oftentimes leads to reduced detection rate, due to the lighting variation. It is therefore important to stabilize lighting before applying enhancement.

B. General Objectives

The objective here is to attempt a solution to the problem of facial detection under lighting variation making use of some existing enhancement tools and then compare the impact of the enhancement on face detection. One approach in solving the unstable lighting problem is by pre-processing, which can be done by enhancing the image. The global Peak Signal to Noise Ratio (PSNR) is introduced to compare the effect of the enhancement methods on face detection.

Detection was done on the nearly frontal face and in a multi stage approach including: 1) conversion to gray scale, 2) apply an enhancement technique, 3) train the Haar like features using the Viola and Jones AdaBoost based method [3]. Assessment of the different approaches proposed was done using Labelled Face in the Wild (LFW) [4] dataset. To improve detection rate,

the luma standard is used as a lighting stabiliser and Histogram based approach are later applied, as illustrated in Fig. 3.

II. BACKGROUND AND RELATED WORK

Most of the facial detection systems start with image acquisition, followed by the extraction of the facial area before further processing. Given an arbitrary image, the goal of face detection is to determine whether or not there is any face in the image and if present, returns the location of each face found in the image. The challenges associated with image detection include: Pose, Presence or absence of structural component, Facial expression, Occlusion, Image orientation and Imaging condition.

In order to detect any face, it is necessary to choose the proper way to represent data so that it can increase the accuracy of the classification method. Song F et al. [5] stated that, given a data patch, several visual feature sets can be extracted from it. Each feature set is simply a transformation of a set of neighbouring raw pixel values, designed to be invariant to certain changes. Since no single feature descriptor will satisfy all the needs, selections for practice are mostly application-driven and the factors that need to be taken into account include: 1) the invariance properties it provides (e.g., invariance to lighting changes or to variations in scale, orientation, and other affine transformations); 2) the information encoded and the discriminability preserved; 3) the computational efficiency.

Hemalatha G and Sumathi C P. [6] proposed a classification of face detection techniques based on Knowledge based method, Facial invariant method, Template matching method and Appearance based method. A more simple but expressive classification approach has been given by Ce Zhan [7] where classification is on Knowledge based method and learning based method. In this, he states that the knowledge-based method attempts to depict our prior knowledge about the face pattern with some explicit rules, such as the intensity of faces, elliptic face contour, and equilateral triangle relation between eyes and mouth. Learning based method in the other hand, tries to model patterns with distribution or discriminant functions under a probabilistic framework. Some of these methodologies include Support Vector Machines (SVM), Artificial Neural Networks (ANN) and Bayesian-rule methods.

This learning based method is very effective with complex cases compared to knowledge based approach, however its drawback is that powerful computations are required due to learning model and large training data set with great variations. Recently this approach obtained better successes as the use of boosting has proven to be effective in reducing computational time. This is seen in Viola and Jones [3] where a real-time approach based on AdaBoost, was proposed. This has been the breakthrough of learning-based methodology.

A. Viola and Jones Face Detection Algorithm [3]

In Viola and Jones [3] approach, face detection is based on:

1) Feature representation: Haar like feature

Features used are reminiscent of Haar basis functions. Each feature is composed of a number of "black" and "white" rectangles joined together. After the approach of Viola and Jones succeeded, an extended set of Haar-like features were added to the basic feature set by Rainer Lienhart and Jochen Maydt [8]. This set contains more than 15 kinds of Haar like feature types.

2) Boosting

They used boosting as features classifier and selector, which converts a weak classifier to a strong one by combining a number N of weak classifier of $h_j(X)$ to form the strong classifier H(X) using the formula:

$$H(X) = \sum_{j=1}^{N} w_j h_j(X) \qquad (1)$$

Zhou Z H et al [9], proposed three modifications to the original Adaboost algorithm. These include: Gentle-, Logit-, and Real. The aim of boosting is to improve the classification performance while reducing the training duration.

3) Attentional Cascade

Viola and Jones [3] defines a weak classifier w(x) with features f, parity p which indicates the direction of the inequality sign and some threshold σ as follows:

$$w(x) = \begin{cases} 1 \text{ if } pf(x) < p\sigma \\ 0 \qquad \text{otherwise} \end{cases} \qquad (2)$$

Where, x is a sub-window of an image. This means that if the value of some feature exceeds some threshold, which has to be learned, the image is classified as positive.

They proposed the selection of a subset of relevant features which are sufficiently informative to model a face. The idea is that a very small subset of features can be combined to form an effective classifier. Adaboost classifier has been found to be appropriate for that purpose. Adaboost searches for a small number of good features classifier and then constructs a strong classifier as a linear combination of weighted simple weak classifiers as shown in Fig. 1.

In Fig.1, the structure reflects the fact that within any single image the majority of sub-windows are negative. As such, the cascade attempts to reject as many negatives as possible at the earliest stage.
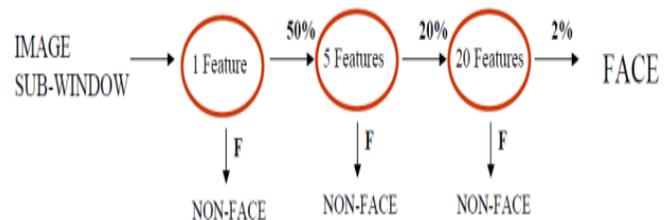

Fig. 1. Cascade structure of Duncan S. and Velthuis T [10].

Duncan S. and Velthuis T. [10] states that for each sub-window, the order of complexity is $O(FS_f)$, where, F is the number of Haar-like features used by the cascade and S_f is the number of pixels contained by each feature. Using the cascade most sub-windows can quickly be discarded and only True Positive and False Positives will run the whole cascade. The total number of operations is order $O(NFS_f)$, where N is the total amount of sub-windows and FS_f, is a function of the image size and number of scales included in the search.

B. Image Representation in Various Colour Space

Generally, a colour space can be seen as a mathematical representation of a set of colours of which the three most popular colour models are Red Green Blue (RGB) used in computer graphics; Luminance, Inphase Quadrature (YIQ), YUV, or Luminance Chominance blue and Chrominance red (YCbCr) used in video systems as stated by Huanzhao Zeng [11]. Other known colour models include the Hue Saturation Intensity (HSI) and Hue Saturation Value (HSV) which could be used to simplify programming, processing, and end-user manipulation. For colour adjustment, luminance is the major concern when compared to Saturation and Hue.

1) Numerical analysis of colour

Huanzhao Zeng [11] states that a 3-dimensional Euclidean space can better represent the human perception of colour since based on the cone analysis three numbers can be used to represent the three types of cones. Author [11] further states that each wavelength λ stimulates each of the three types of cone cells to a known extent, these extents may be represented by three functions, $s(\lambda)$, $m(\lambda)$, $l(\lambda)$ corresponding to the response of the S, M, and L cone cells, respectively. Since a beam of light can be composed of many different wavelengths, to determine the extent to which a physical light $C(\lambda)$ stimulates each cone cell, we must calculate the integral (with respect to λ) over the interval [λmin , λmax] as follows:

$$S = \int_{\lambda_{min}}^{\lambda_{max}} C(\lambda).s(\lambda)d\lambda \tag{3}$$

$$M = \int_{\lambda_{min}}^{\lambda_{max}} C(\lambda).m(\lambda)d\lambda \tag{4}$$

$$L = \int_{\lambda_{min}}^{\lambda_{max}} C(\lambda).l(\lambda)d\lambda \tag{5}$$

In order to simplify the problem, the CIE (Commission Internationale de l'Eclairage) defined a standard colour space CIEXYZ which uses a set of tristimulus values called X, Y, and Z. These X, Y, and Z values are used instead of the tristimulus values the S, M, and L responses of the human eye. In that regard the CIE has defined a set of three colour-matching functions, called $x(\hat{\lambda}), y(\hat{\lambda})$ and $z(\hat{\lambda})$, which can be thought of as the spectral sensitivity curves of three linear light detectors that yield the CIEXYZ tristimulus values X, Y, and Z as described in Huanzhao Zeng [11]. Furthermore, Huanzhao Zeng [11] states that some variations of CIEXYZ were latter deduce from the standard including CIELAB and CIELUV.

2) CIE uniform colour spaces (CIE L*a*b* , CIE L*u*v)

Huanzhao Zeng [11] also states that to improve the visual uniformity of colour, several colour spaces have been used, the most common being the CIELAB colour space which is illustrated as follow:

$$L^* = 116f\left(\frac{Y}{Y_n}\right) - 16 \tag{6}$$

$$a^* = 500\left[f\left(\frac{X}{X_n}\right) - f\left(\frac{Y}{Y_n}\right)\right] \tag{7}$$

$$b^* = 200\left[f\left(\frac{Y}{Y_n}\right) - f\left(\frac{Z}{Z_n}\right)\right] \tag{8}$$

Where,

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{for } t > \left(\frac{24}{116}\right) \\ \left(\frac{841}{108}\right)t + \frac{16}{116} & \text{otherwise} \end{cases} \tag{9}$$

XYZ are the tristimulus values of a colour and $X_n Y_n Z_n$ are the tristimulus values of the reference white being used. L^*, represents the lightness, a^* and b^* represent the chroma coordinates. The chroma is computed as $c = \sqrt{a^2 + b^2}$. An implementation of CIELAB is given by opencv whereby.

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.412453 & 0.357580 & 0.180423 \\ 0.212671 & 0.715160 & 0.072169 \\ 0.019334 & 0.119193 & 0.950227 \end{pmatrix} \text{ and }$$

$$L = \begin{cases} \frac{255}{100}\left(116Y^{\frac{1}{3}}\right) & \text{for } Y > 0.008856 \\ \frac{255}{100}(903.3Y) & \text{for } Y \le 0.008856 \end{cases} \tag{10}$$

$$a* = 500[f(X) - f(Y)] + \Delta + 128 \tag{11}$$

$$b^* = 200[f(Y) - f(Z)] + \Delta + 128 \tag{12}$$

Where, Δ takes 128 value for 8 bits value and 0 for floating point. The function f is defined by:

$$f(t) = \begin{cases} t^{\frac{1}{3}} & \text{for } Y > 0.008856 \\ 7.787t + \frac{16}{116} & \text{for } Y \le 0.008856 \end{cases} \tag{13}$$

3) HSV/ HSL

Hue-Saturation-Value (HSV) space is also a popular colour space as it is based on human colour perception. Similarly, HSI (I-Intensity) and HSL (L-Lightness) colour spaces are derived from HSV. The intensity, value and lightness are related to the colour luminance. Acharya Tand Ray A [12] stated that HSV provides absolute brightness information. Opencv proposed an implementation of HSV where the value is given as the maximum of the R,G,B component of the pixel. That is,

$$V = \max(R, G, B) \tag{14}$$

HSL is another Hue based colour space where the lightness value is the mean of the minimum and maximum RGB values given as

$$L = \frac{V + \min(R,G,B)}{2} \tag{15}$$

C. Gray Scale Conversion of Luminance Components

Luminance is the weighted sum of the RGB components of a colour image. The luminances are obtained using three defined constants C_R, C_G, and C_B define by (16) given as:

$$Y = C_R \times R + C_G \times G + C_B \times B \tag{16}$$

The C_R, C_G, and C_B values are given according to the International Telecommunication Union (ITU) standards illustrated in Table 1, also $C_R + C_G + C_B = 1$.

The ITU-R BT.601 [13] conversion rule is the widely used rule to convert the luminance components of the YIQ, YUV and YcbCr systems. In Table 2, a summary of some luminance based gray scale standard is given. R', G' and B' are obtain respectively from R, G and B using Gamma transform.

D. Enhancement Techniques

Bedi S.S. and Rati Khandelwal [16] explained that the aim of enhancing is to improve the visual appearance of the image for future automated processing, such as analysis, detection, segmentation and recognition and Anish Kumar Vishwakarma [17] did classified Image enhancement techniques into two broad categories: Spatial based domain image enhancement and Frequency based domain image enhancement. Author [16] sub classified spatial based category into point processing operation and spatial filter operations. In this paper, focus is on image enhancement technique in a spatial domain, especially histogram equalization based technique as described below.

1) The histogram equalization

Gonzalez R C and Woods R E [18] explains that the idea behind histogram equalization is to enhance an image such that its histogram will be flat, in other words a transformation of the form $Y = T(X)$ could be created to produce a new image Y with a flat histogram. That is for a digital gray image $X = \{x\}$ with gray levels in the range $[0, L-1]$, $Y = \{y\}$ the enhance image and let n_i be the number of occurrences of gray level *i*. Then the Probability Distribution Function (PDF) of the image which is the probability of an occurrence of a pixel of level *i* in the image can be computed as:

$$p_x(i) = p(x = i) = \frac{n_i}{n} \quad 0 \le i < L \tag{17}$$

TABLE I.   ITU STANDARD TO REPRESENT GRAY SCALE IMAGES

ITU Standard	C_R	C_G	C_B
ITU-R BT.601 conversion[13]	0.299	0.587	0.114
ITU-R BT.709 conversion[14]	0.2126	0.7152	0.0722
ITU-R BT.2020 conversion[15]	0.2627	0.678	0.0593

TABLE II.   GRAY SCALE STANDARD USED

Standard	Gray value
Luminance1	$\mathbf{0.299 \times R + 0.587 \times R + 0.114 \times B}$
Luminance2	$\mathbf{0.2126 \times R + 0.7152 \times G + 0.0722 \times B}$
Luminance3	$\mathbf{0.2627 \times R + 0.678 \times G + 0.0593 \times B}$
Luma1	$\mathbf{0.299 \times R' + 0.587 \times G' + 0.114 \times B'}$
Luma2	$\mathbf{0.2126 \times R' + 0.7152 \times G' + 0.0722 \times B'}$
Luma3	$\mathbf{0.2627 \times R' + 0.678 \times G' + 0.0593 \times B'}$
Luminance HSV	$\mathbf{Max(R, G, B)}$
Luminance HSL	$\mathbf{\dfrac{Max(R, G, B) + min(R, G, B)}{2}}$
Luminance CIELAB/CIELUV	Equation $(\mathbf{2.10})$

The Cumulative Distribution Function (CDF) can be computed as follows:

$$cdf_x(i) = \sum_{j=0}^{i} p_x(j) \tag{18}$$

And to create the mapping transformation one has to linearize the CDF across the value range, i.e.

$$cdf_y = iK \tag{19}$$

The histogram equalized image Y will be defined by,

$$Y(i, j) = (L-1) \sum_{n=0}^{X(i,j)} p(x = n) \tag{20}$$

The motivation for this transformation comes from thinking of the intensities of the continuous random variables X, Y on $[0, L-1]$ with Y defined by

$$Y = T(X) = (L-1) \int_0^X p_X(x)dx \tag{21}$$

T is the cumulative distributive function of X multiplied by $(L-1)$. Assume for simplicity that T is differentiable and invertible. It can then be shown that Y define by T(X) is uniformly distributed on $[0, L-1]$, namely,

$$p_Y(y) = \frac{1}{L-1} \tag{22}$$

If the histogram of any image has many peaks and valleys, it will have peaks and valleys after equalization but the peaks and valleys will be shifted. This technique improves contrast but the goal of histogram equalization is to obtain a uniform histogram. Anish Kumar Vishwakarma [17] defined three main types of Histogram Equalization, namely, Global Histogram Equalization (GHE), Adaptive Histogram Equalization (AHE) and Block-based Histogram Equalization (BHE).

- *Global Histogram Equalization (GHE)*:

According by Agaian et al. [19], the GHE attempts to alter the spatial histogram of an image to closely match a uniform distribution. It uses the same transformation derived from the image histogram to transform all pixels. This approach works well when the distribution of pixel values is similar throughout the image. However, when the image contains regions that are significantly lighter or darker than most of the image, the contrast in those regions will not be sufficiently enhanced. The drawbacks of GHE as identified by Bedi and Rati [16] include:

1) Problem of being poorly suited for retaining local detail due to its global treatment of the image

2) Over enhancing the image, resulting in an undesired loss of visual data, of quality and of intensity scale.

In fact, global histogram modification treats all regions of the image equally and, thus, often yields poor local performance in terms of detail preservation. This motivated the introduction of AHE.

- *Adaptive Histogram Equalization (AHE)*:

In an attempt to treat the aspect of details preservations, the AHE was built. The idea behind this technique as described by Pizer S M et al. [1] is to transform each pixel with a transformation function derived from a neighbourhood region. That is each pixel is transformed based on the histogram of a square surrounding it but the derivation of the transformation

functions from the histograms is exactly the same as for ordinary histogram equalization.

Arun R et al. [20] explains that though this method produces better result, it has the shortfall of over amplifying noise in relatively homogeneous regions of an image. Such that noise in dark regions could be over amplified until it becomes visible. The result is that often times, Block effect is observed in the output after sub-image processing. Thus the Contrast Limit AHE of Zuiderveld K [2] was developed as a means to address the problem to some extent.

- *Contrast Limited AHE (CLAHE)*:

Zuiderveld K [2] describes the CLAHE technique, to operate on small regions in the image, called *tiles*, rather than the entire image, and that each tile's contrast is enhanced, so that the histogram of the output region approximately matches the histogram specified by the CDF. The value, at which the histogram is clipped, known as clip limit, depends on the normalization of the histogram and hence on the size of the neighbourhood region. The part of the histogram that exceeds the clip limit is not discarded rather it is redistribute equally among all histogram bins as illustrated in Fig. 2.

Fig. 2(a) shows a CDF with a slope that can lead to amplification of tiles, so it is important to limit the slope and hence the of the transformation function. The redistribution of Fig. 2(b) will push some bins over the clip limit again resulting in an effective clip limit that is larger than the prescribed limit. The neighbouring tiles are then combined using bilinear interpolation to eliminate artificially induced boundaries. The contrast, especially in homogeneous areas, can be limited to avoid amplifying any noise that might be present in the image.
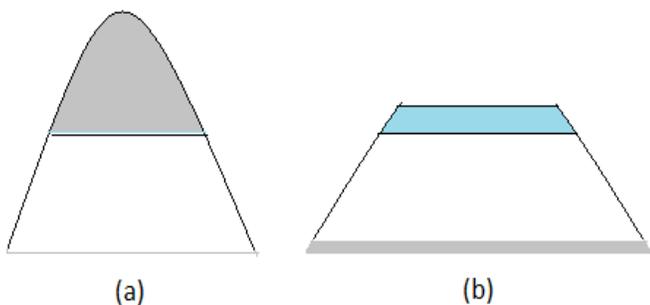

Fig. 2. Illustration of the clip limit effect.

E. Related Work

Many authors have tried to assess the performance of image enhancing techniques accuracy face detection. These include:

Li Tao [21], who proposed image enhancement technique to improve the visual quality of digital images that exhibit dark shadows due to the limited dynamic ranges of imaging and display devices which are incapable of handling high dynamic range scenes. The proposed technique processes images in two separate steps:

1) Dynamic range compression is able to enhance the luminance in dark shadows while keeping the overall tonality consistent with that of the input image, and

2) Local contrast enhancement which uses a neighbourhood dependent local contrast enhancement method to enhance images following the dynamic range compression.

To detect face they used a convolution face finder which is based on Convolutional Neural Network (CNN). These were tested on sample face images containing one up-right frontal human face in each image from the Phillips P J et al. [22] FRGC dataset. A comparative study showed their method to perform better that the GHE method.

Venkatesan M.E and Srinivasa Rao [23] proposed a Hybrid algorithm to detect faces using Ant Colony Optimization (ACO) and Genetic programming algorithms. The enhancement method used is the Histogram Equalization. The Ant colony is used to search the face area while the genetic programming is used to select and generate chromosome features used by ant colony. Though no real database was used for performance comparison, the author acknowledges that the method is time consuming compared to Viola and Jones [3].

Deepak Ghimire and Joonwhoan Lee [24], proposed a method to detect human faces in colour images in which three steps were involved:

1) Colour image enhancement is performed using a nonlinear transfer function that is based on a local approach.

2) Two separate processes are carried out and this included the luminance enhancement in which dynamic range compression is performed on the V channel image using a nonlinear transfer function, and a contrast enhancement in which each pixel in the image is further enhanced to adjust the image contrast, depending upon the centre pixel and its neighbourhood.

3) Finally, the original H and S component images and enhanced V component image are converted back to the RGB.

The method was tested on an unknown database and produces a detection rate of 85.96%.

Table 3 presents a global comparison of these different enhancement approaches, their features descriptor, data classification and some of the associated drawbacks. The belief is that if light is stabilized before applying these techniques, the detection accuracy will improve. This is the main concern of this research.

TABLE III.   COMPARISON OF RELATED WORK

Authors	Enhancement techniques	Descriptor	Classifier	Drawback
Viola Jones (2001)	Global Histogram equalization	Haar-like features	AdaBoost	Over enhancement
Li Tao (2006)	Dynamic range compression and local contrast enhancement	Feature map	Convolutional Neural network (CNN)	Two step processing and uses CNN which takes much time in training.
Venkatesan M.E. (2010)	Global Histogram equalization	Chromosome based features	Ant Colony Optimisation and Genetic Algorithm	Time consuming due to OCA classifier
Deepak Ghimire (2013)	Luminance + contrast enhancement in HSV	Skin Colour segmentation and Edges	Linear classifier	Enhancement done in more than one step with more than one colour.

III. METHODOLOGY

Phillips P J et al. [22] made it clear that the conventional methods like histogram equalization and Gamma transform generally have very limited performance due to the global processing scheme especially when lighting varies. It is therefore necessary to improve and stabilise lighting condition before enhancing. This is the main concern of this research as outlined below.

A. System Overview

In our work, Gamma transform is first applied to stabilize lighting, before other enhancing techniques are applied.

Fig. 3 gives a global view of our system. The Gamma transform is first apply to the RGB image to stabilise lighting variation; AHE is applied to enhanced the global system, though it still produces unnatural image it has the advantage of better highlighting facial features.

For implementation we are going to use, opencv platform with C++ language, the CLAHE implementation is the one given by opencv. The clip limit value for CLAHE is 0.02 and the Gamma transform coefficient of luma is 0.2.

We have train haar like classifier using the various gray scale standards.
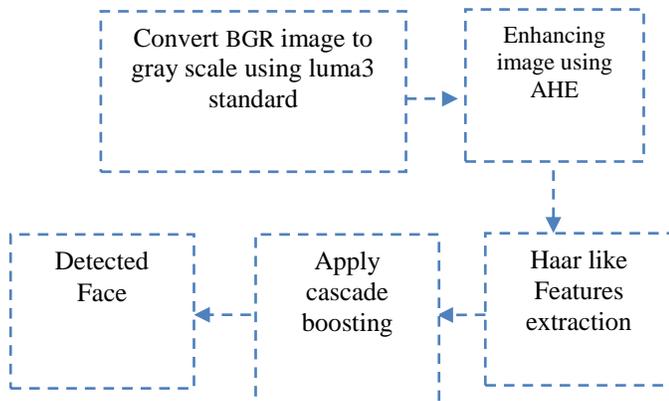
Fig. 3.   Luma based face detection system.

B. Training Stages

For training we used the following stages as describe by author [3] and implemented in the Opencv:

STEP 1: Collecting Image Database

We collected about 3000 positive images, that is, images containing at least one face from the internet and about 1000 negative images that is images without a face. Each set of images is placed in a folder.

STEP 2: Convert images to gray scale and apply enhancement

Positive and negative images were converted to gray scale using the 8 standards described in Section 2.3. After which three enhancements techniques are applied to these gray scale standards in separated patches.

STEP 3: Crop and Mark Positive Images

We need to create a data vector file that contains the names of positive images as well as the location of the objects in each image. The format of the vector file is of the form

(x and y represent the top left vertex):

image_path number_of_face x-coord y-coord width length

STEP 4: Creating a vector of positive images

In this stage we create a vector file using a .vec format, which is a data file containing face location.

STEP 5: Haar-Training and XML file

For training, the following characteristics: 20 stages, Haar like features and Gabor Adaptive Boosting are used. At the end of the training stage, a .XML file is obtained.

A comparison of this method is given in the next section.

C. Performance Criteria

A study will be done to compare image quality and face detection rate. For image quality two metrics will be used which are all based on:

- *MSE (Mean Square Error)*:

It is the cumulative squared error between the enhanced and the input images. Sonam Bharal [25] used it to assess image quality of some enhancement techniques. It is defined as follows:

$$\text{MSE} = \frac{1}{MN} \sum_{m=1}^{M} \sum_{n=1}^{N} (I_1(m, n) - I_2(m, n))^2 \qquad (23)$$

Where, I_1 and I_2 denote the original image and the improved image, respectively. The size of the input and output pictures should be same. If MSE is very large then it means that the quality of the image is very low.

- *Peak Signal to Noise Ratio (PSNR)*:

PSNR defines the ratio between an original image and a new one obtained by introducing error during compression. The PSNR can be obtained using the following formula:

$$\text{PSNR} = 10 \log_{10} \frac{(2^b - 1)}{\text{MSE}} \qquad (24)$$

Where, b is the number of bits per pixel (bpp) of the original image and MSE is the mean-square-error. Hitam et al. [26] showed that PSNR with high value indicates better quality of the image with less noise. They also declared that good methods are those that result to lower MSE and higher PSNR values.

To assess the relation between image quality and face detection, *Algorithm 1* is used; the algorithm uses the standard deviation of each image PSNR and compute a global PSNR of the image folder. Local PSNR is obtained using (24).

The next thing to asses is face detection, different metrics are used for that purpose. In this article we will evaluate the True Positive Rate (TPR) and the False Positive Rate (FPR). The evaluation uses the following matching metrics as defined by Richard Szeliski [27]:

True Positives (TP): Number of objects correct detected.

False Negatives (FN): Object that were not correctly detected, in other word matches identified as negative while they are supposed to be positive.

False Positives (FP): Proposed matches that are incorrect, in other word matches considered as positive while they are really negative.

True Negatives (TN): Non-matches that were correctly rejected. Based on the previous definition, we can define these numbers in terms of unit rates as follows:

Hit rate, sensitivity or True Positive Rate (TPR):

$$TPR = \frac{TP}{TP+FN} = \frac{TP}{P} \qquad (25)$$

Miss rate or False Positive Rate (FPR):

$$FPR = \frac{FP}{FP+TN} = \frac{FP}{N} \qquad (26)$$

Algorithm 1: Compute Image Global PSNR

Input: image folder
Output: global PSNR
1. *Initialization: global PSNR =0*
2. *For each image ∈ images folder*
   - *Read the image and Convert the RGB to gray scale image called **ImRef***
   - *apply an enhancement technique to **ImRef** to obtain **ImEnh***
   - *Process the **localPSNR** using **ImRef** and **ImEnh***
3. *endFor*
4. *Compute the standard deviation of all the **localPSNR** found in the folder.(**global PSNR=standard deviation of all the localPSNR)***
5. *return **global PSNR**,*

IV. RESULTS AND DISCUSSION

A. Experiments and Results

To carry out experiments the Labelled Face in the Wild (LFW) [4] database which shows poor image quality and great appearance datasets is used.

Using the set of 3000 positive images and 1000 negative images, data were trained following the stages given in the section *B*. The output of the training process gave a set of 48 .XML classifiers, which were later used for experiment.

TABLE IV. COMPARISON OF ENHANCEMENT TECHNIQUE WITH THE USE OF PSNR

Gray scale standard	Enhancement Method	LFW Dataset		
		TPR	FPR	PSNR
Luminance1	Not Enhanced	98.26	24.13	
	AHE	95.35	16.20	10.25
	CLAHE	**98.91**	**21.76**	**29.91**
	GHE	97.93	23.82	17.35
Luminance2	Not Enhanced	98.35	20.12	
	AHE	96.90	18.53	10.21
	CLAHE	**98.49**	**17.85**	**29.84**
	GHE	97.86	20.36	17.26
Luminance3	Not Enhanced	98.52	28.26	
	AHE	96.28	16.32	10.33
	CLAHE	**98.80**	**23.30**	**29.93**
	GHE	97.78	24.28	17.36
Luminance HSL	Not Enhanced	98.43	22.39	
	AHE	96.46	14.72	10.31
	CLAHE	**98.92**	**19.26**	**29.78**
	GHE	98.39	22.74	17.22
Luminance CIELAB/ CIELUV	Not Enhanced	98.31	29.92	
	AHE	96.79	21.79	10.32
	CLAHE	**98.79**	**21.60**	**30.14**
	GHE	98.15	20.46	17.96
Luminance HSV	Not Enhanced	98.27	29.38	
	AHE	94.27	19.03	10.60
	CLAHE	**98.51**	**24.74**	**31.54**
	GHE	97.84	26.47	19.01
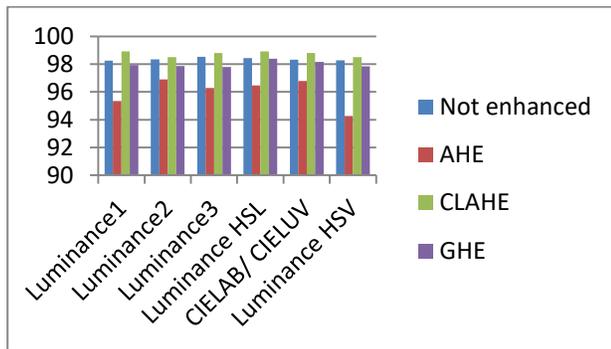
Fig. 4.   Comparison of enhancement techniques for face detection on LFW dataset.

TABLE V.   COMPARISON OF LUMA STANDARD

Gray scale standard	Enhancement Method	LFW Dataset		
		TPR	FPR	PSNR
Luma1	Not Enhanced	0	100	
	AHE	98.77	21.62	2.46
	CLAHE	0	100	0
	GHE	97.52	19.22	3.93
Luma2	Not Enhanced	0	100	
	AHE	99.16	25.40	2.41
	CLAHE	0	100	0
	GHE	97.78	27.16	3.95
Luma3	Not Enhanced	0	100	
	AHE	**99.31**	**23.89**	**2.41**
	CLAHE	0	100	0
	GHE	97.36	22.07	3.94

Fig. 5.   Luma3 Gray scale standard: (a) Not enhanced, (b) AHE, (c) CLAHE, (d) GHE.

To assess the result of the classifiers, LFW dataset is used and applied in each classifier. The result is given in Table 4 and Fig. 4. From the experiments carried out across the various gray scale standards, it was found that the higher the PSNR, the higher the accuracy of the detection method, using Haar like features.

Observation showed the CLAHE outperform the other enhancement techniques at face detection as seen in Fig. 4. The experiments also showed CLAHE to produce greater PSNR and higher detection rate. Yet for over lighted or under lighted condition it produces very low result as illustrated in Table 5.

Table 4 shows that CLAHE outperform in TPR, both AHE and GHE, also the PSNR of CLAHE is greater than the one used by AHE.

To stabilize lighting before enhancing we proposed a technique based on luma as illustrated in Fig. 3. In this technique, Gamma transform is first applied to the image before enhancing using the equalization techniques. The results showed that AHE applied on luminance3 standard performed better. Though it produced unnatural image it better highlighted the facial features than the other method.

B. Discussion

To justify the approach, other techniques are used to compare performance. Experimental trials showed the proposed method, based on Gamma transform and adaptive histogram equalisation performed better than the other techniques. The enhancement method used is a mixture of global enhancement (Gamma Transform) and local enhancement (AHE).

The global enhancement stabilised the light while the local enhancement let the system to easily discriminate the main features, that eyes, mouth, and nose can be easily extracted using rectangle features. Also, the AHE by performing a local enhancement with no clip limit makes the image to appear, though unnatural but with visible features.

The PSNR result shows CLAHE produces better quality image than any other image, yet the quality of image has little impact in very low lighting condition and feature based detection method as seen in Fig. 5. That is why, though the proposed method produces a very low global PNSR (2.41dB) it has a better detection. This is due to the unnatural image given by AHE, which in contrary has better result (99.31% of TPR).

V.   CONCLUSION AND FUTURE WORK

In this article the aim was to study the environment of facial feature detection under unstable lighting condition, we have presented a framework for real time detection of faces. The face detection is based on Viola Jones algorithm, this study focus most on how to improve face detection using luminance components of colour space. Different colour spaces were studied to choose gray scale standard. We have chosen standards which separate luminance to chrominance; these standards include CIELAB/CIELUV, YCbCr, Y'CbCr, HSV, and HSL…. The most important result obtain is the enhancement using gamma based luma technique as a lighting stabilizer before applying AHE. The enhancement method used is a mixture of global enhancement (Gamma transform) and local enhancement (AHE). The global enhancement stabilised the light while the local enhancement let the system to easily discriminate the main features, like eyes, mouth, and nose. The method leads us to a true positive rate of 99.31% with relative high false positive rate (23.89 %).

A new algorithm to calculate global PSNR of an image folder was proposed. For experiments LFW dataset was used, though no rigid general correlation between global PSNR and gray scale standard was found, experiments has proven that enhancing using CLAHE will produce better TPR than enhancing using GHE and AHE, since the over amplification found in GHE and AHE is not found in CLAHE. The global PSNR algorithm result shows that CLAHE produces better quality image than any other image(because it has the highest PSNR), yet experiments has shown that the quality of image has little impact in very low lighting condition using feature based detection method as seen in Fig. 5. Therefore, for an enhancement technique to be used it should be able to discriminate the facial features.

The work done in this article can be extended in many ways. Some possible directions include, assessing the global PSNR and their face detection accuracy using other enhancing method, also our study was applied on Viola Jones haar like features only, this study can still be expanded to other features like the LBP, the HOG, SIFT… to also study the impact of gray scale standard on detection rate. The proposed method produces a significant false positive rate; other research can be done on how to reduce this false positive rate.

REFERENCES

[1] Pizer S M, Amburn E P, Austin J D, Cromartie R, Geselowitz A, Geer T, Romeny B H, Zimmerman J B, and Zuiderveld K. Adaptive histogram equalization and its variations. Computer Vision, Graphics and Image Processing, September 1987, 39(3): 355-368.

[2] Zuiderveld K. Contrast limited adaptive histogram equalization. In: Graphics Gems IV. Academic Press Professional, Inc. 1994, pp. 474–485.

[3] P. Viola and M J Jones. Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE CVPR, 2001.

[4] Huang G B, Ramesh M, Berg T. Learned-Miller, Labeled faces in the wild: A database for studying face recognition in unconstrained environments, University of Massachusetts, Amherst, 2007, Tech. Rep. 07-49.

[5] Song F, Tan X, Chen S, Zhou Z. A literature survey on robust and efficient eye localization in real-life scenarios, Pattern Recognition, 2013, http://dx.doi.org/10.1016/j.patcog.2013.05.009i.

[6] Ming-Husan Yang,David J.Kriegman,Narendra Ahuja ,"Detecting Faces in Images:A survey" IEEE Transaction on Pattern Analysis and Machine Intelligence",Vol.24,No.1,Jan 2002..

[7] Ce Zhan. Facial Expression recognition for multilayer online games. MCompSc Thesis, School of Computer Science and Software engineering, University of Wollonghong. 2008.

[8] Rainer Lienhart and Jochen Maydt. An Extended Set of Haar-like Features for Rapid Object Detection. Submitted to ICIP, 2002.

[9] Zhou Z H and Yu Y. Adaboost. In X. Wu and V. Kumar, ed. The Top Ten Algorithms in Data Mining. Chapman & Hall, Boca Raton, FL, 2009.

[10] Duncan S. and Velthuis T. Nao detection with a cascade of boosted weak classifier based on Haar-like features. Bachelor thesis, Faculty of Science, University of Amsterdam, 2014.

[11] Huanzhao Zeng. Preferred Skin Colour Reproduction, PhD thesis, The University of Leeds, September, 2011.

[12] Acharya T, Ray A. Image processing: principles and applications Wiley-Interscience, 2005.

[13] Charles Poynton, Digital Video and HDTV, Chapter 24, pp. 291–292, Morgan Kaufmann, 2003.

[14] Recommendation ITU-R BT.2020-1, Parameter values for ultra-high definition television systems for production and international programme exchange. ITU, 2014.

[15] Recommendation ITU-R BT.709-5, Parameter values for the HDTV standards for production and international programme exchange. ITU, 2002.

[16] Bedi S.S. and Rati Khandelwal. Various Image Enhancement Techniques. A Critical Review. International Journal of Advanced Research in Computer and Communication Engineering 2013 2(3),.

[17] Anish Kumar Vishwakarma. Color Image Enhancement Techniques: A Critical Review. International Journal of computer science and engineering, 2012 3(1).

[18] Gonzalez R C and Woods R E. Digital Image Processing, Third Edition, 2008.

[19] Agaian, SOS S., Blair Silver, Karen A. Panetta, "Transform Coefficient Histogram-Based Image Enhancement Algorithms Using Contrast Entropy", IEEE Transaction on Image Processing, Vol. 16, No. 3, March 2007.

[20] Arun R, Madhu S. Nair, R. Vrinthavani and Rao Tatavarti. "An Alpha Rooting Based Hybrid Technique for Image Enhancement". Online publication in IAENG, 24th August 2011.

[21] Li Tao, Ming-Jung Seow and Vijayan K. Asari. Nonlinear Image Enhancement to Improve Face Detection in Complex Lighting Environment. International Journal of Computational Intelligence Research. 2006, 2(4), pp. 327-336.

[22] Phillips P J, Flynn P J, Scruggs T, Bowyer K W, Chang J, Hoffman K, Marques J, Min J, and Worek W "Overview of the face recognition grand challenge," Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2005,pp. 947-954.

[23] Venkatesan M.E, Srinivasa Rao. Face Detection by Hybrid Genetic and Ant Colony Optimization Algorithm. International Journal of Computer Applications, 2010, 9(4).

[24] Deepak Ghimire and Joonwhoan Lee, A Robust Face Detection Method Based on Skin Colour and Edges, Journal of Inf Process Syst, 2013, 9(1).

[25] Sonam bharal. L*a*b based contrast limited adaptive histogram equalization for underwater images. International Journal of Computer Application (2250-1797), 2015, 5(4).

[26] Hitam, M.S.; Yussof, W.J.; Awalludin, E.A.; Bachok, Z. Mixture contrast limited adaptive histogram Equalization for Underwater Image Enhancement. International Conference on Computer Applications Technology. 2013, 1, 1-5.

[27] Richard Szeliski. "Computer Vision:Algorithms and Applications". September, 2010 pp 228.