Synthetic remote PPG models for validation and improvement of the automatic gain tuning algorithm

Madanda, R.

Award date:
2016

Technische Universiteit Eindhoven University of Technology
Department of Electrical Engineering.
Electronic Systems Research Group.

Synthetic Remote PPG Models for Validation and Improvement of the Automatic Gain Tuning Algorithm

Master Thesis

Author: Richard Madanda

Supervisors:
Prof. Dr. Ir. Gerard de Haan
Dr. Ir. Sander Stuijk

Version Alpha 2.0

Eindhoven, August 2015

Abstract

Remote photoplethysmography (rPPG) is a method of measuring cardiovascular activity using a video camera. Currently, development of rPPG signal extraction methods is based on real videos from human volunteers. In this research, we develop synthetic videos based on 3D morphable models, which contain physiological signals of predictable amplitude and periodicity which mimic the human facial skin regions. With this model, reproducible and realistic experiments under different illumination and motion scenarios are developed. Using state-of-the-art rPPG algorithms, we prove that the heart rate signal can be extracted precisely as encoded using an RGB camera on a normal LCD. Synthetic models with dynamic illumination are used to validate an Automatic Gain Tuning algorithm (AGT) which aims at increasing SNR by reducing pixel clipping. Results show that the AGT leads to subtle changes in SNR with gains only being noticeable and consistent with high clipping levels. We propose and implement improvements to the AGT which increases gain tuning freedom by extending the implementation to include automatic exposure. Adjusting exposure shows superior improvement in SNR compared to independent gain tuning.

Acknowledgement

In a special way, I thank Prof. Gerard de Haan, Dr Sander Stuijk and Mark van Gastel for giving me the opportunity, conducive environment, guidance and all necessary support to carry out my graduation thesis at Philips Research at the High Tech Campus, Eindhoven, the Netherlands. I am also forever indebted to my parents Mr. James Mafabi and Mrs Regina Nambozo.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background

The human cardiovascular system consists of the heart, blood vessels and blood. In clinical and health monitoring systems, the cardiovascular system activity is used to detect vital human signs such as heart rate (HR), oxygen saturation (pulse oximetry), respiration rate (RR), blood pressure, cardiac output and in diagnosis of peripheral vascular diseases etc. Detection of cardiovascular activity can be done by either contact-based methods such as electrocardiogram (ECG), photoplethysmography (PPG), or contactless methods such as radar, thermal imaging and remote photoplethysmography (rPPG). ECG is a classical contact-based technique used to measure cardiac activity. In ECG, shown in Fig 1.1, the heartbeat signal is obtained by measuring the heart's electrical impulses generated by polarization and depolarization of cardiac tissue using electrodes attached to the body. These impulses are translated into a waveform whose periodicity corresponds to the heart rate.

PPG is the optical measurement of blood volume changes [1]. It is a popular non-invasive contact-based method used for measuring several human vital signs including; pulse rate, oxygen saturation and respiration[2]. In PPG, shown in Fig. 1.1, fluorescent body parts such as the skin are illuminated with illumination of different wavelengths. The intensity of reflected or transmitted light is measured using a photodiode. The measured intensity contains time varying signals, AC components superimposed onto DC components. The time varying signal is as a result of variation in the amount of blood in microvascular bed tissue[2],[3]. The DC components are as a result of the properties of the illuminated skin surface of subjects and other lower frequency components due to other body rhythmic processes such as respiration[2].

Contactless measurement of the heart rate (HR) can be achieved by using radar, thermal camera or a video camera. The most promising of these methods is the use of the video camera because of its inexpensive implementation. This implementation referred to as remote PPG (rPPG), has motivated research interest in recent years as demonstrated in the papers [1], [4], [5], [6], [7]. In rPPG, shown in Fig. 1.1, the minute optical absorption changes caused by blood volume variations in the skin are detected by a video camera. Remote PPG is most desirable in cases where contact has to be prevented because of extreme sensitivity (e.g. neonates), or when unobtrusiveness is essential or desirable e.g. in surveillance, fitness etc [7].

Subject motion and illumination changes affect the perceived color of skin pixels. To retrieve a clean pulse signal independent of other changes other than the physiological changes, motion and illumination robust methods are needed. Efforts have been made towards motion and illumination robust rPPG by using various techniques including; Blind Source Separation (BSS) techniques [5],[8], Chrominance method [6] and blood volume based pulse signature method[7]. The effects of dynamic illumination on rPPG signal measurements have also been studied by Liu *et al* [9]. To improve the performance of the heart rate extraction methods under dynamic illumination, Andreas Papageorgiou [10], a master student at Philips research, developed a

Figure 1.1: Contact-based HR measurement

Figure 1.2: Contact less HR measurement

camera based Automatic Gain Tuning algorithm (AGT) which uses a ROI based white and black clipped pixels measurement to predict camera gains which give minimum clipping.

Development and benchmarking of current rPPG algorithms is based on actual video recordings. Currently, it is impossible to reproduce an experiment with exactly the same physiological signals. Therefore, this research's aim was to create synthetic videos, with realistic physiological and skin optical properties, which mimic the optical properties of the facial skin regions under various illumination and motion conditions. This addresses the need for reproducible experiments. The synthetic videos are benchmarked using current state-of-the- art rPPG algorithms, the CHROM [6] and PBV methods.[7], and are consequently used to validate and improve the AGT algorithm developed at Philips.

In summary, this paper's main contributions are; i) Modelling synthetic rPPG videos based on 3D morphable models (3DMM), ii) Benchmarking and validating a ROI based real time gain tuning algorithm and, iii) Extending the implementation to automatic gain and exposure tuning. The suggested improvements are based on signal-to-noise ratio (SNR) measurements and clipping percentage benchmarking of the original AGT.

1.2 Research Motivation

The periodicity and amplitude of the pulse signal of any given subject are dynamic even in the most controlled experiment conditions [1]. This typical nature is shown in Fig. 1.3. The observed trend in amplitude and frequency can either increase or decrease over time. Consequently, it is unlikely to reproduce an experiment in which the extracted PPG signal is the same. This complicates the testing of specific algorithms for experiments which involve varying the subject's pose, skin types and dynamic illumination.

Detection of the rPPG signal involves pixel-based processing applied to video frames. Similar to other computer vision research areas like face detection, skin detection, feature extraction etc, a reliable and consistent database of videos for rPPG algorithm development, testing and training is required. The first step of developing such a data set is to realise a synthetic human face. In computer vision, realistic face models have been developed based on 3D morphable models (3DMM)[11]. The use of 3DMM in generating synthetic faces is promising although, it still faces challenges of generating very realistic faces using reasonable computation resources. Many researchers believe that 3DMM constitutes the state-of-the-art for face image-based analysis[12]. Based on this premise, 3DMM can be used to generate reliable and reproducible synthetic experimentation videos for rPPG research.

The primary goal of this research is therefore, to bridge a gap in the area of rPPG experimentation by developing synthetic and yet realistic videos from 3DMM with the required skin attributes which contain the relevant vital signals. This shall address the need to have experiments which are reproducible with the exact physiological signals.

The first step in calculating a heart rate from videos is to extract the mean raw RGB signals which are obtained by spatial averaging of pixel values in the ROI. These raw signals contain minute color variations due to changes in blood volume in the blood vessels and variations due to an external light source i.e intensity and color changes. The current state-of-art algorithms i.e. CHROM and PBV are robust under different light and color variations. It is important to note that

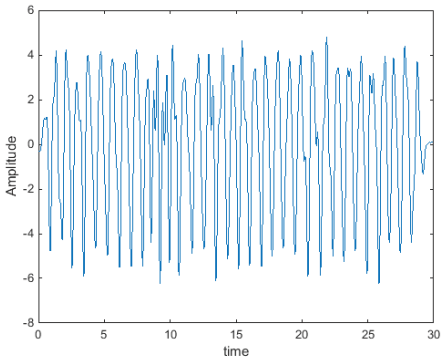

Figure 1.3: Dynamic nature of the rPPG signal

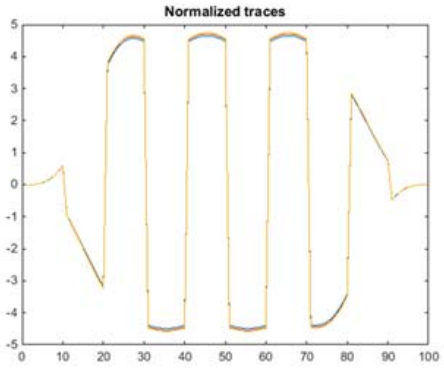

Figure 1.4: Clipping of RGB normalised traces due to very high and low light intensity changes

large changes in intensity (very high or very low intensity) will lead to clipped pixels (Pixels over 255 or under 0 for uint8 RGB pixel representation are clipped). The effect of clipping RGB sine signals traces is shown in Fig. 1.4. The SNR of signals extracted by current rPPG methods reduces when pixels are clipped because the minute color changes are no longer present. To improve the performance of the PPG extraction methods, Andreas [10], implemented an automatic gain tuning algorithm (AGT) with the CHROM algorithm. This algorithm was tested under different light conditions (constant light, extreme ambient light and varying lighting mimicking cloud cover) on one subject. The results from these experiments were not conclusive because the experiments in the two scenarios "AGT-OFF" and "AGT-ON" were not identical. Therefore, the secondary goal of this research is to use the synthetic models to validate the AGT under different illuminations scenarios. The performance results are expected to be more accurate and conclusive. These results shall also inform the areas of improvements to increase the robustness of the AGT.

1.3 Research Objectives

The main research objectives in this paper are;

1. Develop realistic human face model. These models should contain realistic human features

2. Implement physiological and optical properties of the skin. Model the skin using a suitable skin color representation to represent the DC component and a suitable pulse model to represent the AC component.

3. Study and model the effects of illumination on the rPPG signal. By using a suitable light and shading model to reproduce various dynamic illumination.

4. Model different pose and motion scenarios. The motion scenarios include; Translation, rotation and zoom

5. Benchmark the models for consistency and accuracy with state-of- the-art rPPG algorithms and the Automatic Gain Tuning (AGT) algorithm developed by Papageorgiou. In addition, validate and suggest/implement improvements to the Automatic Gain Tuning algorithm developed at Philips.

6. Implement a user interface for rendering videos with different pose, illumination, pulse rates and motion.

1.4 Related Work

At the time of this writing, synthetic videos of human faces containing pulsatile signal information have not been published in any accessible literature. It is important to note though that in numerous medical applications, synthetic phantoms have been developed e.g. the phantoms with tissue like properties for the human skin presented in [13] and other medical imaging phantoms for CT scan based research.

Morphable models for 2D and 3D image processing have been developed in various studies; by Blanz and Vetter [11], Paysan *et al*[12], Chung *et al*[14], Lee *et al*[15]. These generative models provide pose and illumination invariance. They achieve shape, pose and skin color from real world data captured using high definition 3D scanners. The morphable models can be used to generate numerous random faces by using statistical model fits to real data.

Modelling different skin types requires a comprehensive skin model. Kakumanu *et al* [16], presents a synopsis of skin modelling and detection methods. It is noted that representation of skin color in digital images depends largely on spectral reflectance, prevailing illumination conditions and camera characteristics. It also presents a variety of color spaces used in skin representation including; basic color spaces RGB, normalised RGB, CIE-XYZ; perceptual color spaces HSI, HSV, HSL and TSL; orthogonal color spaces YCbCr, YIQ, and YUV. A mixture of the different color spaces has also been used in skin models such as the RGB-H-CbCr skin model presented by Anwar *et al* [17].

Yang *et al*[18] presents a statistical human skin model using a multivariate normal distribution in the normalised color space under specific lighting conditions. For dynamic lighting conditions, an adaptive model is proposed. Hayit *et al* [19] proposes a model for face color modelling which uses a mixture of Gaussians for robust representation of the skin colors under different conditions of shadows and illumination. Ming-Hsuan *et al* [20] uses a Gaussian mixture model for human skin color modelling.

Skin color has also been modelled using illumination reflection. This offers sufficient accuracy in mimicking the skin color as recorded by a camera [21]. Human skin reflectance is an important parameter in skin representation and detection in the visible and near-infrared imaging [22]. The dichromatic reflection model used to characterize a variety of reflectance properties of materials proposed by Tominaga [23] provides such a representation. Weyrich *et al*[24], developed a skin reflectance database and model to achieve Torrance-Sparrow and Blinn-Phong analytic BRDF face model for different skin types according to the Fitzpatrick scale [25] given in Table 3.2.

Hüelsbusch and Blazek [1], elaborate the dynamic nature of the pulse signal obtained by PPG as shown in Fig. 1.3. Allen [2], presents the characteristics of the PPG signal and its relation in appearance to the ECG waveforms shown in Fig. 3.14. Martin-Martinez *et al* [26], present a stochastic model of the PPG signal based on the shape parameterization of the PPG wave and non-stationery model of its time evolution. The PPG signal is reproduced by a mixture of different Gaussian-shape like waves which vary with time.

To accurately represent the actual lighting conditions, scene illumination models are needed. Existing illumination models represent unidirectional or point light source types. In [27] a unified general light source model is presented which models illumination of different types within a single unified framework.

1.5 Paper Overview

The rest of this paper is organised as follows; In **Chapter 2**, PPG-based heart rate detection methods are elaborated. The existing attempts to reduce the effect of dynamic illumination and motion in the current state-of-the-art algorithms are also given. In **Chapter 3**, the synthetic video development and implementation details are discussed. The implementation of dynamic illumination and animations are further detailed. Finally the limitations of the implemented models are presented. In **Chapter 4**, the AGT implementation, the proposed and implemented improvements are discussed. In **Chapter 5**, the experiment results for the implementations,

"agt-off", "agt-on" and improved "agt-exp" and "agt-imp" are discussed. The results are based on the measured SNR and reduction in the clipping percentage and finally **Chapter 6** contains the general conclusions and observations.

Chapter 2

PPG-Based Heart Rate Detection

2.1 Introduction

In this chapter, the theory and implementation of of heart rate extraction from videos is presented. Special attention is given to the effect of light and motion on PPG measurement. This gives a strong background to the AGT benchmark experiments under different illuminations presented in chapter 5.

2.2 PPG Measurement Principle

PPG measurements require a light source to illuminate a body part with blood vessels. Depending on the wavelength of the light source, the light can penetrate up to a depth of 3 mm into the skin and be partially backscattered [1]. The reflected light is detected by a camera sensor. The reflection consists of a diffuse (body) reflection which has penetrated the skin and specular reflection which is directly reflected from the surface of the skin see Fig 2.1. The diffuse reflection shows the color variations due to the rythmic movement of the blood in the micro-vascular tissues under the skin and a stationery component due to the skin colour. The specular reflection contains a constant color component due to the light source and no pulse signal. Using a camera, the sum total of the observed color of a measured skin contains the diffuse and specular reflection.

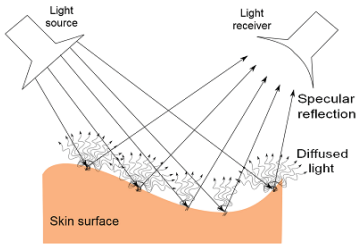

Figure 2.1: Specular and diffuse Reflection

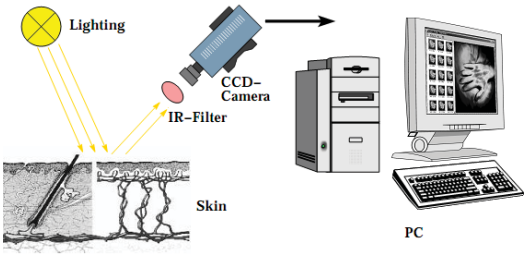

Figure 2.2: PPG Imaging

2.3 Effect of Light and Motion on PPG

The amplitude of the reflected light depends on the wavelength of the light source. Near infra-red light (NIR) penetrates the skin (epidermis) deeper whereas visible light i.e red, green and blue have a lower skin penetration. Additionally, the absorption of haemoglobin is about 10 times higher than that of bloodless tissue. Therefore, there is high illumination absorption if the blood contents

in the skin is high and vice versa. Using the normal RGB camera, the pulse signal amplitude is highest in the green channel compared to the red and blue because of the high absorption and of green light by haemoglobin. PPG measurements are influenced by angles between the camera, skin and the light source. Changes in position of the skin, camera and the light source affect the relative contribution of specular and diffuse reflection to the total skin reflection.

2.3.1 Effect of illumination Illumination Colour(wavelength of light)

Measurement of PPG is effected by color variations of the illumination source. The variation of measured rPPG amplitude with wavelength of the light source is shown in Fig. 2.3 and Fig. 2.4. The green $495 - 570$ nm component of the PPG signal from an RGB camera has a higher amplitude than that of the red $620 - 750$ nm and blue $450 - 495$ nm channels. It has also been established that above 590 nm, physiological pulsations are more apparent in the red channel than that of the green and blue channels [9].

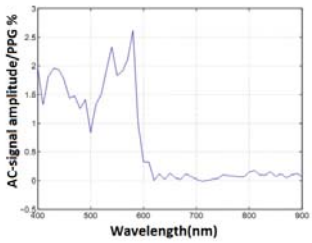

Figure 2.3: PPG amplitude as a function of wavelength

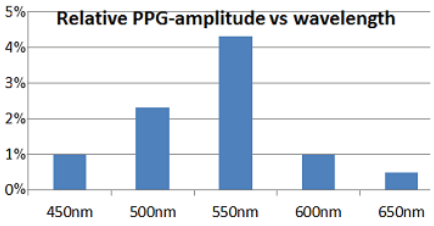

Figure 2.4: PPG signal amplitude vs illumination wavelength

2.3.2 Effect of Light Intensity

In chapter 3, and in [6] the Dichromatic model [23] is used to mimic the appearance of skin pixels in terms of intensity of the light, the stationery component of reflection(DC), the pulsatile component (AC) and the specular reflection. From this model, it is predicted that increase in light intensity leads to an increase in the PPG amplitude. It is also obvious that this increase also leads to an increase in the DC and specular reflection. In the next section, current rPPG extraction methods which attempt to improve motion robustness are presented.

2.4 Heart Rate Extraction Methods

Several methods have been proposed in literature for heart rate extraction from RGB videos. The existing methods include; i) Blind Source Separation-based methods i.e. Independent Component Analysis (ICA) based method by Poh [5] and Principal Component Analysis method by Lewandowska [4], ii) CHROM method of de Haan and Jeanne [6] and iii) PBV method of de Haan and Leest [7].

2.4.1 BSS-based methods

BSS is a technique for noise removal from physiological signals. BSS refers to the separation of a set of mixed signals (with or without little information about the source or the mixing of the signals). This is used where a source signal is recorded using a set of sensors where each sensor receives a different combination of the source signals. In rPPG, BSS has been applied by either using ICA or PCA. The Independent Component Analysis (ICA) is a statistical and computational technique used to separate independent signals from a set of observations that consist of linear mixtures of underlying sources [28]. PCA is a statistical transformation that identifies patterns in

data and expresses the data in such a way that it highlights the similarities and differences. The extraction of a signal from videos using BSS-methods is shown in Fig. 2.5.

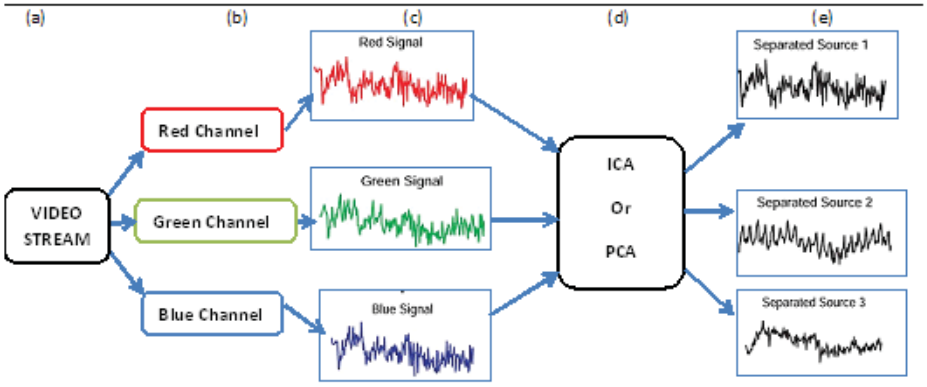

Figure 2.5: Pulse rate extraction using ICA and PCA

Signal extraction using BSS-based methods i.e the ICA involves; (a) Face recognition within a suitable face recognition algorithm and automatic detection of the ROI, (b) the ROI is decomposed into red, green and blue channels which are spatially averaged to obtain the raw signals, (c) the raw signals are detrended and normalised, (d) Independent component analysis is applied to separate the three independent sources. (e) The pulse rate is obtained from one of the sources i.e the second source. The The pulse-signal $\vec{S}$ obtained with the BSS-method is written as a linear combination of the individual mean-centered normalised color channels C_n; $\vec{S} = \vec{W}C_n$, where the linear weighting $\vec{W}$ with $\vec{W}W^T = 1$ is obtained using the BSS methods (ICA and PCA). Each of the rows of the $3XN$ matrix C_n contains N samples of the mean-centered normalised color signals, $\vec{R_n}$, $\vec{G_n}$, $\vec{B_n}$. These are given by;

$$\vec{R_n} = \frac{1}{\mu\vec{R}}\vec{R} - 1, \quad \vec{G_n} = \frac{1}{\mu\vec{G}}\vec{G} - 1, \quad \vec{B_n} = \frac{1}{\mu\vec{B}}\vec{B} - 1 \tag{2.1}$$

Where μ corresponds to the temporal mean (over N frames), and the vectors $\vec{R}, \vec{G}, \vec{B}$ contain the spatial mean of the red, green and blue pixels.

2.4.2 Chrominance method

Extraction of a pulse signal from a video using the CHROM method consists of five major steps shown in Fig 2.6 [6].

1. Color Normalisation: To produce a pulse signal that is independent of the stationary color of the light source, as well as its brightness levels, each color channel is normalised by dividing its samples with their mean over a temporal interval. This color normalization makes the algorithm robust under different color and intensity variations of the ambient light. The length of the temporal interval must guarantee that it contains at least one pulse period.

2. Color signal definition: The color signals $\vec{X}$ and $\vec{Y}$ are defined by the following equations:

$$\vec{X} = 0.77\vec{R_n} - 0.51\vec{G_n}, \vec{Y} = 0.77\vec{R_n} + 0.51\vec{G_n} - 0.77\vec{B_n} \tag{2.2}$$

   Using color difference signals (chrominance), ensures that the light component that is directly reflected from the surface of the skin is eliminated. To eliminate motion which affects chrominance signals identically, a ratio of the two chrominance signals is taken.

3. Skin-tone standardization: To enable correct functionality under different colored light sources, a standard skin tone is used to normalize the color channels with the fixed standardized vector [R,G,B]=[0.7682,0.5121,0.3841].

4. Distortion Minimization: The human heart rate ranges between $\approx 40-240$ beats per minute (BPM), distortions out of this range are removed by using a band-pass filter.

5. Overlap Add The overlap-add procedure makes the algorithm more adaptive to quick changes in the Heart rate which occur under different-intensity or motion intervals. It improves the PPG signal by separately optimizing partially overlapping time intervals. The resulting pieces are stitched together in an overlap-add fashion which is accommodated with the use of a Hann window. In practice, after normalizing and minimizing the signal, it is multiplied with a hanning window of a specified size Hann-size = 32. A step size of; Step-size = Hann-size/4 is used along the time axis and this procedure is repeated. To construct the final signal the partially overlapping signals are added.

The pulse signal is computed from the normalised color channels as

$$\vec{S} = \vec{X} - \alpha\vec{Y}, \alpha = \frac{\delta(\vec{X})}{\delta\vec{Y}} \tag{2.3}$$

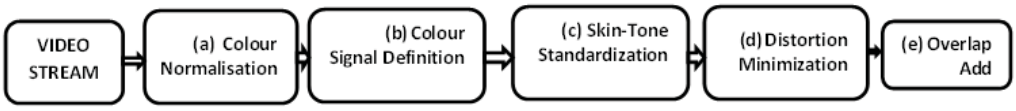

Figure 2.6: CHROM algorithm

2.4.3 PBV-method

The Pulse Blood Volume(PBV) method is based on the unique 'signature' of blood volume changes expressed as a vector. The $\overrightarrow{P_{bv}}$ vector is the relative PPG-amplitude in the normalised RGB-color channels is given by;

$$\overrightarrow{P_{bv}} = \frac{[\sigma(\vec{R}_n), \sigma(\vec{G}_n), \sigma(\rightarrow n)]}{\sqrt{\sigma^2(\vec{R}_n) + \sigma^2(\vec{G}_n) + \sigma^2(\vec{B}_n)}} \tag{2.4}$$

In [7], the pbv vector is found to be relatively stable under white illumination for different skin tones. The pulse signal is built as a linear combination of normalised color channels. To achieve this, a $\vec{W}PBV$ is sought, that gives a pulse-signal for which the correlation with the color channels $\overrightarrow{R_n}$, $\overrightarrow{G_n}$ $\overrightarrow{B_n}$ equals $\vec{P}_{bv}$ according to the expression in equation 2.5.

$$\vec{S}{C_n}^T = k\vec{P}_{bv} \Leftrightarrow \overrightarrow{W_{PBV}}C_n{C_n}^T = k\vec{P}_{bv} \tag{2.5}$$

The scale factor k is chosen such that $\vec{W}PBV$ has unit length.The ovelap-add procedure similar to the CHROM method is used to construct the pulse signal.

2.5 Summary

In this chapter, effects of light and motion on the PPG signal have also been discussed. The existing heart rate extraction methods have also been presented. In the next chapter, we present the modelling of the synthetic face model with PPG signals.

Chapter 3

The PPG Face Phantom

3.1 Introduction

Synthetic videos with physiological and skin optical properties, which mimic the human facial skin under various illumination and motion conditions developed and implemented in this research, are based on synthetic 3D Morphable models(3DMM). In computer vision synthetic 3DMM have become a powerful tool in face recognition tasks. 3DMM was first proposed and developed by Blanz and Vetter of the University of Basel as a sibling of 3D AAM [11]. 3DMM statistically builds a joint model based on 3D shape and texture in 2D. Human face phantoms based on 3DMM are generated from several 2D images or 3D face scans by using feature points on the human face. The current state-of-the art methods used involve face registration (triangulation) and parametrisation. Triangulation/reconstruction is a process by which a point in 3D space is calculated from two or more images where that point is visible [29]. 3DMM parametrisation involves specifying a model to represent the shape and the skin color of the human face. Parametrisation can also be useful in changing the shape and texture attributes of the face in a natural way to generate other faces. In most cases it is sufficient to use the cartesian (X, Y, Z) or spherical co-ordinate system (r, θ, ϕ) to define a point in space(the shape form). The skin color model used varies depending on the purpose of the 3DMM. Commonly used skin color representation are presented in the section 3.3. 3DMM is suitable for 3D face animation and potentially useful for 3D face recognition tasks. Construction of PPG simulating videos requires animation of 3DMM adding parameters (model) which mimics the skin colour changes with time.

3.2 The Basel Face Model

In this research, the Basel Face Model (BFM) presented in [12] shown in Fig. 3.1 was used. This model assumes independence between human shape and texture (color). Dense triangulations are used to achieve higher resolution and detailed facial features and landmarks like the cheek elevations. These dense triangulations produce a realistic face shape. The BFM model uses a triangulation of 53,490 vertices each with associated RGB color. Each vertex is associated with a position vector $[x, y, z]$ and a color vector, $[R, G, B]$. Since face shape is not as important to this research as the face color variations, the parametrised shape model of the BFM is adopted for generating random faces upon which a unique face color model is superimposed. The BFM shape model, also provides capability to change shape and incorporate various characteristics like age, weight, gender etc. These are of less importance for this research's objectives but they are an added benefit for exploration purposes. The main disadvantage of the BFM shape is that it only achieves a natural expression of the face. Therefore, it is difficult to produce various facial expressions like smile, anger etc. These properties too are of less importance to the research topic.

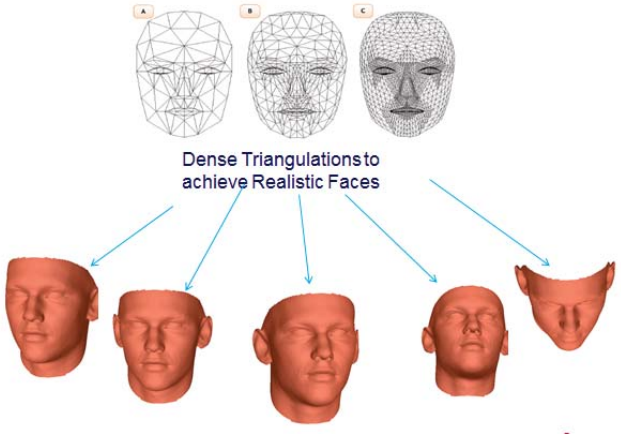

Figure 3.1: The Basel Face Model

3.2.1 BFM Model construction

The construction of a reliable 3DMM requires a training set with a variety of shapes and appearance. The BFM is based on face scans of 100 females and 100 males mostly Europeans with ages ranging between 8 and 62 years with an average age of 24.7 years and weights between 40 and 123 kg with an average of 66.48 kg.

To effectively parameterize data from the 3D scans to achieve a fit of scan data to model correspondence i.e. data from a noise tip should be represented by only one point in the model, registration is performed as shown in Fig. 3.3. Each BFM is parameterised by triangular meshes each with $m = 53,490$ vertices and a shared topology. Each vertex $(x_i, y_j, z_j)^T \in \mathbb{R}$ has an associated color $(r_j, g_j, b_j)^T \in [0,1]^3$. The face is represented by $3m$ dimensional vectors.

$$s = (x_1, y_1, z_1, ...x_m, y_m, z_m)^T \tag{3.1}$$
$$t = (r_1, g_1, b_1, ...r_m, g_m, b_m)^T \tag{3.2}$$

A Guassian distribution is fit to the data using Principle Component Analysis(PCA), resulting into a parametric face model consisting of

$$M_s = (\mu_s, \delta_s, U_s) \text{ and } M_t = (\mu_t, \delta_t, U_t) \tag{3.3}$$

Where $\mu_{s,t} \in \mathbb{R}^{3m}$ are the mean, $\delta_{s,t} \in \mathbb{R}^{n-1}$ the standard deviation and $U_{s,t} = [u_1, u_2, ...u_n] \in (R)^{3mxn-1}$ are the orthonormal basis of principal components of shape and texture. New faces with different attributes of age, weight, height, gender, etc. are generated as shown in Fig.3.3 are generated from this parametric model.

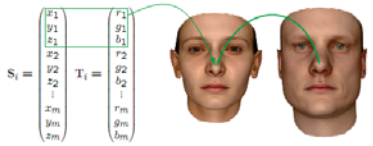

Figure 3.2: Each entry in shape and color vector correspond to the same point on the faces i.e the tip of the nose for this case.

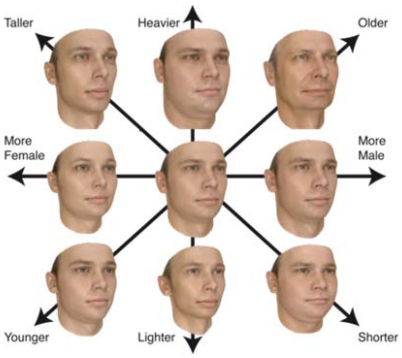

Figure 3.3: The parameterised model is used to produce faces with different attributes

3.3 The Skin Color Model

Although the RGB color space used in the BFM is adequate to accurately represent the color of the face, its skin types are limited to subjects of one ethnicity (mainly Europeans). It has already been noted that a skin model based on skin reflectances can sufficiently reproduce the required skin representation achieving a wide variety of skin types. The dichromatic model is one of the the classical models based on skin reflectance. According to the dichromatic model, the intensity of a given pixel in image number i in each color channel $C \in R, G, B$ is represented as;

$$C_i = I_{ci}(\rho C_{dc} + \rho C_{ic} + S_i) \tag{3.4}$$

Where I_{ci} is the intensity of the light source integrated over the exposure time of the camera in image i for the particular color channel, ρC_{dc} is the stationery part of the reflection coefficient of the skin in that particular color channel and ρC_{ic} is the pulsatile component. The DC and pulsatile coefficients account for the diffuse reflection. The diffuse reflection is due to illumination which has penetrated the skin surface and contains the cardiac cycle due to blood volume changes in blood vessels underneath the skin. S_i is the specular reflection coefficient which accounts for the percentage of light which is reflected directly from the skin surface without skin penetration.

Each substance has its own unique reflectance characteristics [22]. The human skin has lower reflectance at shorter wavelengths ($about 350\ nm$) than at longer wavelengths (about $1050\ nm$). Therefore, in the RGB color space, the skin reflects more red light ($\approx 620 - 740\ nm$) followed by green light ($\approx 495 - 570\ nm$) and least blue light ($\approx 450 - 495\ nm$). Furthermore, for different ethnicities, i.e. Negroid, Mongolid and Caucasoid, these coefficients are different as shown in Table 3.1. With this data, typical skin types as shown in Fig. 3.5 can be rendered.

Weyrich *et al*[24], present a face reflectance model with surface reflectance and sub-surface reflectance which accurately reproduces the appearance of the human skin as seen in a digital photograph. This is particularly relevant for rPPG signal modelling because the pulsatile signal originates from the subsurface reflection. With these skin reflectance and dichromatic models, skin and pulse rate signals are appropriately modelled. The various skin types according to the Fitzpatrick scale, Table 3.2, are also rendered.

Table 3.1: Ethnicity skin reflectances

Ethnicity	Red	Green	Blue
Negroid	$0.1 - 0.4$	$0.05 - 0.08$	$0.01 - 0.05$
Caucasoid	$0.9 - 0.99$	$0.35 - 0.38$	$0.3 - 0.35$
Mongoloid	$0.9 - 0.99$	$0.3 - 0.35$	$0.3 - 0.35$

Table 3.2: Fitzpatrick Scale

Skin Type	Skin Color
I	Pale white
II	White
III	Cream White
IV	Moderate Brown
V	Dark brown
VI	Dark Skin

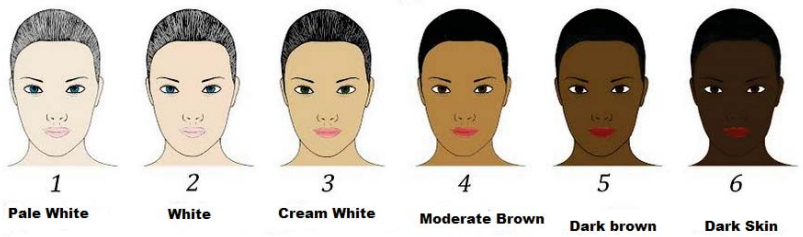

Figure 3.4: Fitzpatrick scale

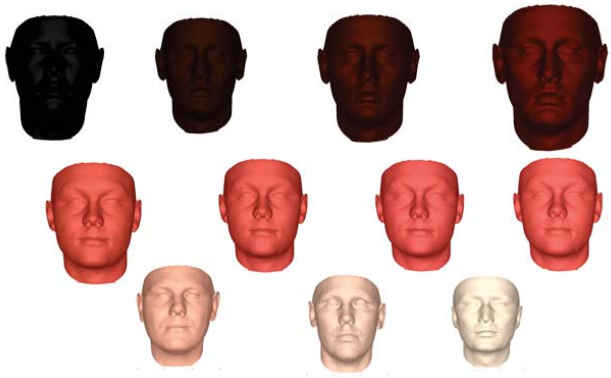

Figure 3.5: Fitzpatrick Skin color modelling using Dichromatic model and Ethnic Reflectances

3.4 Scene Illumination

The light intensity and spectra (different wavelengths) used in a PPG experiment have a measurable effect on the DC and AC components. In [9], it is shown that with an increase in light intensity, the measured DC and AC components increase. Furthermore, it is premised that background scene lighting, whatever color this may be may have significant effect on the rPPG signal. To model the effects of lighting, a suitable light model is needed. Extensive light source models have been used in computer vision and computer graphics, among which the most common are point light source, directional light source and area light sources. It is important to note that in the real world, the situation could be very complex where multiple different types of light sources co-exist in the same scene. Such a scenario is investigated in the rPPG signal measurements in an automotive environment under dynamic illumination conditions by Jeanne *et al*[30]. In Fig. 3.6, the effect of illumination sources of different colours on the skin appearance are illustrated.

Figure 3.6: Different illumination scenarios

3.4.1 Light source

Realism is added to the 3DMM using light objects in Matlab graphics environment. A light object is characterised by its color, position and style. The light color is an RGB triplet $[I_R, I_G, I_B]$ where I is the intensity in the different color channels in the range $[0, 1]$. Depending on the nature of a light source, the light ray direction is either modelled as parallel rays(see Fig. 3.7) emitting from a distant source(infinity) represented using a three element direction vector $[x, y, z]$ or a point source(see Fig. 3.8) defined by the position vector. The light position is defined using either the cartesian system using a three element vector, $[x, y, z]$ or using spherical coordinate system represented by the azimuth(az) and elevation(el) angles in relation to the target 3DMM object.

Figure 3.7: Parallel light represented using light direction, intensity and color

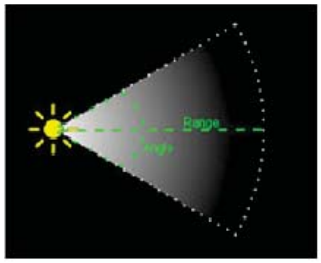

Figure 3.8: Point light source represented using light position, intensity and color

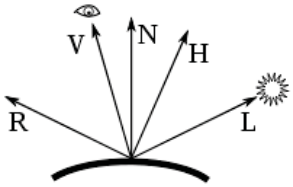

Figure 3.9: Phong shading vectors

3.4.2 Shading model

To render 3DMM objects realistically to approximate specular,diffuse and ambient reflection from the 3DMM triangulated faces, the phong reflection model and interpolation method developed by Tuong Phong [31] is used. The Phong reflection is an empirical model for local illumination. It describes the way a material reflects light as a combination of the diffuse, specular and ambient reflection. Specifically, it is based on interpolation of incident light surface normals across rasterized polygons and computes pixels colors based on the interpolated normals and a reflection model. The illumination of each point in a scene is computed as

$$I_p = k_a i_a + \sum_{m \in lights} (k_d(\hat{L}_m \cdot \hat{N}) i_{m,d} + k_s(\hat{R_m} \cdot \hat{V})^{\alpha} i_{m,s}). \tag{3.5}$$

Where k_s, k_d, k_a and α are material properties defined as; k_s Is a specular reflection constant, the ratio of reflection of the specular term of incoming light,
k_d, Is a diffuse reflection constant, the ratio of reflection of the diffuse term of incoming light (Lambertian reflectance),
k_a, Is an ambient reflection constant, the ratio of reflection of the ambient term present in all points in the scene rendered, and
α, which is a shininess constant for this material, which is larger for surfaces that are smoother and more mirror-like. When this constant is large, the specular highlight is small.

For all lights placed in the scene, the parameters defined are; $\hat{L}_m$, which is the direction vector from the point on the surface toward each light source (m specifies the light source),
$\hat{N}$, which is the normal at this point on the surface,
$\hat{R}_m$, which is the direction that a perfectly reflected ray of light would take from this point on the surface, and
$\hat{V}$, which is the direction pointing towards the viewer (such as a virtual camera).
where the direction vector $\hat{R}_m$ is calculated as the reflection of $\hat{L}_m$ on the surface characterized by the surface normal $\hat{N}$ using $\hat{R}_m = 2(\hat{L}_m \cdot \hat{N})\hat{N} - \hat{L}_m$. The vectors are shown in Fig 3.9.

3.5 PPG Signal Modelling

3.5.1 PPG Signal Waveform Model

The PPG signal waveform is typically a periodic signal with the same periodicity as the ECG signal. Fig. 3.14 shows typical ECG, PPG and rPPG signal waveforms which are extracted using current methods. The appearance of the pulse wave in Fig. 3.11, is defined in two phases: the anacrotic phase being the rising edge of the pulse, and the catacrotic phase being the falling edge of the pulse. The first phase is due to systole and the second diastole events of the cardiac cycle. Martin-Martinez *et al* [26], model the PPG signal waveform as a combination of statistical Gaussian-shaped waveforms which adequately mimic the actual observed PPG signal shape. In this model, the pulse signal corresponding to a measurement period T, is split into single pulses each with t_{on} as the start time and t_{end} the end time of the pulses respectively. As shown in Fig.3.13, each pulse is then modelled using a combination of two Gaussian curves S_1 and S_2 corresponding to the start and end of the pulse respectively. Since the pulses are dynamic, the subsequent pulses over the entire period of measurement T, are constructed from the parametrized Gaussian models S_1 and S_2.

3.5.2 Pulse Signal in RGB-Space

Hülsbusch [32], explained that the measured relative PPG signal amplitude is a function of the wavelength, λ, modelled according to Equation 3.6. This relationship is determined by the contrast between the blood (absorption spectra of oxy-and de-oxy haemoglobin) and the blood-free tissue. In Fig. 3.15, the variation of the PPG amplitude with the wavelength, λ is shown.

$$RPPG(\lambda) = \sigma(PPG(\lambda))/\mu(PPG(\lambda)) \tag{3.6}$$

The sensitivity of the red, green and blue sensors in the RGB camera used in rPPG measurement is shown in Fig. 3.16. A larger area of the AC signal amplitude in Fig.3.15, falls under the green channel of the camera spectral sensitivity. Therefore, the pulse signal amplitude in the RGB color channels is highest in the green channel as compared to the the blue and red channels respectively.

Figure 3.10: ECG

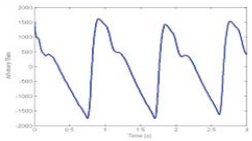

Figure 3.11: PPG

Figure 3.12: rPPG

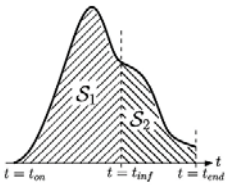

Figure 3.13: Pulse model

Figure 3.14: HB pulsatile waveforms

In [6], it is further shown that the normalised relative amplitudes due to the PPG signal in the RGB color channels can be approximated with a unique vector, [R,G,B] = [0.34, 0.77,0.54]. The AC pulsatile component is modelled based on these observations. For each pixel (vertex) in the BFM, an AC pulsatile waveform in the RGB color space is added to the DC value using three sine waveforms whose amplitudes satisfy the normalised vector and 0.01% of the DC level. As already observed in Section 3.4.1, the PPG signal is not a perfect sine wave, but for purposes of benchmarking the first implementation, a sine wave is used since the sinusoidal peaks mimic the systole and diastole events present in the PPG signal.

$$\begin{bmatrix} Pulse_R \\ Pulse_G \\ Pulse_B \end{bmatrix} = \begin{bmatrix} A_R \\ A_G \\ A_B \end{bmatrix} sin(2\pi ft/fs) \tag{3.7}$$

The simple pulsatile signal model is given by Equation 3.7. Where $A_{R,G,B}$ are the pulse signal amplitudes in each color channel, f is the pulse rate and f_s the sampling frequency.

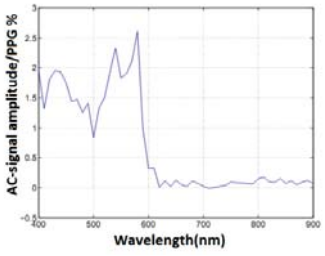

Figure 3.15: PPG amplitude as a function of wavelength

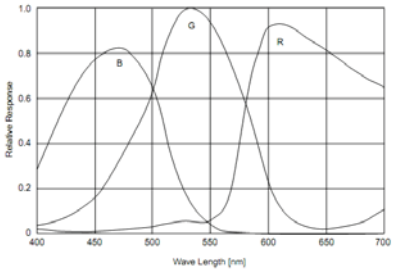

Figure 3.16: Spectral responses from the color channels of an RGB camera

3.6 Motion and Dynamic Lighting generation

To mimic motion and moving light sources, translations are imposed on either the 3DMM object or light object in each movie frame as explained hereafter

1. 3DMM objects can be translated, rotated, zoomed using the cameraview angles. The same effect can also be achieved by changing the $[x, y, z]$ position.

2. Dynamic lighting is achieved by changing the position or direction or light color of the light object per frame. This requires predefinition of a path along which the plotting of the light positions is defined. The position change my occur along the vertical plane, horizontal plane or both.

3. Periodic light flicker is generated by toggling lights in the scene by sampling '1's on a simple square wave whose frequency is pre-set depending on the required flickering frequency according to Equation 3.8

4. Periodic motion is generated by varying the position of the camera view or 3DMM object along a given trajectory(horizontal, vertical or zoom) by sampling different points on a sine wave according to equation 3.9.

$$\text{New Light Position} = \text{Old Light Position} * \text{square}(2\pi f_t/f_s) \tag{3.8}$$

$$\text{New Object Position} = \text{Old Object Position} * \sin(2\pi f_t/f_s) \tag{3.9}$$

where f_t is the frequency of motion and f_s is the sampling frequency.

3.7 Complete model generation

The complete synthetic model is generated according to the steps shown in the flow chart in Fig. 3.17 and Fig. 3.18. The DC level is selected and set to standardized skin. The AC pulse signal is set to 60 BPM in all models. The AC amplitude is set to 0.05 of the DC level and not the empirical value of 0.001. This is because at 0.001, the LCD screen induced noise is comparable to the AC level.

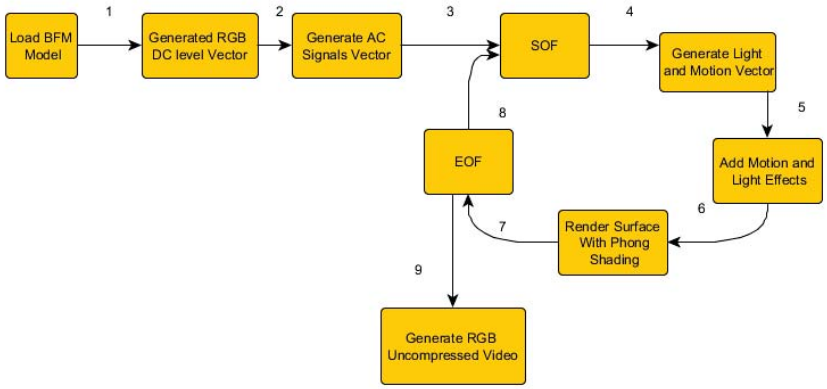

Figure 3.17: steps in generation of the complete model

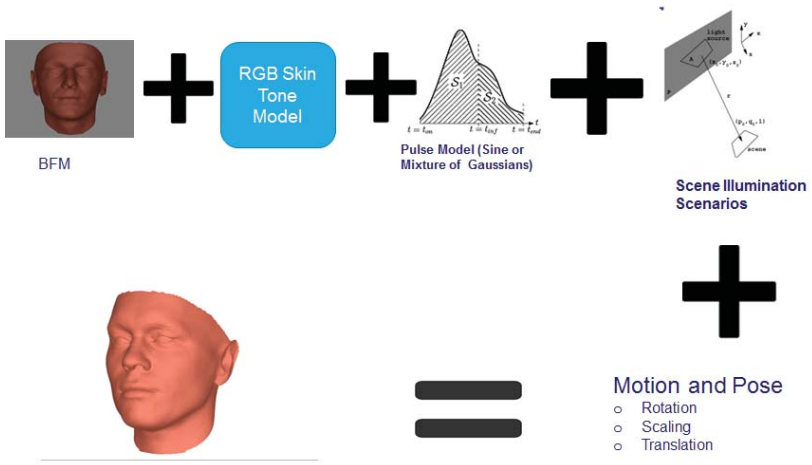

Figure 3.18: Construction of the complete model takes the BFM to which a suitable RGB skin tone is added. A pulse model is added depending to the corresponding skin tone. The rendering of each individual frame depends on each illumination condition for the model. Motion and pose is added to generate realistic motion which may occur during experimentation like head rotation, scaling and translation

3.8 Model Verification

The developed synthetic models are verified using the CHROM and PBV heart extraction algorithms. Sample signal and frequency spectrum using a static ROI are shown in Fig. 3.19 and Fig. 3.20.

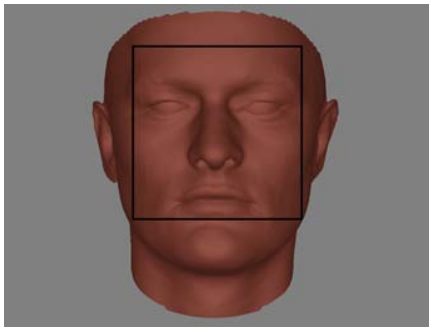

Figure 3.19: ROI

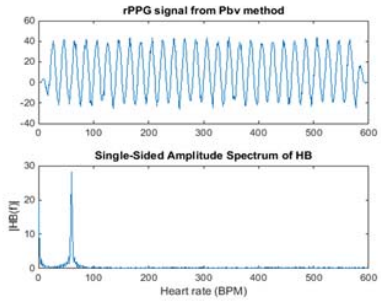

Figure 3.20: Pulse signal and frequency spectrum

3.9 Graphical User Interface (GUI)

This tool is developed as an aid in generation of video sequences with different attributes. To render video and image sequences with different attributes discussed, a prototype GUI implemented in was developed, shown in Fig. 3.21. The major functionalities of the GUI include: setting skin types according to the Fitzpatrick scale, setting illumination intensity, point source position and source color, setting the pulse signal amplitude and periodicity, simulating head motion i.e. translation, rotation, scaling etc. and finally simulating background lighting effects including setting flickering lights.

3.10 Model Limitations

The model implements a simplified version of the pulse signal from the human face. In a realistic situation, the heart signal amplitude varies both spatially and temporally. It is also noteworthy that no noise is added which makes the modelling on the real situation less perfect.

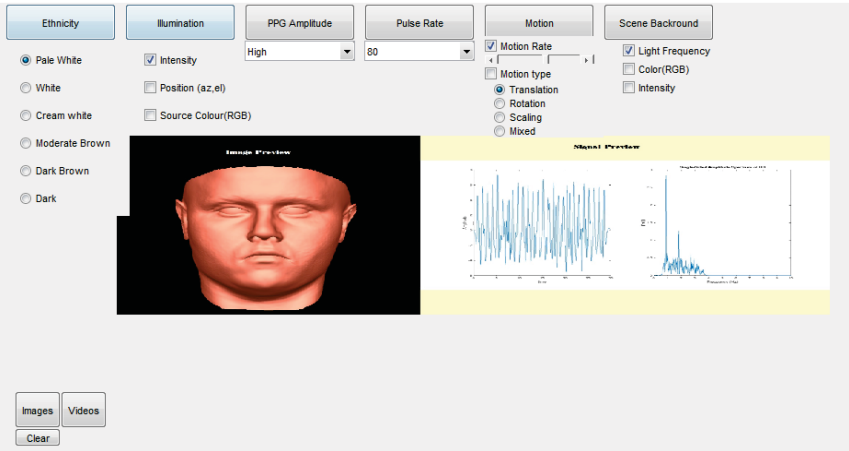

Figure 3.21: GUI prototype

3.11 Summary

In this chapter, we have presented the methods for the 3D generation of the synthetic face, the skin color model used, the AC pulse model and finally the illumination models used. Finally, the steps taken to generate a movie with a PPG signal are explained.

Chapter 4

The Automatic Gain Tuning Algorithm

4.1 Introduction

In this chapter, the concept of camera gains, clipping and the effect thereof on the pulse signal extraction from videos is explained. The steps of the Automatic Gain Tuning (AGT) algorithm developed by Papageorgiou, a masters student at Philips research are also detailed. Based on initial benchmarking results, the possible areas of improvement of the AGT are identified and implemented.

4.2 Camera Gains and Clipping

In digital imaging, a voltage proportional to the amount of incident light is output by a sensor embedded in a camera. To increase image brightness and contrast, this signal can be amplified by an a gain factor before the digitizing process [33]. Depending on the sensor type, a global gain value for all pixels(master gain) or a separate gain value for each color(RGB gain) can be set.

A gain factor is the magnitude of amplification that a given system will produce. It refers to the conversion factor between electrons (e-) recorded by a camera sensor and the number of digital counts. It is expressed as the number of electrons that get converted into a digital number, or electrons per ADU (e-/ADU), i.e. a gain factor of 1.8 e-/count means that the camera will produce 1 count for every 1.8 recorded electron.

Increasing gains increases the brightness of an image, it is reasonable to assume that this magnifies the minute colour changes due to blood volume changes. Most cameras in use today are 8-bit cameras[pixels are represented in the range 0-255]. Therefore, depending on the ambient light conditions, increasing or decreasing gains will increase or decrease the pixel intensity and thus some pixels may fall out of the this range. This phenomenon is called clipping (black clipping for pixels below zero or white clipping for pixels above 255). Since clipping is as a result of very low or very high intensity, it can be naturally caused by intensity changes in the illumination source.

It is hypothesized by Papageorgiou [10] that clipping has a negative impact on the quality of the PPG signal because it renders the minute colour changes absent/lost from the clipped pixels. To reduce this effect, camera gains can be used to reduce or increase the intensity in case of white or black clipping respectively. This can recover the minute colour variations due to PPG.

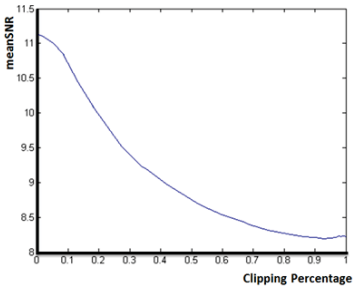

Figure 4.1:
Red:SNR vs clipping %age

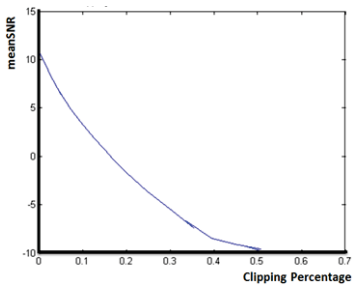

Figure 4.2:
Green:SNR vs clipping %age

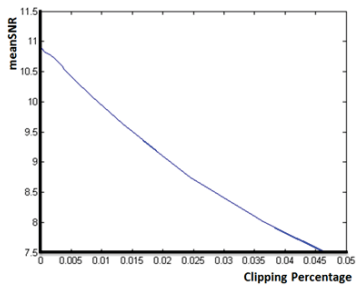

Figure 4.3:
Blue:SNR vs clipping %age

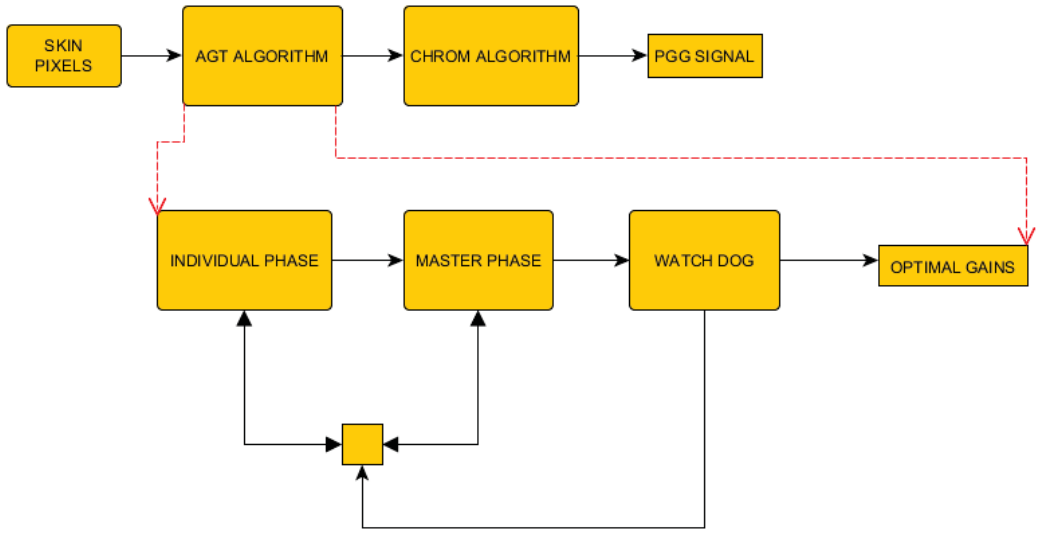

Figure 4.4: AGT algorithm

4.3 Effect of clipping on SNR

To prove the hypothesis, Papageorgiou measures the signal to noise ratio (SNR) metric for different levels of clipping in the different channels, red Fig. 4.1, blue Fig. 4.3 and green Fig. 4.2. In Fig. 4.1, the results for increasing the red gain from 0 to 100, while keeping the green and blue to 0 are given. The mean SNR for each experiment against the white clipping percentage is plotted. The x-axis corresponds to [0 - 100 %] clipping, while the y axis corresponds to the mean SNR value. It is clear that once clipping occurs, the SNR drops. This observation is consistent with the blue and green channels respectively. It is concluded that clipping reduces the quality of the PPG signal. Therefore, setting optimal camera gains can reduce the level of clipping and concurrently amplify the minute colour changes for the PPG signal extraction. The AGT algorithm is elaborated further in the next section.

4.4 The AGT Algorithm

Fig 4.4 visualizes the basic structure of Papageorgiou's AGT algorithm. The AGT runs in real real time for each frame, to predict the optimum gains, before the heart extraction algorithm, the CHROM algorithm for this example. The AGT consists of two major phases; the **1) Individual**

phase where each channels gains are adjusted individually and the **2) Master phase** where all RGB channel gains are adjusted by equal margins and a minor phase the **Idle phase** where no gain changes are made. A watchdog component is added to monitor the change between the 2 major phases.

4.4.1 Individual Phase

In the individual phase, the gains are adjusted individually. This is performed in a few iterations before the master phase because repetitive gain level change would introduce unnecessary noise and potentially corrupt the pulse signal which is constructed from the chrominance signals in the CHROM algorithm. Optimization of each gain initially does not individually affect the CHROM algorithm because it is brightness and hue invariant as long as there is no clipping.

$$\text{White Clip}_{ch} = \frac{\sum_{i=1}^{N}\sum_{i=1}^{M} val(i,j)_{ch} > \text{White Clip Thres}}{\text{Total pixels}} \tag{4.1}$$

$$\text{BlackClip}_{ch} = \frac{\sum_{i=1}^{N}\sum_{i=1}^{M} < \text{BlackClip Thres}}{\text{Total Pixels}} \tag{4.2}$$

$$\text{Total Clip}_{ch} = \text{White Clip}_{ch} + \text{Black Clip}_{ch} \tag{4.3}$$

Where N, M =total number of rows and columns of the captured frame, $val(i,j)_{ch}$, the value of each pixel at the i^{th} and j^{th} column on the ch channel, the White Clip Thres = 245 and Black_Clip_Thres = 10.

The clipping percentage in each channel is calculated according to equations4.1 and 4.2. Using the current gains, the hypothetical clipping at all the different gain levels [0-100] is calculated according to the equation 4.4. The gain factor is an estimated gain multiplier for all the gain levels. Equation 4.4 gives a clipping percentage at all the gain levels. The algorithm selects the maximum gain with the minimum clipping. Gain prediction is subject to estimation errors, therefore gain estimation in the first phase is performed iteratively for 30 frames while checking with the actual clipping before entering the master phase.

$$\text{Hypothetical ClipGain} = \text{k} = \frac{\text{Actual Clipping}}{\text{Gain Factor}} * \text{Gain Factor (Gain = k).} \tag{4.4}$$

4.4.2 Master Phase

In the master phase, the hypothetical clipping is estimated following the same approach in the individual phase. Adjusting all the gains is equivalent to adjusting the brightness level. The white and black clip percentages are calculated according to the equations 4.5 and 4.6. In this phase the hypothetical clipping in all the different channels are not calculated. In the master phase, it is assumed that there are no large variations, therefore a spread threshold = 10 is set with in which an optimal gain is set.

$$\text{Total White Clip} = \mu(\text{White Clip}_R + \text{White Clip}_G + \text{White Clip}_B) \tag{4.5}$$

$$\text{Total Black Clip} = \mu(\text{Black Clip}_R + \text{Black Clip}_G + \text{Black Clip}_B) \tag{4.6}$$

4.4.3 Idle Phase

In this phase, no gain optimization is performed. It is based on the assumption that if there are no changes in the master phase after a given number of iterations, then there is no need to perform expensive calculations.

4.4.4 Watchdog

The watchdog module triggers the change from the masterphase back to the individual phase in case of a monochromatic light variation. In such a case, there is need to optimize the individual colour channel to counter for the changes. The pbv vector introduced in chapter 2 and shown in equation 4.7 is used to monitor for colour changes in the RGB colour space. A change in colour leads to a change in the angle of the vector in equation 4.8. If the angle is larger than an empirically derived threshold (0.015°), then the algorithm triggers a roll back to the individual phase.

$$\vec{V}_{fr} = \frac{[\mu(R), \mu(G), \mu(B)]}{norm[\mu(R), \mu(G), \mu(B)]} \tag{4.7}$$

$$Angle = ArcCos\left(\frac{\vec{V}_{frcur}.\vec{V}_{frprev}}{||V_{frcur}|| \cdot ||V_{frprev}||}\right) \tag{4.8}$$

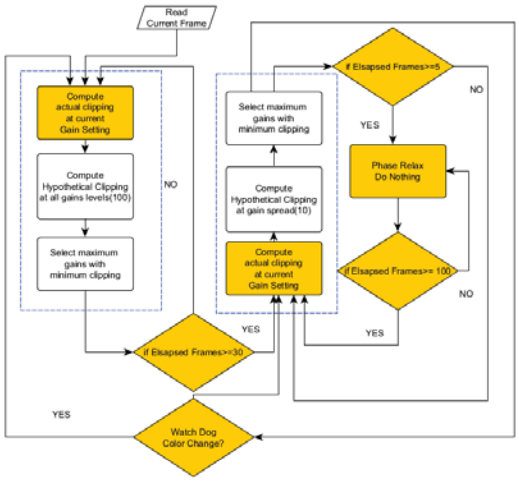

Figure 4.5: AGT by Papageorgiou

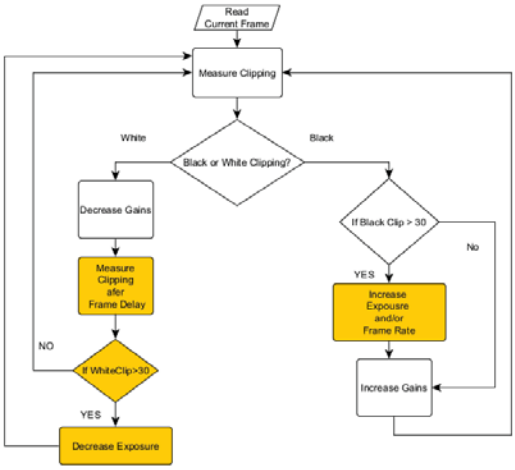

Figure 4.6: AGT with Exposure Adjustments

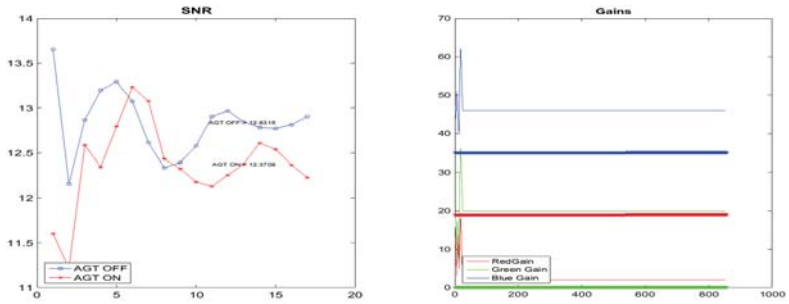

Figure 4.7: Benchmark

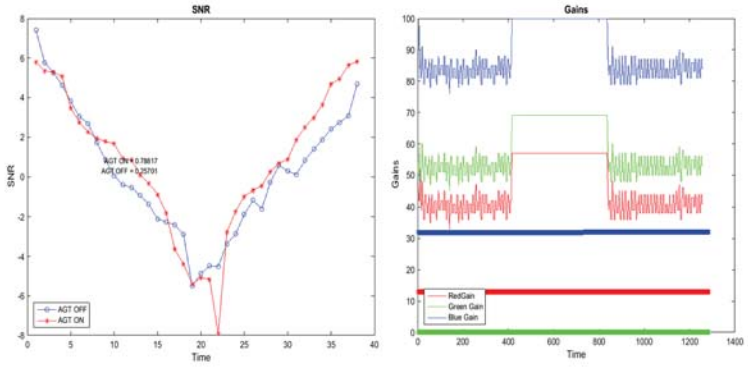

Figure 4.8: Black clipping

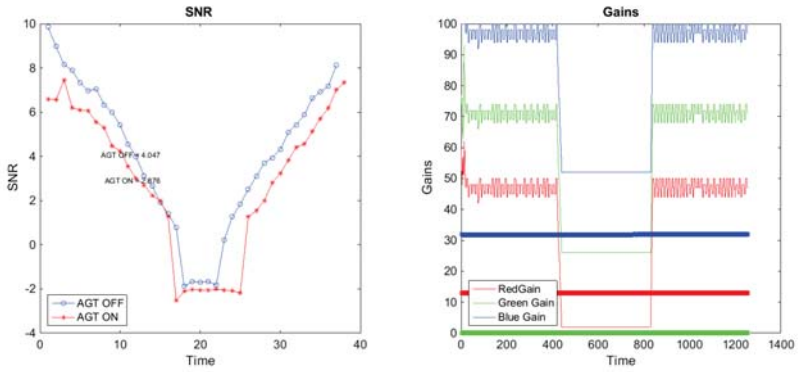

Figure 4.9: White clipping

4.5 Preliminary benchmarking

To evaluate the performance of the AGT algorithm under different illumination scenarios, three synthetic videos are initially developed.

1. A benchmark video: With zero clipping throughout the measuring period.
2. White clip video: With 0% clipping in the first 30 seconds, ≈ 60% white clipping in the following 30 seconds and finally 0% clipping in the last 30 seconds
3. Black clip video: With 0% clipping in the first 30 seconds, ≈ 60% black clipping in the following 30 seconds and finally 0% clipping in the last 30 seconds).

The set up and experiment procedure is detailed in chapter 5. Preliminary results of SNR and the temporal gain settings are shown in figures 4.7, 4.8 and 4.9. These results reveal a major AGT drawback, we shall refer to as the gain trap.

4.6 Gain trap

In Fig. 4.7, there is no deference between the "agt-off" and "agt-on" implementations. The gains are stable throughout the entire recording period because there is no clipping in the video sequence and no need for Auto-gain tuning. In Fig. 4.8, the SNR drops during the clipping interval. Investigation of the gains shows that the blue gain is set to the maximum possible value while the other gains are not at their maximum. The gains cannot be adjusted anymore because there are no individual gain changes in the master phase. This traps the rest of the gains. As a result of the gain trap, the clipping cannot be entirely reduced to zero.

In Fig 4.9, the red gain is trapped to 0%. As a result of this phenomenon, the blue and green gain cannot be adjusted any longer. Thus, the clipping percentage cannot be entirely reduced to zero.

It is also clear that in both scenarios the gain oscillations are periodic. These oscillations are as a result of the minute light changes which are as a result of setting the amplitude of PPG model from 0.001% to 0.05% of the DC level. This leads to increase or decrease of pixels which fall out of the RGB range[0-255]. The oscillation are removed by implementing a higher gain change threshold explained in the improvements in section 4.7.

From the observations, it clear that once in the masterphase, the AGT does not optimize the gains effectively because it cannot trigger a roll back to the individual phase to reset to other gains. It is also evident that depending on the ambient lighting, setting gains alone cannot effectively reduce the clipping percentage. In section , several suggestions are presented to reduce/eliminate the gain trap.

4.7 Eliminating the gain trap

To eliminate the gain trap, the watchdog component which monitors and triggers changes between the phases is modified, the design choices for the delays (duration of each phase) are also modified, and finally adaptive exposure is suggested as shown in the flow diagrams Fig. 4.5 and Fig. 4.6. The major improvements are summarized into 4 major areas,

1. **Adaptive Exposure.** Increasing or decreasing gains cannot eliminate clipping in certain cases of extreme clipping. it is also noted from device manufacturers, that increasing gains also increases/amplifies the noise levels. Therefore, a new improvement shown in Fig. 4.6 is proposed.

   In the adaptive exposure strategy, in case of black clipping, an increase in exposure is first effected before allowing a dynamic gain setting. This is based on the premise that in case of black clipping, the AGT will increase all gains to 100 which is not desirable because this also amplifies the noise. Increasing the exposure also allows for a more dynamic gain change instead of clipping all gains to one value. In case the maximum exposure is reached i.e (1/Frame rate), a lower frame rate is set(decrease by one) allowing for a longer exposure setting. Decrease of the frame rate should also ensure that the lower limit of the Nyquist sampling frequency for the heart rate is not exceeded. The heart rate is expected to lie between 0.5 and 3.5 Hz, therefore the lowest frame rate is not permitted below 7 Frames per second(7 FPS which is twice 3.5 Hz)

   In case of white clipping, a gain setting is effected before effecting an exposure setting. This is based on the evidence that extremely high gain settings also lead to noise amplification but setting lower optimum gains is desirable. Therefore setting gains takes precedence over

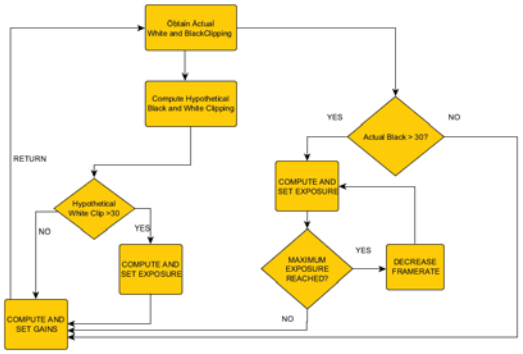

Figure 4.10: Automatic gain and Exposure Tuning Algorithm

Figure 4.11: Additional Master Phase Watchdog

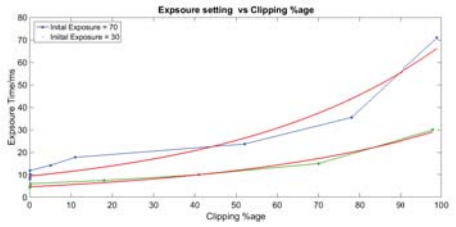

Figure 4.12: Exposure vs clipping

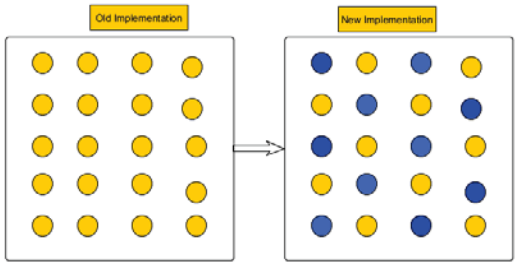

Figure 4.13: Pixelsubsampling

setting exposure. The adaptive exposure methods are illustrated in the flow diagram in Fig 4.10.

The exact profile guiding the selection of the exposure times is shown in Fig 4.12. This graph is based on experiments performed to determine the clipping percentages at different exposure settings. In the first experiment, the clipping percentage is initially set to 98% , at an initial exposure of 71 ms (maximum for frame rate of 14 FPS). It is then repeated with the initial clipping, set using a shutter set to 98 % and an initial exposure of 30 ms. The exposure is then divided by 1,2,3,45... until the clipping achieved is $\approx 0\%$. It is noted that in both scenarios the exposure settings achieve zero clipping at level 5. The exposure vs clipping approximately follows the profile given by;

$$\text{Expsoure} = A \cdot \exp^{\text{B.Clip}\%age}, \quad A = \frac{\text{Exposure at 100\%clipping}}{5}, \tag{4.9}$$

Where A is the value of exposure setting at zero clipping. From experiment $B \approx 0.02$.

2. **Introduce Master watchdog**
   There is no trigger in the master phase to enforce a roll back to the individual phase in case of a significant brightness changes which causes gain trapping. Additional logic is added to watchdog as shown in Fig 4.11. The additional watchdog monitors for a large brightness change by comparing the current and previous clipping levels. If this change is a greater than a predefined threshold (20%), the watchdog triggers a roll back to the individual phase.

3. **Eliminate Idle Phase**
   From the AGT implementation, in Fig 4.5, the algorithm does not effect any gain change

during the idle phase. For 100 iterations at 14 FPS frame rate, this is equivalent to up to 7 seconds without gain optimization. Light changes can occur randomly and therefore this phase is completely eliminated from the algorithm.

4. **Adjust Thresholds:**
   The effective gain threshold is increased to eliminate gain changes due to minute light variations. Trigger from the master phase to any other phase is as a result of light color change or large intensity variation.

4.7.1 Clipping Percentage and Gain Freedom

Fig. 4.14 to Fig. 4.18, show results of experiments performed with clipping percentages between 10 % white and 10% black clipping. The purpose of the experiment is to motivate the fact that with such low levels of clipping, AGT is not desirable. It can be seen that in case of black clipping, the implementations of AGT(AGT Papageorgiou and AGT from Richard) have lower SNR and the average clipping percentage is not entirely reduced to zero in all cases. This stems from the fact that all experiments are performed with an initial gain set up of [RGB=19,0,35] obtained from benchmarking of default IDS camera gains at start up. These initial gains give a smaller dynamic range in case of black clipping where gains are reduced towards zero. In case of black clipping, the gain freedom is higher because gains are increased towards 100. The observations suggest two main conclusions 1) The performance in black clip is superior to the white clip due to the gain freedom, 2) The AGT is not necessary for low clipping percentages(below 10 %).

4.7.2 Pixel Sub-sampling

We have seen that a small change in clipping percentages does not have large effects on the SNR. The AGT is also a computational intensive algorithm because of the computations performed in measuring the hypothetical clipping and the actual clipping. Instead of measuring all pixels in the ROI, we can sub sample the pixels i.e skipping rows or columns or skipping a pixel shown in Fig 4.13 without greatly affecting the clipping percentage predictions.

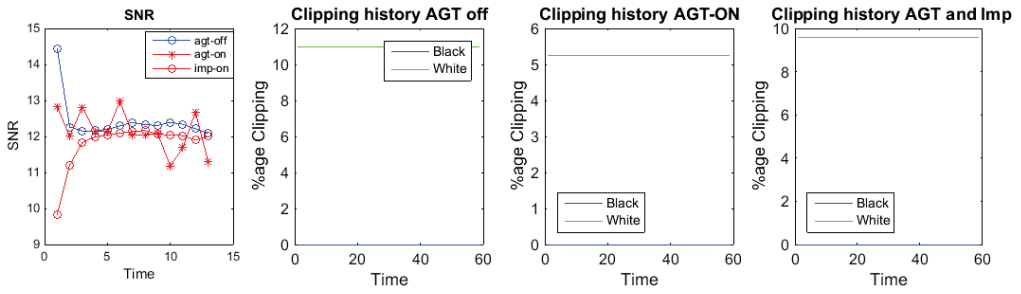

Figure 4.14: SNR and Clipping percentage 10% white clipping

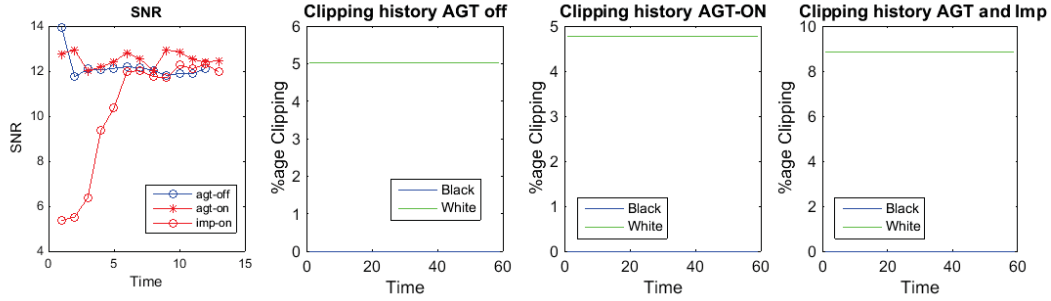

Figure 4.15: SNR and Clipping percentage 5% white clipping

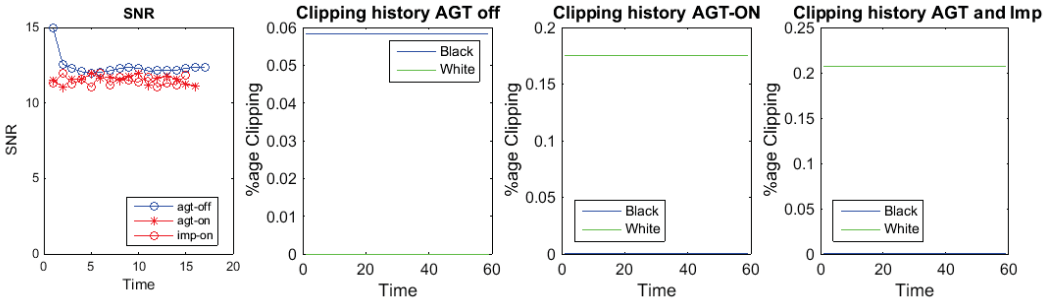

Figure 4.16: SNR and Clipping percentage 0% white and black clipping

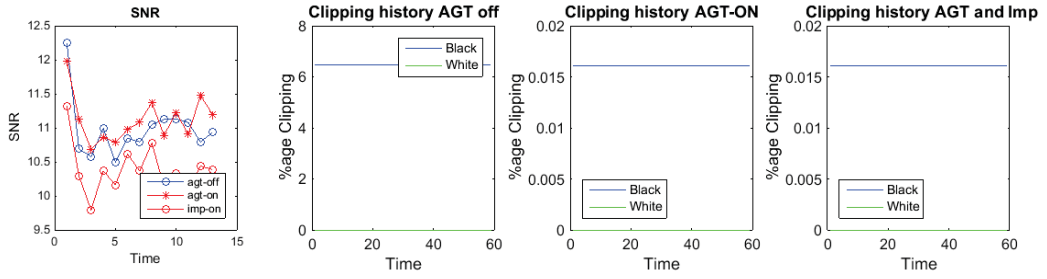

Figure 4.17: SNR and Clipping percentage 5% black clipping

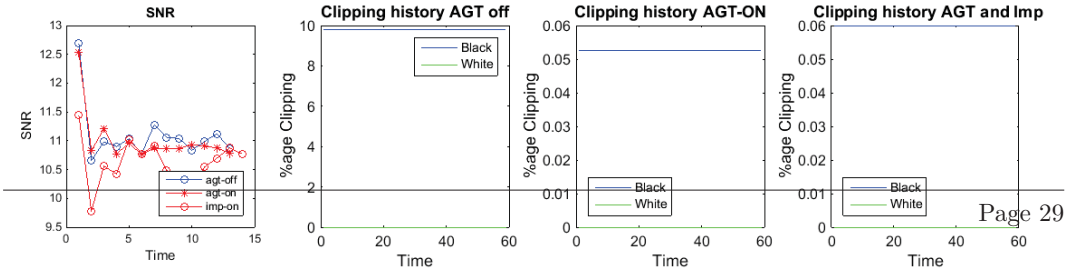

Figure 4.18: SNR and clipping percentage 10% black clipping

4.8 Summary

In this chapter, the concept of clipping and camera gains has been presented. It has been shown that a high level of clipping (black or white) affects SNR while low level of clipping have minute effect on SNR. The AGT algorithm reduces clipping by adjusting the gains. Camera gains can not be adjusted indefinitely [only 1-100 gains] and additionally the current AGT implementation creates a gain trap. To improve the current AGT, adaptive exposure and an additional master phase watchdog is suggested. The current implementation is a computational-resource intensive implementation. The computation overheads can be greatly reduced by sub-sampling, increasing effective gain thresholds and avoiding AGT completely for clipping percentages below 10 %.

Chapter 5

Experimentation and Results

5.1 Introduction

In this chapter, the experiments to validate the different AGT algorithm implementations are discussed. These include; clipping based on exposure, real life scenarios (movement of cloud cover, glare and dynamic room lighting) and finally white and monochromatic flicker. The evaluation of the experiments is based on SNR and clipping percentage. The different AGT algorithm implementations are referred to as;

1. **agt-off**: Experiments with no automatic gain or exposure tuning.
2. **agt-on**: Experiments with Andreas' AGT implementation
3. **exp-on**: Experiments with Andreas' AGT implementation and automatic exposure, and
4. **imp-on**: Experiments with Andreas' AGT and suggested changes based on the initial benchmarking described in chapter 4

5.2 Experiment Procedure

Synthetic videos developed using the approach in chapter 3 are displayed on an LED monitor using a suitable high bit rate video player(VLC) because the data rate from such a video i.e uncompressed AVI is high($\approx 364kbps$). The synthetic video properties are listed in Table 5.1. The experiment set up is as shown in Fig. 5.1. A facial ROI of (230x250) pixels is used for pulse signal extraction implemented using the CHROM algorithm described in chapter 2. Videos are recorded for a minimum of 60 seconds, specific experiment timings depend on the different experiments. To extract the pulse-rate from pulse signals, peak detection in the frequency domain using a 512 point FFT (for stationary scenarios) or 250 point FFT (for dynamic scenarios) using a 20 frames moving window is implemented. For each experiment, SNR, temporal gains and clipping history is measured.

Table 5.1: Synthetic video specifications

Frame width	1000
Frame height	759
Data rate	$364,329kbps$
Total bit rate	$364,329kbps$
Frame rate	$20frames/second$

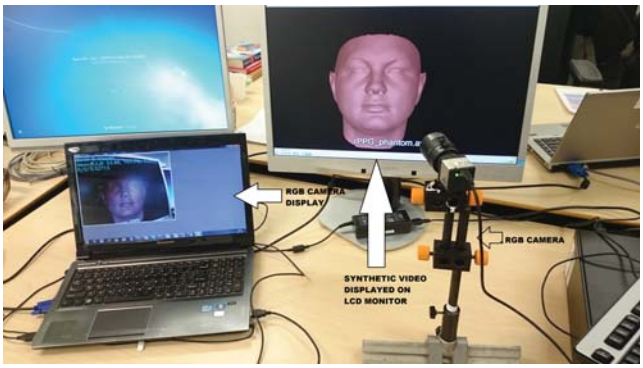

Figure 5.1: Using RGB camera to capture a video displayed on the LCD monitor for PPG

5.3 Measurement Equipment

To display the synthetic video, a desktop computer of with Intel(R) core(TM) i5-4570 CPU @ 3.20 GHz and RAM 4GB and LED monitor 1680x1050 @ 60 Hz) is used. The Ueye IDS CCD camera, model UI-222x/UI-622x is used, The camera auto functions: i) Auto gain tuning, ii) Auto white balance, iii) auto shutter and iv) Color correction are disabled to prevent any camera color corrections. The Frame rate is set to 14 FPS to reduce processing overheads at higher frame rates. The initial static gains are set to [RGB = 19 0 35] based on benchmarking steady state camera gains under zero clipping ambient lighting conditions.

5.4 Measurement Metrics

The first metric used is the percentage of clipped pixels in the ROI of interest. The measurement of this metric is described in chapter 4. The second metric is the SNR which is computed as the ratio of the energy around the fundamental frequency and the energy around the first two harmonics contained in the rest of the spectrum within a frequency range of (0.5 Hz to 3.5 Hz) given as

$$SNR = 10\log\left[\frac{\sum_{f=0.5}^{f=3.5}(\vec{U}_t(f)\vec{\hat{S}}(f)^2}{\sum_{f=0.5}^{f=3.5}(\vec{(}1-U_t(f))\vec{\hat{S}}(f)^2}\right] \tag{5.1}$$

Where $\vec{\hat{S}}(f)$ is the spectrum of the pulse-signal $\vec{S}(f)$, f is the frequency(Hz), and $\vec{U}_t(f)$ is the binary template as shown in Fig. 5.2

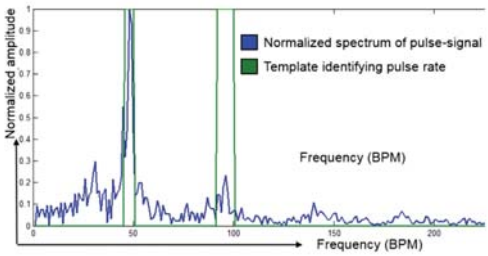

Figure 5.2: SNR template

5.5 Exposure experiments

This set of experiments is based on varying the exposure/shutter to achieve different levels of white (80%, 40%, 20 %, 0%) and black(80%, 40%, 20 %, 0%) clipping. The performance of the different AGT implementations is compared based on SNR and clipping percentages. All gain charts referred to in this section can be found in appendix A.

1. 80 % White Clipping. From Fig. 5.3, the average SNR of agt-on (14.79 dB) and agt-imp (15.23 dB) is greater than exp-on(10.54 dB) and agt-off(11.98 dB). Reducing the clipping percentage increases the SNR. Exp-on achieves 0% clipping while imp-on and agt-on achieve 40% and zero gains in all channels. Exp-on achieves an unexpectedly low SNR because in the process of reducing exposure to achieve 0% clipping, the agt algorithm measures a temporary clipping percentage increasing gains and thus amplifying the sensor noise too.

2. 40 % White Clipping. The average SNR of agt-on(12.73 dB), agt-imp (12.40 dB) agt-off (12.30 dB) and exp-on (11.66 dB). Exp-on achieves 0% clipping but has a worse SNR because in the process of predicting the exposure to achieve zero clipping, the gains are increased, thus amplifying the noise too. It is worth noting that the signal response for 40 % and 80 % is largely due to the first harmonic because of the distortions to the sine wave at these levels of clipping. These distortions are shown in Appendix chapter:allother.

3. 20 % White clipping. All AGT implementations; agt-imp (12.28 dB), agt-on (12.46) and exp-on (11.08 dB) achieve better SNR compared to agt-off (7.48 dB). This is a result of reduction in amount of clipped pixels to $\approx 0\%$. This is different from the above observations because, there is no sine wave signal distortion and only the fundamental frequency is recorded.

4. 0% There are subtle differences between the different AGT implementations and agt-off. As already noted in the initial benchmarking, in chapter 4, gain changes around 0% clipping may lead to a lower SNR or are not beneficial at all.

5. 20% Black Clipping. There are minute differences in SNR measurements although all agt implementations achieve zero clipping. The exp-on has slightly higher average SNR compared to all other implementations because it achieves zero clipping without adjusting all gains to the maximum gain level. Increasing exposure also increases the amount of light making the minute color changes due to blood volume flow more visible and detectable compared. With less lighting, noise is more dominant.

6. 40 % Black Clipping: agt-exp(11.94 dB) achieves a higher average SNR compared to the other implementation, because zero clipping is achieved at a slightly higher intensity since the exposure is increased. All agt achieve zero clipping.

7. 80 % Black Clipping; The agt-exp (12.14 dB) achieves a higher average SNR compared to the other implementations because it achieves zero clipping with lower gain settings compared to agt-on (10.89 dB) and agt-imp(12.14 dB). The agt-on and agt-imp does not achieve total

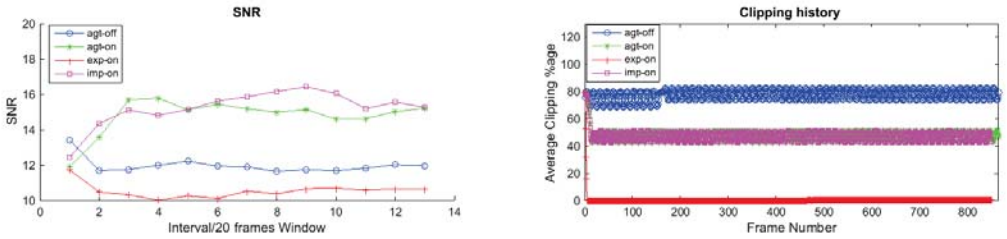

Figure 5.3: 80% white clipping

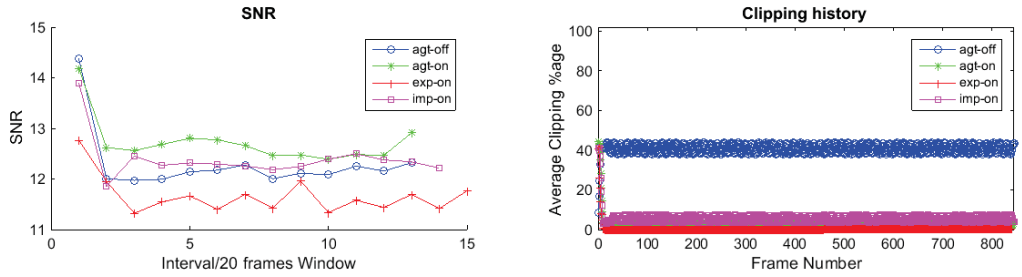

Figure 5.4: 40% white clipping

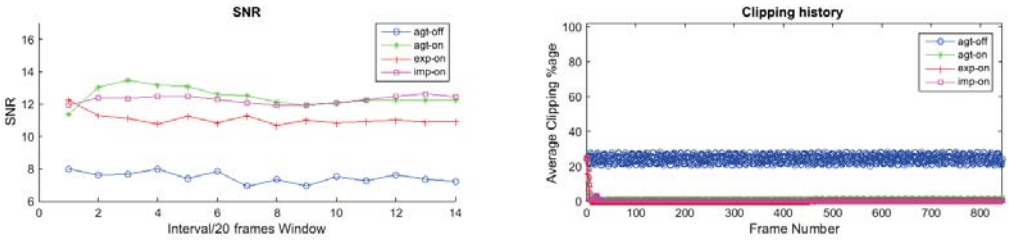

Figure 5.5: 20% white clipping

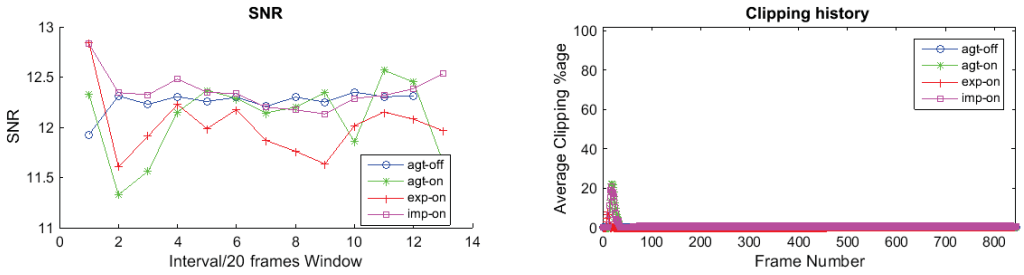

Figure 5.6: 0% Clipping

0% clipping. With exp-on, the light-intensity is increased thus amplifying the minute color changes due to blood flow.

From the experiments above, the AGT has subtle improvements in SNR with high (above 80 %) white and black clipping. At lower levels of clipping (Below 20 %) gain tuning is not beneficial. The rPPG pulse signal amplitude is more dependent on the amount of light falling on the ROI and less dependent on a constant clipping percentage. This is supported by i) The black clip experiments where increasing exposure time(light) greatly increases the SNR compared

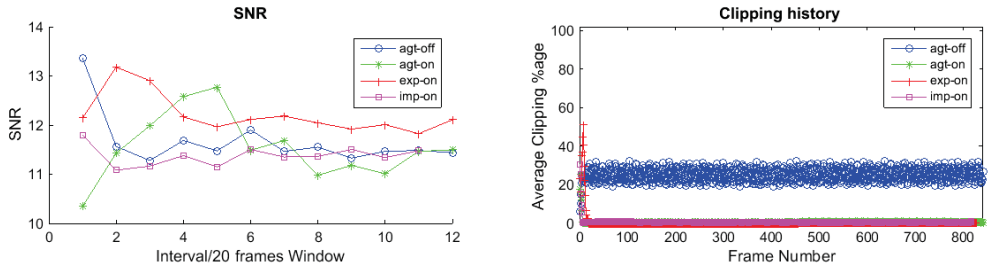

Figure 5.7: 20% black clipping

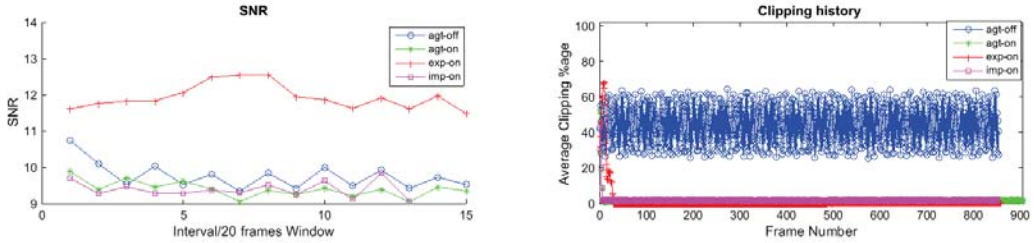

Figure 5.8: 40% black clipping

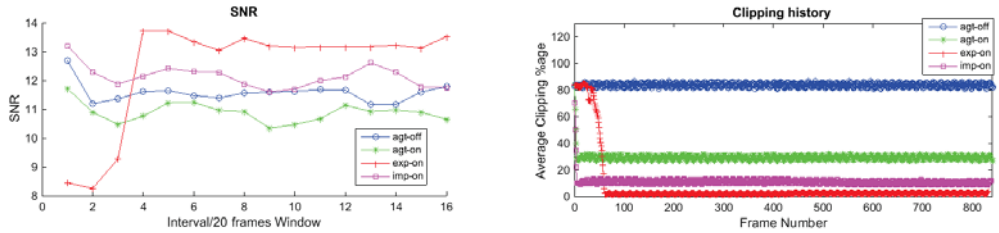

Figure 5.9: 80% black clipping

to increasing gains which have less effect. ii) The white clip experiments where changing gains (i.e at 40 % clipping) has no effect on SNR because the high light intensity is enough to amplify minute color changes due to blood volume change.

5.6 Dynamic Lighting

In real experiments, dynamic lighting manifests itself in several forms;

1. In an exposed environment, light changes occur as a result of cloud cover movement.

2. In an a room environment, light changes occur as result of toggling room lighting, opening or closing doors/windows and movement of a highly reflective object, e.t.c. These experiments are investigated using three main synthetic videos whose results are representative of dynamic lighting in general. Sample frames from these experiments are presented in Appendix B.

1. Cloud Cover: PPG experiments taken in the open are subject to light changes that can arise as a result of the cloud cover movement casting shadows or exposing light rays in the experiment scene. Such light changes are non-uniform over the facial regions. A synthetic video of 90 seconds is rendered in which light changes mimicking a cloud movement leading to decrease and increase in light in the middle 30 seconds are mimicked. In Fig. 5.10, dynamic increase and decrease of clipping reduces SNR in all cases. The agt implementations

show no improvements because in a highly dynamic setting, the gain predictions based on hypothetical clipping are applied with a frame delay of 2 frames. This implies that the gain change doesn't affect the right frames which results in low SNR of agt-on and agt-imp in particular. It is also noticeable that the gain change alone doesnot get rid of the clipping, thus the exp-on achieves the desired effect and has a slightly higher average SNR compared to the agt-on and agt-imp. The agt-imp performs worse because dynamically changing gains results into distortions.

2. Glare: Glare is a strong and dazzling light. Glare can be as a result of a primary light source or reflected light of very high intensity like the sun or focussed light from sources i.e headlights and torches. Glare leads to white clipping. In Fig. 5.11, agt-imp has the lowest average SNR because the predictions lead to amplification of the clipping. This is because allowing a free play of all gains in a dynamic situation can causes an off prediction.

3. Toggling of overhead room lighting: Switching on room lights leads to an sudden increase in light intensity while switching off lights leads to a decrease in light intensity. In both cases white or black clipping is likely to occur. In Fig. 5.12, it is evident that the gradual increase in light intensity leads to increase in clipping which leads to a decrease in SNR. The agt implementations lead to a decrease in clipping and an increase in SNR. The agt-exp achieves zero clipping and thus the highest average SNR after the disruptive lighting event. This is different from the cloud cover and glare experiment because the light changes are gradual and predictable.

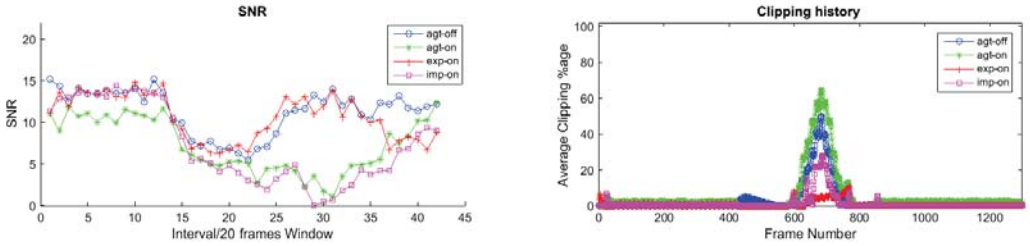

Figure 5.10: Cloud cover

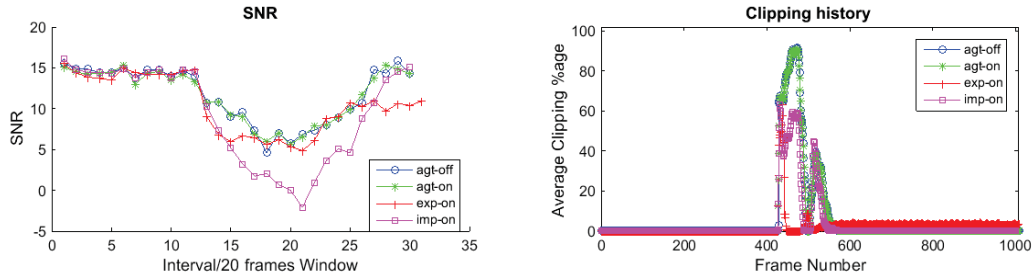

Figure 5.11: Glare

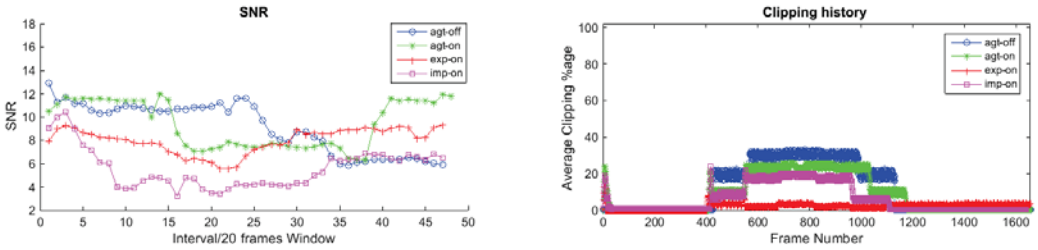

Figure 5.12: Dynamic Room Lighting

5.7 Periodic Light and Monochromatic Light changes

Light changes can also occur as a result of periodicity in lights causing flickering. This flickering can occur directly in the face or in the background. In addition, changes in lighting can occur in only one color channel. These lighting conditions are investigated in a three experiments;

1. Face flicker: In this experiment, facial illumination frequency is varied from 0.05 Hz to 0.50 Hz in a single video to investigate the performance of the agt under different frequencies. In Fig. 5.13, flickering decreases the SNR because of clipping (the flicker causes a clip of $\approx 30\%$). The general trend observed shows that below 0.5 Hz, the agt-off is lower than agt-on, therefore gain tuning corrects for the changes in illumination. Beyond 0.5 Hz, at a frame rate of 14 FPS, there are 2 or more clipping changes in one frame therefore, clipping based prediction cannot help to estimate the correct gains even when the clipping percentage is the same.

2. Background Flicker: In this experiment, the background flicker frequency is set to 0.1 Hz. From Fig. 5.14, there is no large difference in SNR values for all implementations because in the implementation of our model, a flicker in the background has no effect on the facial ROI because only the primary reflection from the background is modelled. Therefore, since no clipping occurs in the ROI during the entire experiment time, there is no gain change

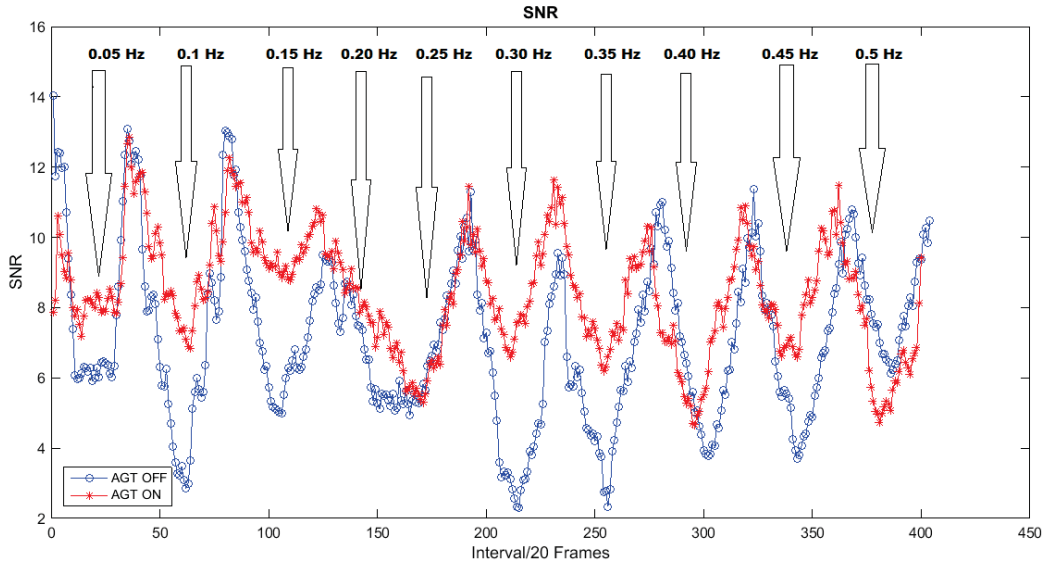

Figure 5.13: Increasing Flicker Frequencies from 0.05 Hz to 0.5 Hz

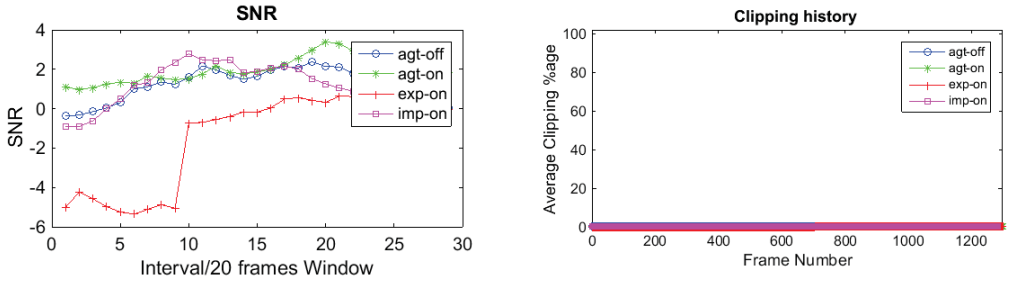

Figure 5.14: Background Flicker

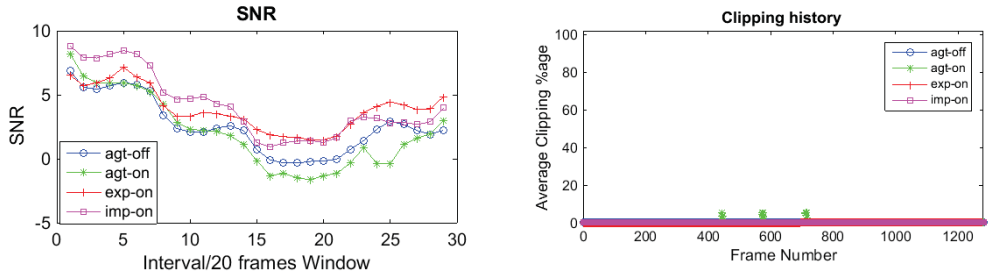

Figure 5.15: Monochromatic Facial Flicker

effected. In a realistic situation, background flicker may cause secondary reflections from surfaces which will affect the facial ROI pixel values.

3. Monochromatic face flicker. A low frequency blue light of 0.05 Hz is used to simulate a scenario in which a light color change occurs. From Fig. 5.15, the SNR in all implementations decreases as a result of a facial flicker despite the 0% lipping percentage in ROI. Since the

agt-performs a cross-check on the color changes, an update of gains is performed. This update nonetheless has no effect on SNR.

Illustrations of these scenarios are shown in the Appendix B

5.8 Validation with real-world subjects

To draw a valid conclusion from the different AGT implementations, the synthetic experiments are replicated in realistic scenarios. From Fig 5.16 to Fig 5.20, we observe that generally, exp-on implementation has the best SNR performance because it reduces clipping to zero by not only adjusting the gains but by also increasing light. At 0 % clipping , gain tuning is not desirable, this is evidenced by the SNR value of agt-off in Fig. 5.18. Gain tuning in all cases has minute gains in terms of SNR if there are low levels of clipping(Below 50 % clipping). Gain tuning is only profitable at high levels of clipping.

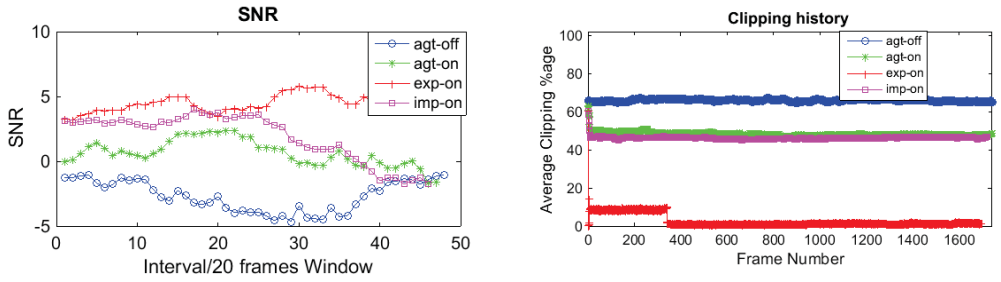

Figure 5.16: 70% white clipping

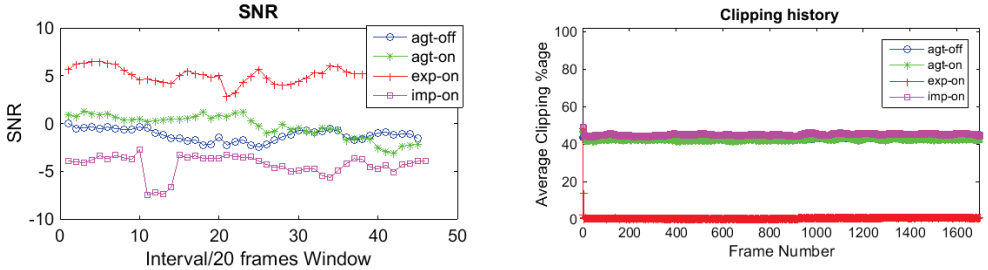

Figure 5.17: 40% white clipping

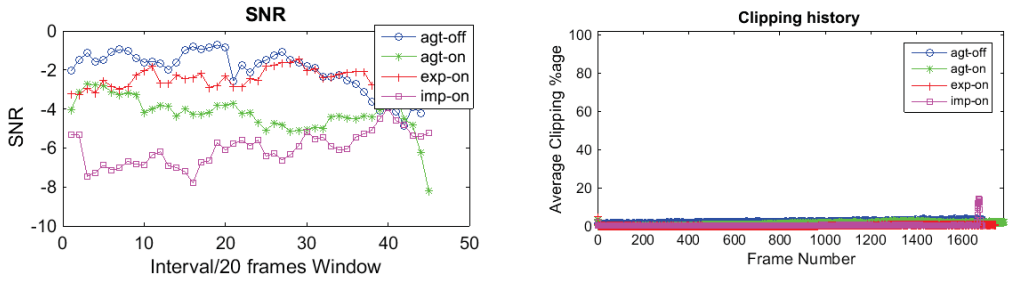

Figure 5.18: 0% Clipping

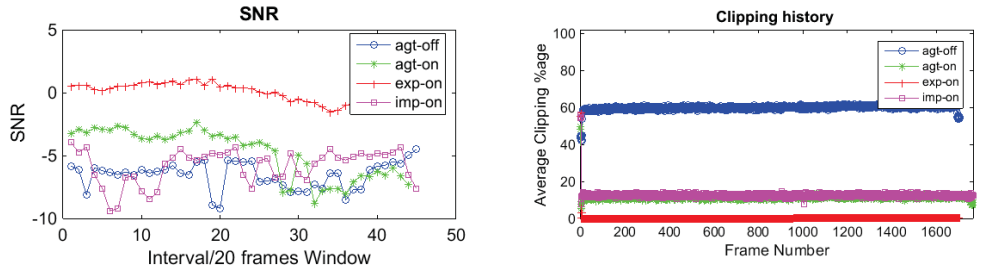

Figure 5.19: 50% Black clipping

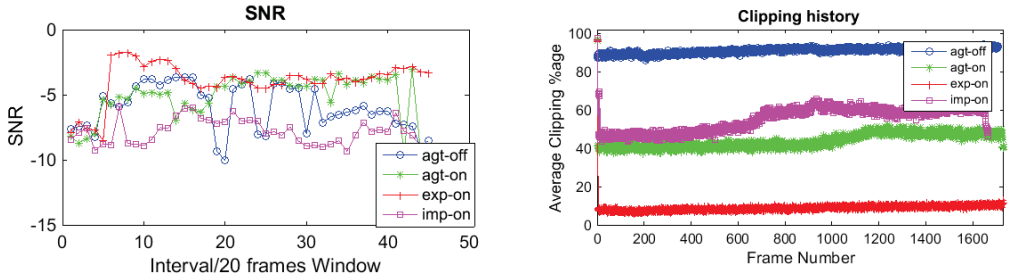

Figure 5.20: 80% Black clipping

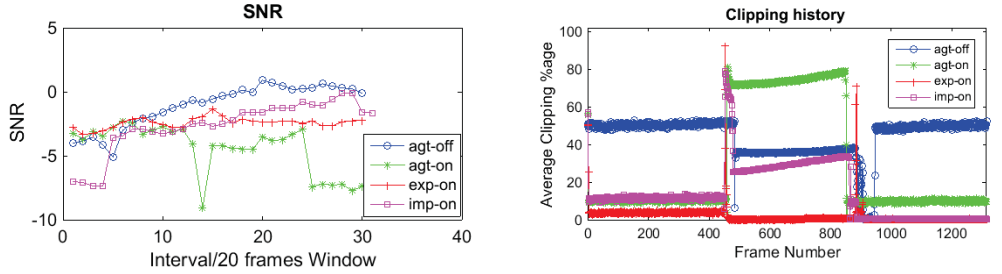

Figure 5.21: Switching on Lights

5.9 Summary

In this chapter, the effect of dynamic lighting, gain and exposure tuning has been investigated by using a series of realistic synthetic videos. It has been experimentally shown that adjusting gains reduces clipping, but as seen in extreme clipping scenarios, this does not reduce the clipping

percentage completely in most cases because gains cannot be adjusted indefinitely (They range between 0 and 100). Adjusting exposure times is inevitable and preferred in reducing the clipping percentages to zero. The effect of gain changes is high if gains have a higher dynamic range in which they can be adjusted i.e, if initially gains are set to zero and the desired effect is an increase, then a gain change is very effective, otherwise it is not. The SNR, in general, shows subtle changes with low clipping irrespective of whether gain tuning is performed or not. Increase in SNR under AGT is more noticeable with a high initial level of clipping. Changing gains individually during the experiment is not desirable because in most cases it leads to a lower SNR as evidenced in our 'improved version' agt-imp. We have also shown through a series of experiments that highly dynamic lighting leads to a decrease in SNR in all scenarios even if gain tuning is applied. Gain tuning my lead to a decrease in SNR in dynamic situations because of gain predictions applied to the wrong frames.

Chapter 6

Conclusions

6.1 Synthetic rPPG Model

In this paper, we have created realistic synthetic models which contain physiological signal using 3D morphable models. It is possible to embed different minute color changes in pixels of 3D models, whose variations contain a specific reproducible frequency mimicking the pulse rate. Using CHROM and PBV, we have shown that the minute color variations in the 3D models can be detected from rendered lossless AVI movies. We have also shown how to add realistic motion and illumination effects to the synthetic models. These models can be used to develop new rPPG algorithms aimed at improving motion robustness without requiring real world experiments.

6.2 Automatic Gain Tuning

A comprehensive benchmark of the Automatic Gain Tuning algorithm developed by a masters student at Philips has been presented. The benchmark mark is based on the synthetic models developed in this research. The aim of the AGT is to reduce black and white clipping. It is hypothesised that this increases the SNR because reducing clipping increases the count of pixels with physiological signals. Based on initial benchmarks of white and black clip videos, it is established tha the effect of gain tuning depends on the ambient light conditions and the initial gains. Gain tuning will have significant effect on clipping if the initial ambient condition for example are; a bright environment and the initial gains are high or a dark environment and the initial gains are low. In both cases, the gain change counteracts the current ambient conditions to reduce clipping. We refer to this phenomenon as gain freedom, the higher the gain freedom the better the performance of the AGT. In its implementation, the AGT does not reduce clipping by a large margin because of a gain trap. This occurs when all gains are changed simultaneously such that the first gain to get to 0 or 100 traps the other gains and thus no further gain change can be made. Improvement are suggested to avoid the gain trap. It is observed that from the suggested improvements, allowing individual gain changes during PPG experiments does not improve the AGT performance. Adjusting camera exposure is preferred and indeed achieves better SNR compared to adjusting individual gains. Finally, very low clipping percentages i.e below 20 % for synthetic videos and below 50% yield only subtke difference in SNR for all agt implementations. Reducing clipping achieves better SNR at high clipping levels, otherwise only subtle differences occur.

6.3 Future work

Future work should focus on improving the robustness of the synthetic model to accurately represent the specular and diffuse reflection and implementation of an accurate spatial

representation of the PPG signal amplitude over the different facial regions. The exposure prediction model should be improved empirically.

Bibliography

[1] Markus Huelsbusch and Vladimir Blazek. Contactless mapping of rhythmical phenomena in tissue perfusion using ppgi. In *Medical Imaging 2002*, pages 110–117. International Society for Optics and Photonics, 2002. 1, 2, 4, 7

[2] John Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1, 2007. 1, 4

[3] Puneet Kumar Jain and Anil Kumar Tiwari. Heart monitoring systemsa review. *Computers in biology and medicine*, 54:1–13, 2014. 1

[4] Magdalena Lewandowska, Jacek Ruminski, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcama non-contact method for evaluating cardiac activity. In *Computer Science and Information Systems (FedCSIS), 2011 Federated Conference on*, pages 405–410. IEEE, 2011. 1, 8

[5] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774, 2010. 1, 8

[6] Gerard de Haan and Vincent Jeanne. Robust pulse-rate from chrominance-based rppg. 2013. 1, 2, 8, 9, 17

[7] Gerard de Haan and A van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014. 1, 2, 8, 10

[8] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008. 1

[9] He Liu, Yadong Wang, and Lie Wang. The effect of light conditions on photoplethysmographic image acquisition using a commercial camera. *Translational Engineering in Health and Medicine, IEEE Journal of*, 2:1–11, 2014. 1, 8, 14

[10] Andreas Papageorgiou. Adaptive Gain Tuning for Robust Remote Pulse Rate Monitoring under changing light conditions, 2014. 1, 3, 21

[11] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, pages 187–194. ACM Press/Addison-Wesley Publishing Co., 1999. 2, 4, 11

[12] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *Advanced Video and Signal Based Surveillance, 2009. AVSS'09. Sixth IEEE International Conference On*, pages 296–301. IEEE, 2009. 2, 4, 11

[13] Manuela Lualdi, Ambrogio Colombo, Bruno Farina, Stefano Tomatis, and Renato Marchesini. A phantom with tissue-like optical properties in the visible and near infrared for use in photomedicine. *Lasers in surgery and medicine*, 28(3):237–243, 2001. 4

[14] SoonKee Chung, J-C Bazin, and Inso Kweon. 2d/3d virtual face modeling. In *Image Processing (ICIP), 2011 18th IEEE International Conference on*, pages 1097–1100. IEEE, 2011. 4

[15] WonSook Lee, Jin Gu, and Nadia Magnenat-Thalmann. Generating animatable 3d virtual humans from photographs. In *Computer Graphics Forum*, volume 19, pages 1–10. Wiley Online Library, 2000. 4

[16] Praveen Kakumanu, Sokratis Makrogiannis, and Nikolaos Bourbakis. A survey of skin-color modeling and detection methods. *Pattern recognition*, 40(3):1106–1122, 2007. 4

[17] Nusirwan Anwar bin Abdul Rahman, Kit Chong Wei, and John See. Rgb-h-cbcr skin colour model for human face detection. *Faculty of Information Technology, Multimedia University*, 2007. 4

[18] Jie Yang, Weier Lu, and Alex Waibel. *Skin-color modeling and adaptation*. Springer, 1997. 4

[19] Hayit Greenspan, Jacob Goldberger, and Itay Eshet. Mixture model for face-color modeling and segmentation. *Pattern Recognition Letters*, 22(14):1525–1536, 2001. 4

[20] Ming-Hsuan Yang and Narendra Ahuja. Gaussian mixture model for human skin color and its application in image and video databases. In *Proc. SPIE: Storage and Retrieval for Image and Video Databases VII*, volume 3656, pages 458–466, 1999. 4

[21] Iztok Kramberger. Reflectance measurement system for skin color modeling in chromaticity color space. In *Systems, Signals and Image Processing, 2009. IWSSIP 2009. 16th International Conference on*, pages 1–4. IEEE, 2009. 4

[22] Yusuke Kanzawa, Yoshikatsu Kimura, and Takashi Naito. Human skin detection by visible and near-infrared imaging. In *IAPR Conference on Machine Vision Applications*, 2011. 4, 13

[23] Shoji Tominaga. Dichromatic reflection models for a variety of materials. *Color Research & Application*, 19(4):277–285, 1994. 4, 8

[24] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, et al. Analysis of human faces using a measurement-based skin reflectance model. In *ACM Transactions on Graphics (TOG)*, volume 25, pages 1013–1024. ACM, 2006. 4, 13

[25] Thomas B Fitzpatrick. The validity and practicality of sun-reactive skin types i through vi. *Archives of dermatology*, 124(6):869–871, 1988. 4

[26] Diego Martin-Martinez, Pablo Casaseca-de-la Higuera, Marcos Martin-Fernandez, and Carlos Alberola-López. Stochastic modeling of the ppg signal: A synthesis-by-analysis approach with applications. *Biomedical Engineering, IEEE Transactions on*, 60(9):2432–2441, 2013. 4, 16

[27] Wei Zhou and Chandra Kambhamettu. A unified framework for scene illuminant estimation. *Image and Vision Computing*, 26(3):415–429, 2008. 4

[28] Aapo Hyvärinen and Erkki Oja. Independent component analysis: algorithms and applications. *Neural networks*, 13(4):411–430, 2000. 8

[29] Israel Vite-Silva, Nareli Cruz-Cortés, Gregorio Toscano-Pulido, and Luis Gerardo de la Fraga. Optimal triangulation in 3d computer vision using a multi-objective evolutionary algorithm. In *Applications of Evolutionary Computing*, pages 330–339. Springer, 2007. 11

[30] Vincent Jeanne, Michel Asselman, Bert den Brinker, and Murtaza Bulut. Camera-based heart rate monitoring in highly dynamic light conditions. In *Connected Vehicles and Expo (ICCVE), 2013 International Conference on*, pages 798–799. IEEE, 2013. 14

[31] Bui Tuong Phong. Illumination for computer generated pictures. *Communications of the ACM*, 18(6):311–317, 1975. 15

[32] Markus Hulsbusch. *An Imge-based functional method for opto-electronic detection of skin perfusion*. PhD thesis, RWTH Aachen department of EE, 2008. 16

[33] uEyeCamids development camera basics; camera parameters: Gains and offset, howpublished = `https://en.ids-imaging.com/manuals/ueye_sdk/en/ueye_manual/index.html`, note = Accessed: 2015-06-23. 21

Appendix A

Camera Gain History

A.1 Gains at different Exposure Settings

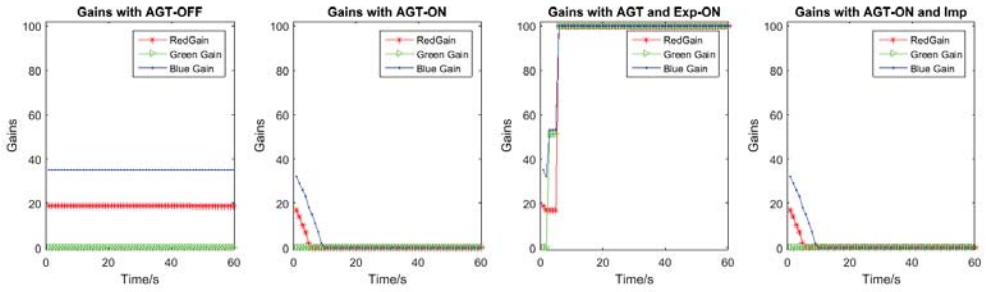

Figure A.1: 80% white clipping

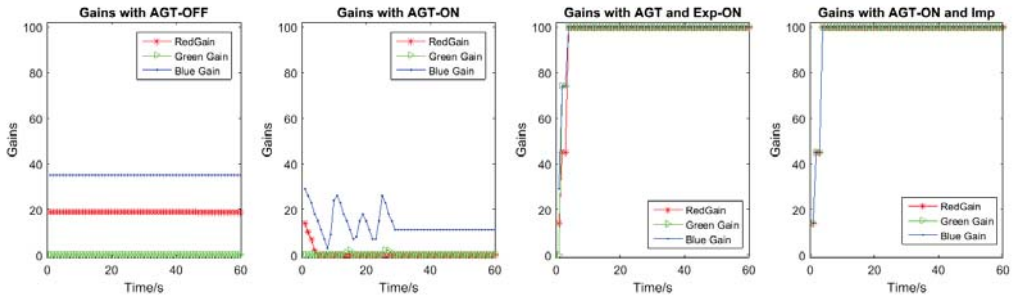

Figure A.2: 40% white clipping

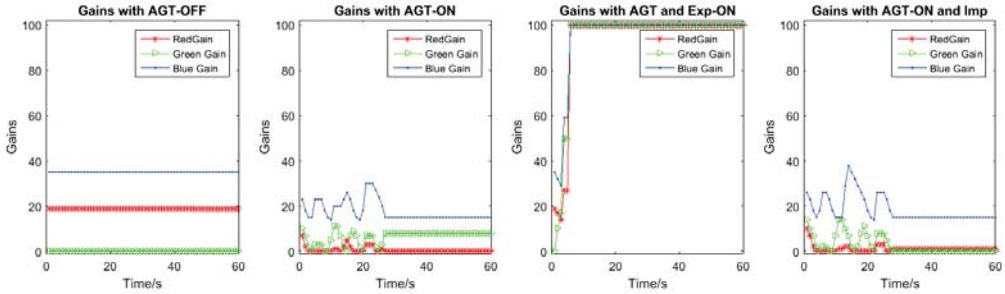

Figure A.3: 20% white clipping

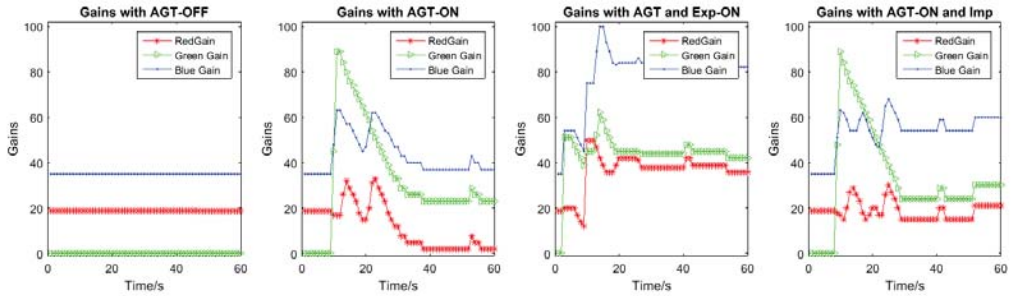

Figure A.4: 0% Clipping

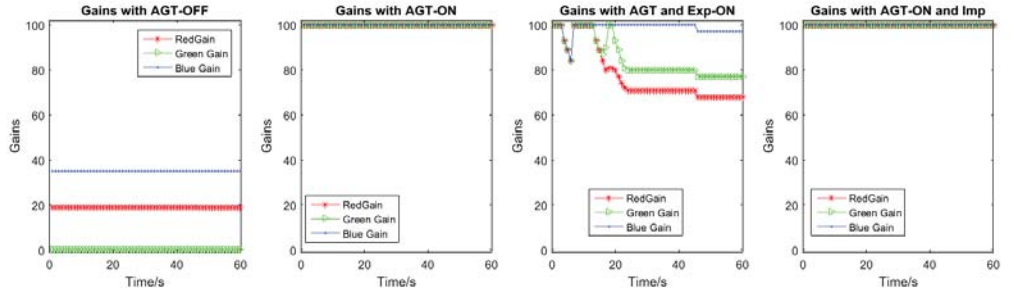

Figure A.5: 20% black clipping

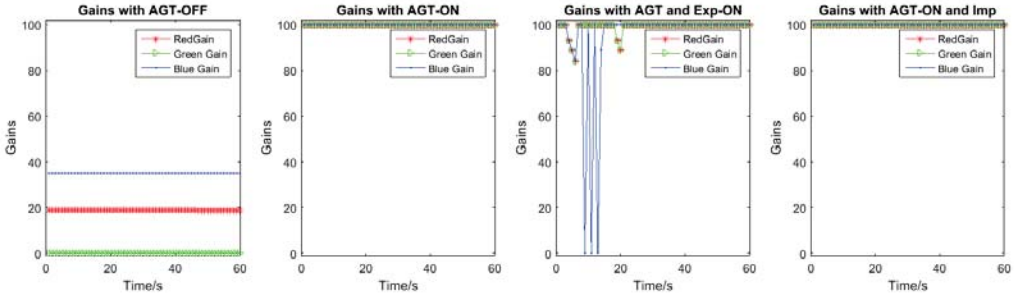

Figure A.6: 40% black clipping

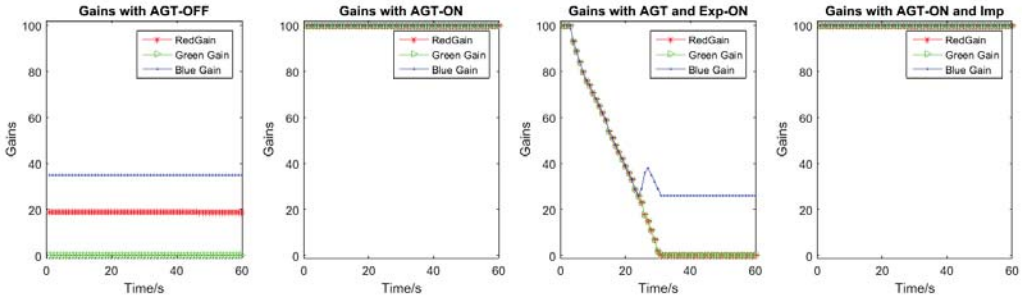

Figure A.7: 80% black clipping

A.2 Dynamic Lighting

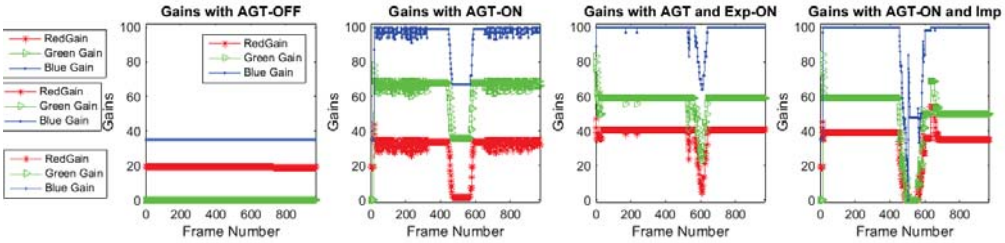

Figure A.8: Cloud cover

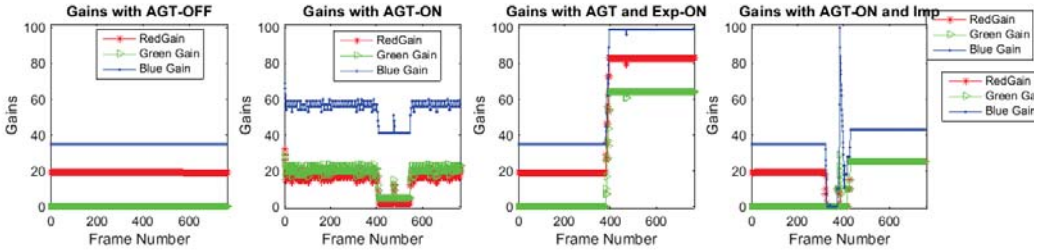

Figure A.9: Glare

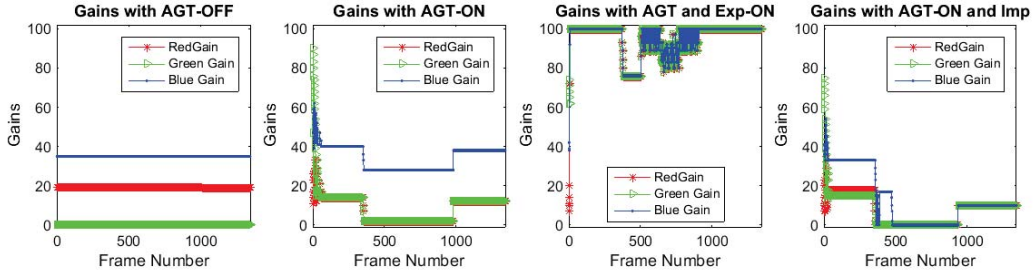

Figure A.10: Dynamic Room Lighting

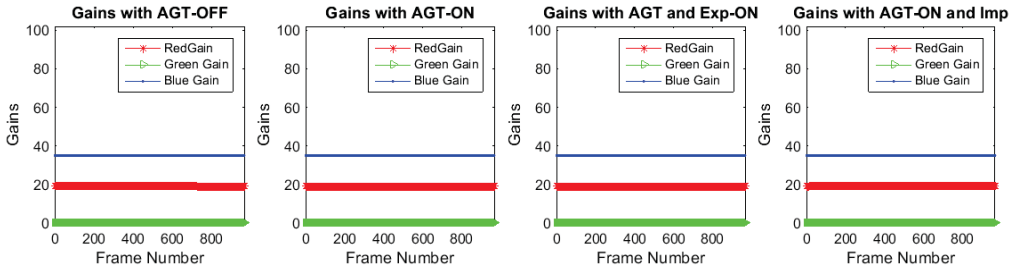

Figure A.11: Background Flicker Gains

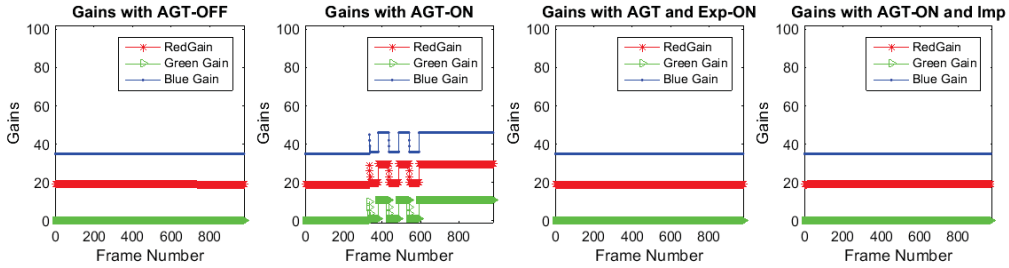

Figure A.12: Monochromatic Facial Flicker

A.3 Real Experiments

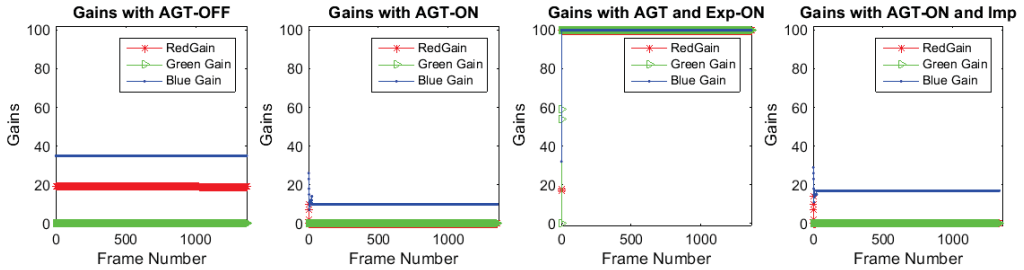

Figure A.13: 70% white clipping

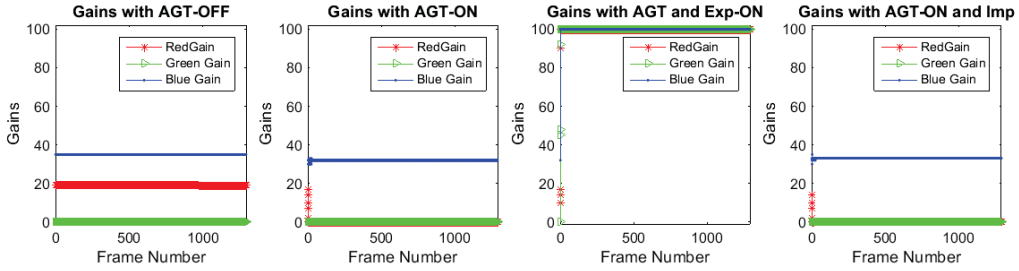

Figure A.14: 40% white clipping

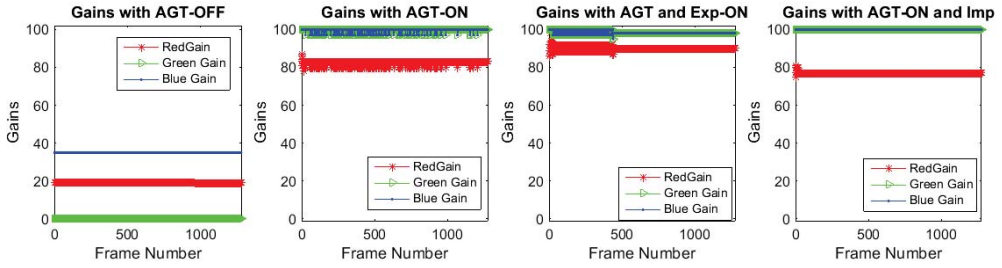

Figure A.15: 0% Clipping

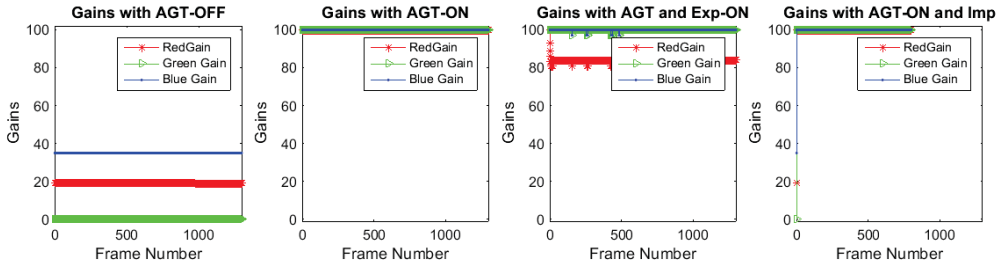

Figure A.16: 50% Black clipping

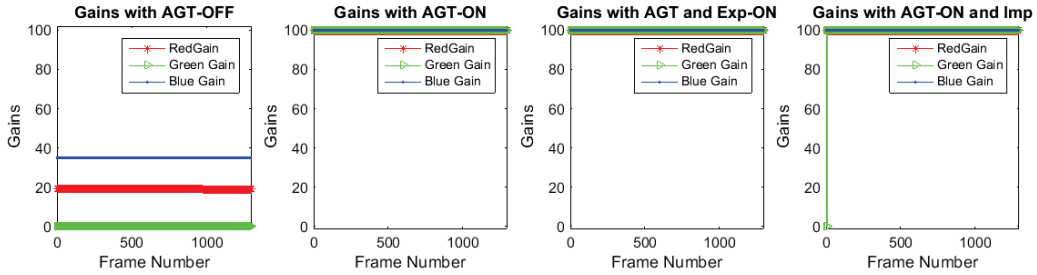

Figure A.17: 80% Black clipping

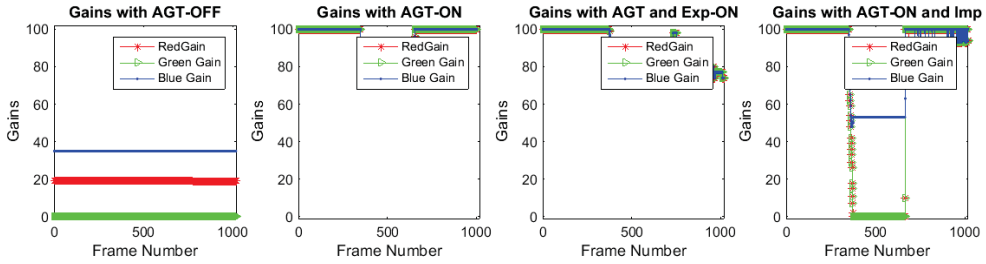

Figure A.18: Switching on Lights

Appendix B

All Experiments

B.1 Experiment Phantoms

B.1.1 Black clip

Figure B.1: First 30 seconds

Figure B.2: Middle 30 seconds

Figure B.3: Last 30 seconds

Investigation using black clip synthetic video

In this experiment, the response of the AGT to black clip conditions during experimentation is tested. A synthetic video of 90 seconds with black clip was rendered using the model explained in chapter. In the video, the ambient Light intensity is set to $[0.8, 0.8, 0.8]$, A headlamp (camera headlight) of intensity [1,1,1] is used for the initial and final 30 seconds. These conditions reproduce standard room light conditions. During the black out, the intensity of ambient lighting is reduced to [0 0 0] and the intensity of the headlamp is reduced to $[0.5, 0.5, 0.5]$. The artificial black out is introduced in the middle 30 seconds of the synthetic video. The positioning of the test conditions(black out) in the middle of normal conditions enables a valid comparison between the normal conditions and the black out.

B.1.2 White clip

The white clip video sequence is illustrated in Figure **??**. The first 30 seconds of the video sequence have 0 % white clip. The middle 30 seconds have white clip which is set with different desired percentages depending on the light intensity of the face facing light added to the model.

Figure B.4: First 30 seconds

Figure B.5: Middle 30 seconds

Figure B.6: Last 30 seconds

Figure B.7: Frames from cloud cover video sequence

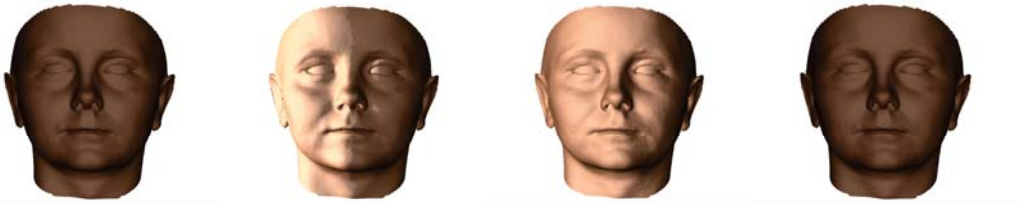

Figure B.8: Synthetic video with Glare

B.1.3 Cloud Cover Experiment

B.1.4 Glare

B.1.5 Face Flickering Light

Figure B.9: Face Flicker

B.1.6 Background light Flicker

B.1.7 Room lighting switch

Figure B.10: Effect of switching on and off of lighting

B.2 Monochromatic light Flicker

Figure B.11: Effect of Monochromatic Flicker

B.3 Effect of Gain Tuning on Raw RGB signals

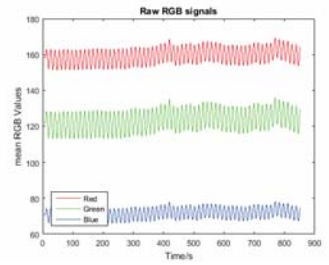

Figure B.12: Raw signals RGBgains=[0,0,0]

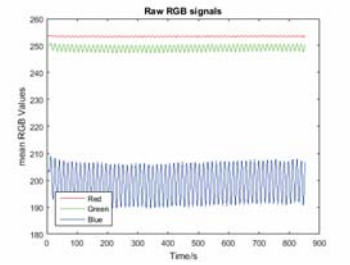

Figure B.13: Raw signals RGBgains=[50,50,50]

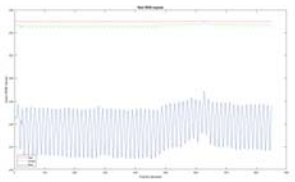

Figure B.14: Raw signals RGBgains=[100,100,100]

B.4 SNR at different Frequencies

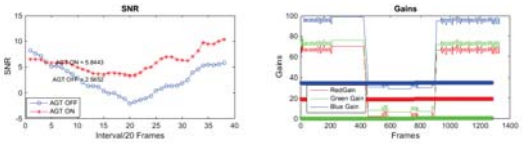

Figure B.15: 0.05 Hz

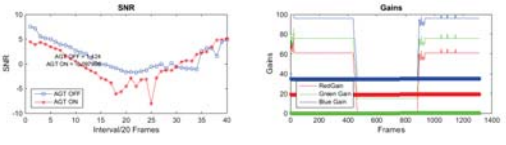

Figure B.16: 0.10 HZ

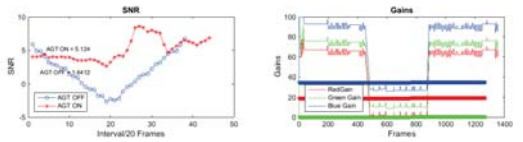

Figure B.17: 0.15 Hz

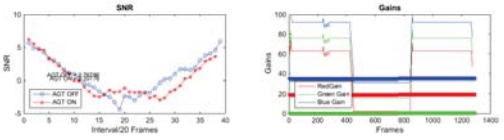

Figure B.18: 0.20 Hz

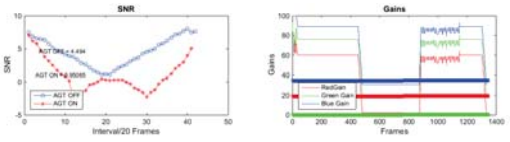

Figure B.19: 0.25 Hz

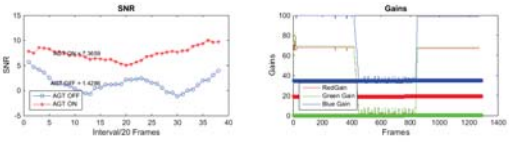

Figure B.20: 0.30 Hz

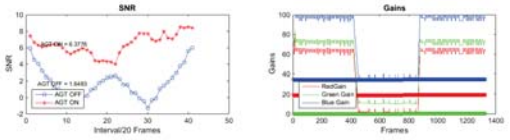

Figure B.21: 0.35 Hz

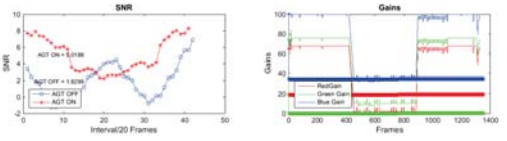

Figure B.22: 0.40 Hz

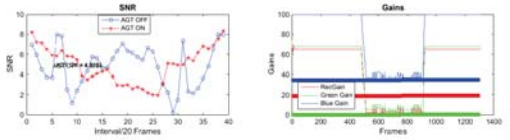

Figure B.23: 0.45 HZ

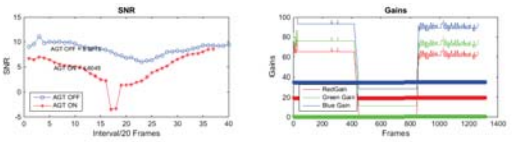

Figure B.24: 0.50 Hz