Graduate Theses, Dissertations, and Problem Reports

2014

Human Recognition from Video Sequences and Off-Angle Face Images Supported by Respiration Signatures

Xue Yang

Follow this and additional works at: https://researchrepository.wvu.edu/etd

Recommended Citation

Human Recognition from Video Sequences and Off-Angle Face Images Supported by Respiration Signatures

Xue Yang

Thesis submitted to the

Benjamin M. Statler College of Engineering and Mineral Resources

at West Virginia University

in partial fulfillment of the requirements for the degree of

Master of Science

in

Computer Science

Thirimachos Bourlai, Ph.D., Chair

Yuxin Liu, Ph.D

Jeremy Dawson, Ph.D.

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia

2014

Keywords: Human Respiratory Waveform Extraction, Human Identity Recognition, Respiratory Waveform Feature Extraction, Component-Based Face Recognition

Abstract

Human Recognition from Video Sequences and Off-Angle Face Images Supported by Respiration Signatures

Xue Yang

In this work, we study the problem of human identity recognition using human respiratory waveforms extracted from videos combined with component-based off-angle human facial images. Our proposed system is composed of (i) a physiology-based human clustering module and (ii) an identification module based on facial features (nose, mouth, etc.) fetched from face videos. In our proposed methodology we, first, manage to passively extract an important vital sign (breath), cluster human subjects into nostril motion vs. nostril non-motion groups, and, then, localize a set of facial features, before we apply feature extraction and matching.

Our novel human identity recognition system is very robust, since it is working well when dealing with breath signals and a combination of different facial components acquired in uncontrolled luminous conditions. This is achieved by using our proposed Motion Classification approach and Feature Clustering technique based on the breathing waveforms we produce. The contributions of this work are three-fold. First, we collected a set of different datasets where we tested our proposed approach. Specifically, we considered six different types of facial components and their combination, to generate face-based video datasets, which present two diverse data collection conditions, i.e., videos acquired in head fully frontal position (baseline) and head looking up pose. Second, we propose a new way of passively measuring human breath from face videos and show comparatively identical output against baseline breathing waveforms produced by an ADInstruments device [1]. Third, we demonstrate good human recognition performance when using the proposed pre-processing procedure of Motion Classification and Feature Clustering, working on partial features of human faces.

Our method achieves increased identification rates across all datasets used, and it manages to obtain a significantly high identification rate (ranging from 96%-100% when using a single or a combination of facial features), yielding an average of 7% raise, when compared to the baseline scenario. To the best of our knowledge, this is the first time that a biometric system is composed of an important human vital sign (breath) that is fused with facial features in such an efficient manner.

Acknowledgments

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

Human identity security is becoming increasingly important for our society. Nowadays, there are examples of real-life cases indicating that the traditional authentication techniques based on passwords, identity documents or physical tokens fail to provide enough safety. In order to authenticate personal identity with comparatively high security, it would be more convenient and safe for users to employ a set of biometric features derived directly from their physical or behavioral characteristics (also known as traits or identifiers) [6]. There already exists several typical biometric authorization systems [7] or simple biometric features (such as fingerprint [8], face [9], hand geometry [10], iris [11], signature [12], etc.) in use for diverse applications [13].

Compared to traditional authorization schema (e.g., password), biometric-based identification systems have presented significant advantages in identity security aspects [6]. Biometric traits are inherently more reliable, as they cannot be forgotten or lost. While simple passwords are easy to remember, and also easy to guess by others. While complicated passwords can be lost or forgotten. On the other hand, it is difficult for others to share or duplicate the inherent characteristics of human beings, but passwords can be stolen by hackers and presented on public. And it is a comparatively strict requirement that the authentication process has to be done only when the person is present at the point of authentication. Also, it is quite difficult to change

or erase the personal biometrics, since more time and money as well as advanced technology are needed to complete it. On the other hand, modifying a password only requires a few seconds. Therefore, designing an authentication system based on critical biometric characteristics is supposed to be an effective replacement for traditional authentication approaches.

The identity security industry demands the promotion of further research on discovering and developing novel biometric-based identification systems that can help supplement or replace the traditional ones. For example, lately we have seen a number of studies demonstrating the capability of using the electrocardiogram (ECG) signal for human identity recognition [14][15][16][17], since human individuals present certain distinctive patterns in their ECG signals (such as wave shape, amplitude, interval of two adjacent local maximum pointsl).

Inspired by the analysis of ECG-based identification systems, we propose a novel video-based biometric approach that is composed of two major components:

- A physiology-based human clustering module (passive human respiratory monitoring and analyzing using a visible camera).

- An identification module based on facial features (nose, mouth, eye, eyebrow) fetched from face videos.

A typical application of such an approach is face recognition (FR), on laptops and/or cell phones, which can be used for different purposes, including login access etc. Actually several techniques are used for respiratory monitoring [18], such as spirometers [19], nasal thermocouples [20], transthoracic inductance, impedance plethysmography [21], strain gauge [22], etc. While each of these techniques requires a dedicate device attached to some region of human body. The method here proposed derives respiratory waveforms directly from videos recording of human faces passively, i.e. without any physical contact.

The face recognition component-based approach reveals promising aspects in various security applications [23]. The main effect of using components is more tolerant of head pose changes by considering a flexible geometrical relation between the selected facial parts during the recognition stage [24]. In [23] face recognition studies were conducted by combining a set facial components into a single feature vector and classifying using linear Support Vector Machines (SVM) [25]. An improved technique was proposed by combining 3D morphable models

with component-based recognition that computed the 3D face models from input face images, which were used for training the proposed component-based face recognition system [26]. Also some previous studies [24][27] discussed the approaches of selecting the discriminatory facial components to categorize subjects. Finally, there are other face recognition component-based approaches when operating in the mid-wave infrared (MWIR) band [28]. In these papers, face matching is performed utilizing fiducial points on the face when using either the whole or sub-regions of the human face. Face matching was also reported to be independent of the operational band [29].

In our study, we mainly focus on human face identification using the nose component and its combination with other typical facial components (e.g. eyes, mouth, eyebrow). The pre-processing stage for our component-based FR system is composed of a physiology-based classification module, where the features used for classification are generated by respiratory waveforms produced from our breathing waveform extractor, and the subjects are grouped into nostril motion vs. nostril non-motion groups before matching is performed that significantly increases the identity recognition efficiency and reduces process computation. The purpose of the classification module is to improve the identification performance of our hybrid biometric system, as we will discuss in the following sections.

1.2   Goals and Contributions

The goals of this work include two major aspects as following:

- The design of a non-contact measurement of the human respiratory waveforms based on face videos that can guarantee the comparative performance to the traditional human breathing signal monitoring devices.

- The development of effective pre-processing methodologies can accurately group each of the original biometric gallery and probe datasets into smaller subsets (such as nostril-motion specific cohorts or top 5 most similar feature sets) prior to applying single- or multi-facial component matching, with the purpose of improving identification performance.

In this regard, we first propose a respiratory waveform extraction schema. The objective is, for each frame of our available subject-specific face video, to detect and localize the nose and each nostril, and then, measure a set of pre-defined features, i.e. the width and height of each nostril in terms of pixel distance. As we will discuss in more detail in our methodology section, the measurements obtained by our respiratory extraction schema consequentially produce a client-specific de-noised set of breathing signals representing a 60-second long breathing waveform that combines all four measured features. In order to facilitate this, the Root Mean Square (RMS)[30] calculation is applied respectively to both left and right nostril feature values. Thus, for each frame, these values are averaged to one output generating a hybrid temporal respiratory waveform.

Second, we propose a motion classification approach when obtaining the respiratory waveform from any given video. For this purpose, we perform respiratory waveform peak and bottom detection. Then, based on the detectable number of local maximums and minimums, the given waveforms are categorized into two groups, i.e., nostril motion and non-motion group. In order to further reduce the number of potential candidates to match with in the identity recognition process, we define and extract six features from each input breathing waveform, and employ the modified K-Nearest Neighbor algorithm [31] to find the top five most similar subjects within each classified group. The key characteristics of the proposed pre-processing procedure are the following:

- Instead of comparing all subjects in the testing session, while conducting identity recognition, we narrow down the range of potential identical subjects in advance with comparatively high accuracy, hence ensure that the computational complexity of the recognition algorithm is low.

- The proposed pre-processing approaches of both motion classification method and feature clustering technique achieve very high classification accuracy, as demonstrate by the experimental results.

Moreover, the images employed during the face recognition process are six different combinations of facial components acquired randomly from videos under two collection scenarios:

[Scenario One] - Uncontrolled condition of variant luminance and unfixed head pose; and [Scenario Two] - Semi-controlled condition with variant luminance and fixed head position. The proposed recognition system is tested under three experiments:

- No pre-processing methodology is involved and directly match images (it works as baseline scenario).

- Pre-processing stage contains only motion classification method.

- Pre-processing stage includes both motion classification and feature clustering approach.

And all three experiments work on both original images and waveform-based normalized (WA) [32] datasets. Standard face recognition academic algorithms are used including, Local Binary Patterns (LBP) [33] and Local Ternary Pattern (LTP) [34]. Thus we argue that the proposed recognition procedure can achieve high matching performance, especially when baseline experiments shows comparatively lower identification rates.

1.3 Structure of the Thesis

The rest of this thesis is organized as follows:

- **Chapter 2:** initially gives a description of the human electrocardiography (ECG) identification techniques that inspired our research work, the existing techniques that monitors and measures human breathing signals and how our designed respiratory waveform extraction approach is different from the traditional devices.

- **Chapter 3:** describes the experimental devices, conditions employed during the data collection stage, the detailed information about the data collection stage, the face video datasets and the component-based human face image databases applied in this work.

- **Chapter 4:** provides a summary of the proposed respiratory waveform extraction approach, the designed pre-processing techniques (including the proposed motion classification and feature clustering methods), the face normalization technique employed, and the face recognition algorithms applied.

- **Chapter 5:** introduces the detailed design and the purpose of the three experiments used to test the research hypothesis, displays the output from each step of pre-processing approaches, and analyzes the experimental results under different experiments of two different types of input face image databases.

- **Chapter 6:** summarizes the conclusions of this thesis and future work.

Chapter 2

Related Work

This chapter generally describes the human electrocardiography (ECG) identification techniques, especially the specific features defined and extracted from the ECG curves that have shown the identity distinctness for human recognition [14][15][16][17]. In this chapter, we also list the existing techniques that monitors and measures human respiratory signals, and explain how our implementation of breathing waveform extractor is different from the traditional devices.

2.1 Human ECG for Biometric Recognition

Human electrocardiogram (ECG) is used to record and measure different electrical potentials of the heart, which was initially developed by Willem Einthoven in the early 1900s. The analysis of ECG data are traditionally served as a tool for clinical diagnosis, and the researches have shown the link between cardiac function and the expression of the ECG trace. Recently, a lot of studies have proposed the potential probability of employing ECG data as a novel biometric attribute for human identity recognition purpose. In [35], the study has supported the probability of applying ECG signals for human biometric identification based on the fact that both the physiological and geometrical differences of the heart from the distinct individuals have presented some certain uniqueness from their ECG data.

A typical ECG signal of a normal heartbeat is mainly consists of the following components: a P wave, a QRS complex and a T wave shown in Fig. 2.1 [2]. Recent researches have indicated
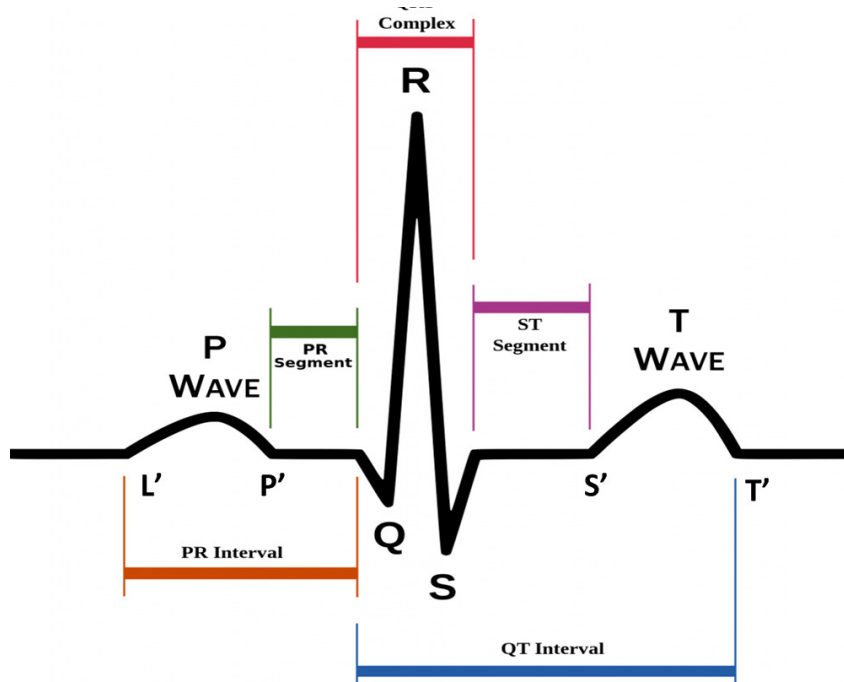
Figure 2.1: Main components of an ECG signal [2].

that human individuals present distinct patterns of their ECG curve in regards to the wave shape for an ECG cycle, the amplitude for PRT peaks, the PT interval, because of the physical conditions of individual's heart [36].

In [37], a SIEMENS ECG equipment is used to extract a set of temporal and amplitude attributes directly. Then a feature selection algorithm is applied to remove unnecessary and useless features based on the analysis of correlation matrix (i.e. the features with a relatively high correlation with other features are removed). The empirical feature selection algorithm reduced the attributes number from 30 to 12 (i.e. 30 is the original number of features delivered from the SIEMENS device). The identification system was tested on a database of 20 subjects, and successfully obtain 100% identification accuracy. One critical limitation of this method is that both the feature measurement approach and the feature selection procedure is not automatic and has to be dependent on specific equipment.

In [38], the proposed of human identity recognition was consists of two main steps based on input one-lead ECG. The correlation coefficient for comparing two QRS complexes based on the specific template matching method as to find the potential candidates for the matching process. Finally, the identity recognition stage was completed by applying the decision-based neural network (DBNN) technique for the potential identical candidates selected from the com-

pleted template matching procedure. The final accuracy can achieve 100% identification accuracy when combines the template matching (i.e. 95% of correct verification rate) and DBNN method (i.e. 80% correct rate).

Israel et al. [17] proposed an extensive set of ECG descriptors that characterizes the trace of the heartbeat in detail. An input ECG signal is firstly applied a bandpass filter to remove non-signal components. The peaks (i.e. points P, R, T) were established by finding the local maximums that are surrounding each of the P, R, T complexes, and the base points were localized by tracking downhill and discovering the location of the minimum radius of the curvature. And in total 15 features listed in Table 2.1 were extracted from each heartbeat signal, which is defined within the normalized distance between the two fiducials. And a stepwise canonical correlation that applied the Wilkes lambda as a divergence measure worked for the feature selection process. A database of 29 subjects was used for testing the identification system, and achieved 100% human identification accuracy and about 81% heartbeat recognition rate. In this method, the automatic feature extraction and identity recognition process can be achieved.

In general, the existing works about ECG biometric identification measure different critical components of input ECG signal and utilize feature vectors for further classification procedure. In our study, we referenced this research schema that define and extract several attributes from each breathing cycle curve and establish the six-dimension feature vector for further clustering procedure.

Table 2.1: Extracted 15 attributes. The features list labels are employed the normalized distance between the two fiducials.

Feature ID	Feature Definition	Feature ID	Feature Definition
1	RQ distance	2	RS distance
3	RP distance	4	RL distance
5	RP' distance	6	RT distance
7	RS' distance	8	RT' distance
9	P width	10	T width
11	ST distance	12	PQ distance
13	PT distance	14	LQ distance
15	ST' distance		

2.2 Human Respiratory Signal Measurement

Detection and measurement of human respiratory rate (RR) works as a significant component of initial and continuous evaluation of actual patients for clinical diagnosis. The automatic recording of respiratory signal techniques can be applied to monitor a number of physiological variables. There exists several typical equipment that can be used to achieve the monitoring and recording of human respiratory signals [18], which are listed as follows:

- **Spirometers and nasal thermocouples:** these two devices keep measuring the air flow in and out of the lungs directly during inhalation and exhalation, which can produce detailed breathing signals. However, they may interfere with the respiration process that may influence the accuracy of the actual breathing waveform.

- **Transthoracic inductance and inpedance plethysmographs, pneumatic respiration transducers, whole-body plethysmographs, and strain gauge measurement of thoracic circumference [18]:** these equipment can also be used to indirectly monitor respiratory signals based on the measurement of the variation of body volume.

All these techniques mentioned above are required some particular devices attached to some specific component of human body. Travaglini et al. [18] proposed a novel method that derives reliable detection of respiratory waveform from input human original ECG signals. This technique doesn't require additional equipment, which is only based on some signal processing approaches, and works particularly effective when the input information is only limited on human ECG data. In this study, eight leads of ECG signals were utilized and worked as the input data for the conversion of breathing signal, and the whole technique was based on the hypothesis that the points representing the human respiratory waveform can be aligned around the preferred direction in 8-D space. A database of 10 subjects have been used to test this technique under several breathing conditions, such as deep breath and short breath, and the results have been compared with the signals generated from the strain gauge measurement of the thoracic circumference. This approach can be applied to existing equipment and help analyze the ECG signals in order to obtain addition information without utilizing other additional devices.

In our study, the respiratory waveforms are detected and extracted directly from face videos based on some video-processing methods and signal-processing techniques. This approach doesn't require any additional devices or monitors attached to human body, and only a digital camera is necessary to conduct the video recording from a specific distance from the human face. Also, it can be incorporated into real-time monitoring system.

Chapter 3

Experimental Setup

Two unique facial video databases, i.e. DB_FF (fixed fully frontal face pose) and DB_HU (head-up face pose) were manually collected and considered to facilitate the proposed study. Each database consists of two sets of video recordings (session 1 and session 2) of the same 30 subjects. The data collection of session 1 and session 2 were conducted on different days in order to guarantee the facility and validity of the gallery and probe. For each session, five 1-minute long videos were recorded for each subject (i.e., totally 150 videos are included in each set). From each video database and for the purpose of conducting the face component-based recognition experiments, two sets of six different face image databases (DB_FF1~DB_FF6 and DB_HU1~DB_HU6) of diverse facial component combinations (i.e., eye, eyebrow, mouth, and successively group with nose to build the image databases) were generated.

The following subsections provide more detailed information and settings about each database as well as the equipment used for acquiring the facial videos and the baseline breathing waveform data from device.

3.1 Equipment

- **Cannon 5D Mark II:** This digital SLR camera has a 21.1-megapixel full-frame CMOS sensor with DIGIC 4 Image Processor and a vast ISO range of 100-6400. It supports Live View HD video recording with up to 39 frames per second (FPS). In this study, the Mark II is used to obtain ultra-high resolution facial videos (about 1-minute for each recording)

in the visible spectrum.

- **ADInstruments Respiratory Belt Transducer:** This Respiratory Belt Transducer contains a piezo-electirc device that measures changes in thoracic or abdominal circumference during respiration, which indicate inhalation, expiration, breathing strength and can be used to derive breathing rate. In this study, the Respiratory Belt Transducer is used to track and measure the up-and-down movement of the subject's chest accompanying the inhalation and exhalation while taking deep breath and convert it to the respiratory waveform signal served as the baseline data, which helps assess the performance of our developed respiration waveform extraction system.

- **ADInstruments PowerLab 4/35:** This PowerLab device is a high-performance data acquisition hardware physically connected with Respiratory Belt Transducer. It is capable of recording at speed of up to 400,000 samples per second to disk continuously with four analog input channels. In this study, we set up the sampling rate to be 40 samples per second, which is the most compatible option to our video recording frame rate (39 frames per second). And the generated datasets are stored in MAT format.

Table 3.1: Utilization information about each database. All the videos were collected under natural light that was in uncontrolled illumination condition and might vary according to the different collection days.

Database	Cameras	# of Subjects	# of Sessions	# of Data (/subject /session)	Head Pose
DB_FF	5D Mark II	30	2	1	Full Frontal Face
DB_HU	5D Mark II	30	2	5	Head up at $30° - 40°$
DB_FF1~DB_FF6	5D Mark II	30	2	5	Full Frontal Face
DB_HU1~DB_HU6	5D Mark II	30	2	5	Head up at $30° - 40°$

3.2 Databases

Thirty (30) subjects were involved into the data collection over two sessions conducted on two different days in order to guarantee the facility and validity of the gallery and probe data. The

following is a detailed description for each database utilized in our experiments. Additional

information can be found in Table 3.1.

Table 3.2: Six different types of facial-component combinations and their corresponding covered regions employed during the identity recognition stage, which is to eliminate the potential influence on the matching performance caused by other facial features not included within the specific combination.

Database	Facial Components	Image Size	Covered Area
DB_FF1, DB_HU1	Nose	180*150	Top_Width:55; Top_Height:40; Bottom:17;
DB_FF2, DB_HU2	Nose, Eye	330*180	Top_Width:0; Top_Height:0; Bottom:17;
DB_FF3, DB_HU3	Nose, Mouth	330*260	Top_Width:135; Top_Height:42; Bottom:22;
DB_FF4, DB_HU4	Nose, Eye, Eyebrow	330*250	Top_Width:165; Top_Height:20; Bottom:22;
DB_FF5, DB_HU5	Nose, Mouth, Eye	330*340	Top_Width:165; Top_Height:50; Bottom:25;
DB_FF6, DB_HU6	Nose, Mouth, Eye, Eyebrow	330*340	Top_Width:165; Top_Height:0; Bottom:25;

- **DB_FF:** Collected in an indoor environment with semi-controlled conditions of fixed head pose and variant illumination. And high quality of one 10-second video was captured for each subject in each session with a Cannon 5D Mark II with 39 frames per second and 1920 by 1080 pixel per image. Videos were collected at fully frontal head position with 2-meter distance between the camera.

- **DB_HU:** Collected in an indoor environment with uncontrolled conditions of unfixed head position and variant illumination. Five videos of 60-second long were captured for each subject in each session utilizing a Cannon 5D Mark II with 39 frames per second and 1920 by 1080 pixels per image. Videos were acquired at a un-fixed head-up pose, i.e., the pitch (see Fig. 3.1 [3]) was between at 30 to 40 degrees plus (head looking up) with 2-meter distance between the camera.

- **DB_FF1∼DB_FF6:** For each session, 5 visible full frontal face images were extracted (per subject) from the corresponding group in DB_FF. The number 1 to 6 represents a specific component-based face image database as follows. And more detailed information about each type of component-based face image used in our work can be found in

Table 3.2.

- **Number 1:** Image contains only nose part.

- **Number 2:** Image contains nose and eye.

- **Number 3:** Image contains nose and mouth.

- **Number 4:** Image contains nose, eye, and eyebrow.

- **Number 5:** Image contains nose, mouth, and eye.

- **Number 6:** Image contains nose, mouth, eye, and eyebrow.

- **DB_HU1~DB_HU6:** For each session, five visible head-up face images (per subject) were extracted from the corresponding group in DB_HU. The number 1 to 6 represents a specific component-based face image database as same as the combinations of facial parts in DB_FF1~DB_FF6. More detailed information about each type of component-based face image used in our work can be found in Table 3.2.
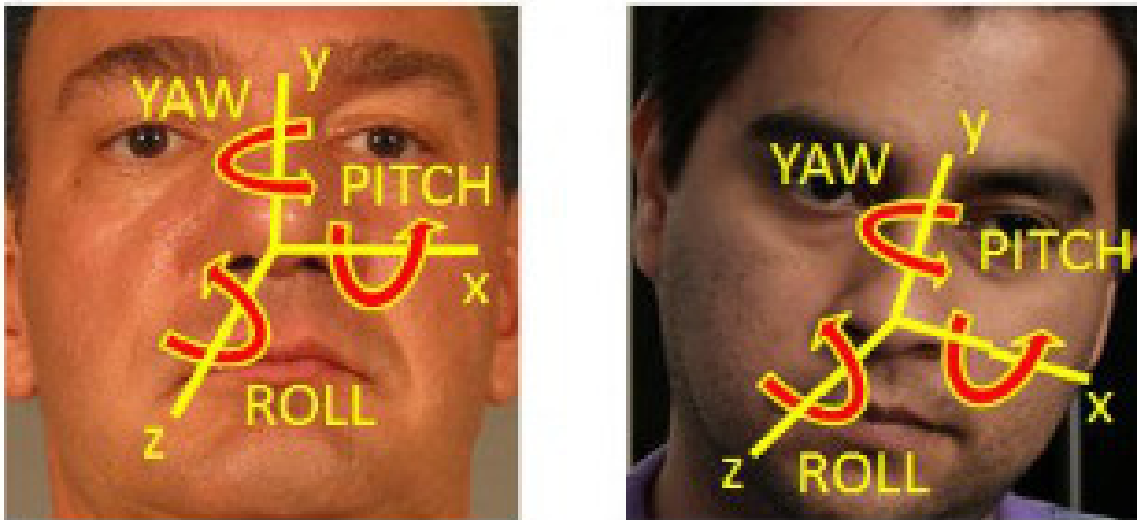
Figure 3.1: Illustration of three types of rotational descriptors on face images acquired under variable conditions using two different sensors [3]. Note that in order to clearly track and record the movement of nostrils, the face videos (DB_UH) used in our study were generated when the pitch descriptor rotated between 30 to 40 degrees, i.e., in head looking up position.

Chapter 4

Methodology

In this section, we outline the techniques used for achieving and manipulating human respiratory waveforms from input face videos, and describes the proposed two data pre-processing methodologies (i.e., motion classification method and feature vector clustering approach) and facial recognition algorithms. The overall work is displayed in Fig. 4.1 and the salient stages of the proposed approaches are described below.

4.1 Breathing Waveform Extraction

In this study, instead of applying traditional techniques of monitoring and measuring human respiratory signals, a novel approach dealing with head-up face videos (DB_HU) for non-contact measurement of human breathing waveform is completely developed based on nostril movements during inhalation and exhalation while taking breath. Throughout data collection, while asking subjects to breathe, a number of them showed an obvious enlargement of the nostrils opening during inspiration, which is a medical symptom named Nasal Flaring [39]. And our designed respiratory waveform extractor is mainly based on the changing opening of the nostrils.

Nasal flaring is the condition where human nostrils are getting dilated compared to normal situation, which usually occurs during inspiration and may occasionally happen during expiration or throughout the breathing cycle. Certain respiratory treatments, such as deep breathing, can cause nasal flaring [40].
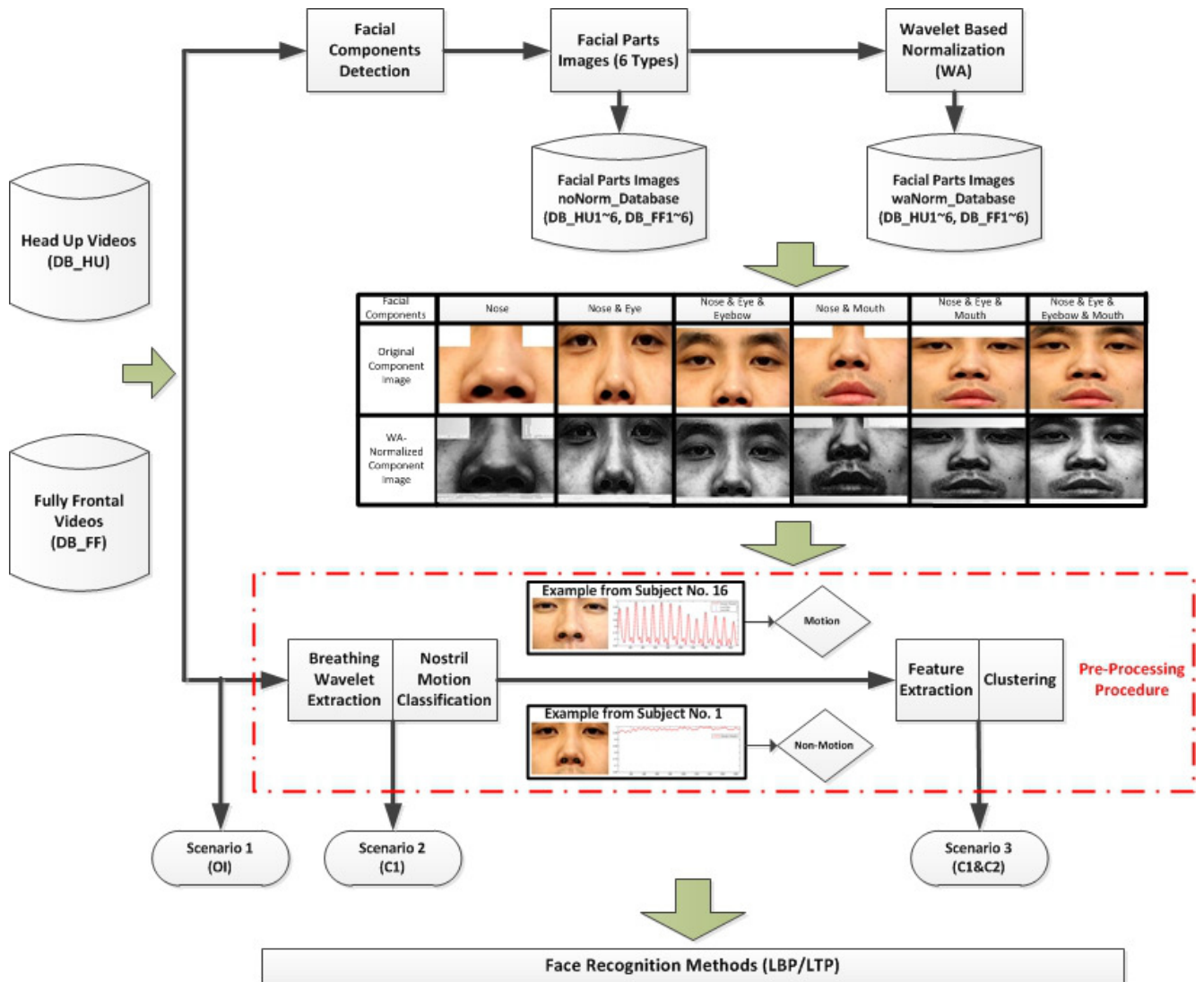
Figure 4.1: Overview of input experimental datasets, the design of our proposed pre-processing approaches based on human respiratory waveforms that supports the human identity authentication system, and the three testing scenarios.

In our case, we exploit this physiological condition (see Fig. 4.2), and for each frame within one given video, the proposed method, firstly detects and localizes the nose component position using the Masayuki Tanaka face parts detection algorithm [5] , then, detects the edges of the facial features and measures the horizontal width W and the vertical height H for both nostrils (i.e., collects four critical feature values: $H_{left}, W_{left}, H_{right}, W_{right}$). Fig. 4.3 illustrates the measurement approach of these four defined features on each given frame from input video. By accumulatively collecting the value for each type of distance produced from any single video, we managed to obtain four sets of data (as discussed above) representing
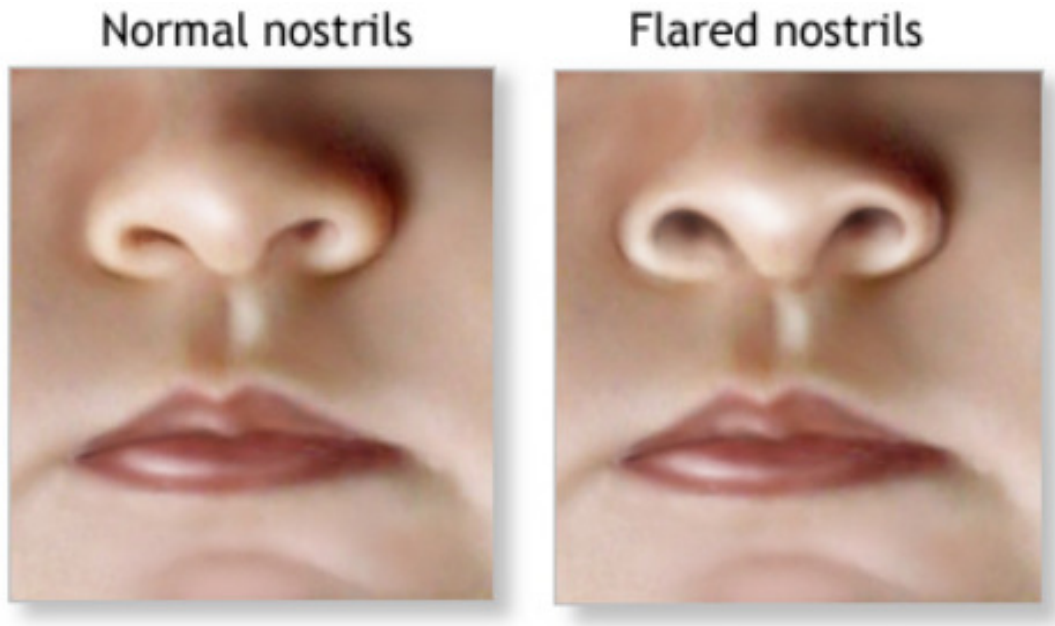
Figure 4.2: A visual example where flared nostrils are observed. This condition is noticed over time on some individuals, and if exploited properly, we can group subjects before performing any face component-based matching that will significantly reduce the computation complexity during recognition process [4].
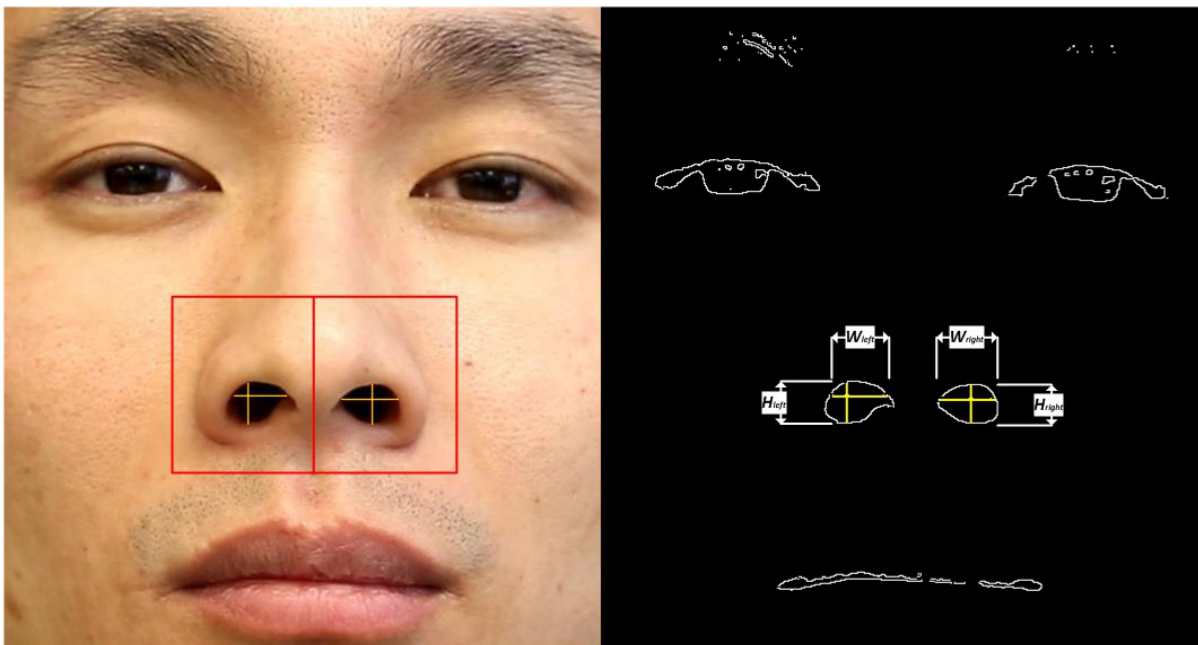
Figure 4.3: Illustration of our nostrils features measurement approach. The nose component is firstly detected and localized (shown as red squares) using the Masayuki Tanaka face parts detection algorithm [5]. Then, 4 features (left height H_{left}, left width W_{left}, right height H_{right} and right width W_{right}) are measured. The picture is an example frame from subject no. 16 (session no. 1, video no. 1, frame no. 12).

18
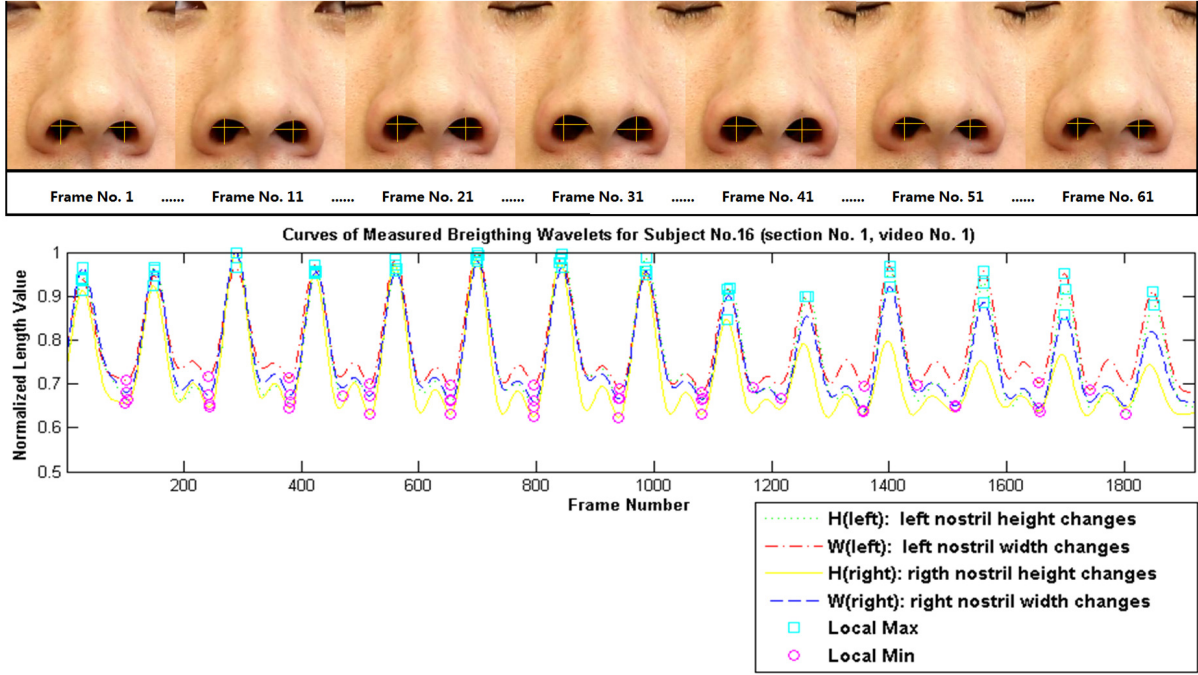
Figure 4.4: Example of extracting breathing waveforms from one of the collected video. The proposed approach keeps measuring four nostril features for each frame along the video, then apply the 0-1 normalization and de-noising algorithms on the collected data for each extracted feature value set. The top Fig. represents the frames selected with 10-frame interval to display the observable nostril movement. Bottom Fig. represents the four curves for the measured nostril features within one video. Both the frames and the curves shown are collected from subject from subject no. 16 (session No. 1, video No. 1).

the nostrils four-dimension movement within 1-minute time. Besides considering the measurement errors (such as the slight difference of the distance between camera and subjects during video shooting process caused by the unconscious body movement), all four sets of feature values were processed using 0-1 normalization by scaling between 0 and 1, using equation (1) and applying a signal de-noising algorithm in order to smooth the signal waveforms. Fig. 4.4 shows the breathing waveform extraction process of converting measured four nostril features of each frame from input face video into consecutive curves with 0-1 normalization and signal de-noising manipulations.

$$Normalized(e_i) = \frac{e_i - S_{min}}{S_{max} - S_{min}} \tag{4.1}$$

Where e_i is the target value to be normalized; S_{max} and S_{min} are the global maximum and

minimum values in the given set of data.

Instead of analyzing four nostril features separately, the Root Mean Square [30] (RMS, also called quadratic mean) measurement is conducted on each frame to combine each pair of two features from one side nostril (i.e., H_{left} and W_{left} for left nostril; and H_{right} and W_{right} for right nostril). And the averaged RMS value for both nostrils is the target output for one given frame, and its accumulative collection of data throughout one input video forms the resulting respiratory waveform. The RMS works as a tool to comparatively reduce the measurement errors, and it is mathematically characterized as:

$$L_{left} = \sqrt{\frac{H_{left}^2 + W_{left}^2}{2}} \tag{4.2}$$

$$L_{right} = \sqrt{\frac{H_{right}^2 + W_{right}^2}{2}} \tag{4.3}$$

$$L_{mean} = \frac{L_{left} + L_{right}}{2} \tag{4.4}$$

Where H_{left}, W_{left}, H_{right}, and W_{right} are the measured nostril feature values for one given frame; L_{left} and L_{right} are the resulting RMS values for left and right nostrils and L_{mean} is the target mean value combining 4 features produced from one input frame. The collection of the L_{mean} value for each frame along one video make up the target dataset representing one breathing waveform.

The resulting breathing waveform for our designed system is generated according to the collected set of the mean value L_{mean} based on left RMS value L_{left} and right RMS value L_{right} for each frame. All the videos were collected with 39 frames per second, thus the sampling rate for the breathing waveform is 39 points per second. However, the ADInstrument device we used to acquire the baseline data works at 40 samples per second. Therefore in order to guarantee the same dimension of two datasets and equally compare these two respiratory curves, the Linear Interpolation is applied to baseline dataset (i.e., from ADInstrument device). One example of the comparison between the output waveform extracted from one input video and its corresponding baseline breathing curve (i.e. real-time waveform detected via ADIn-
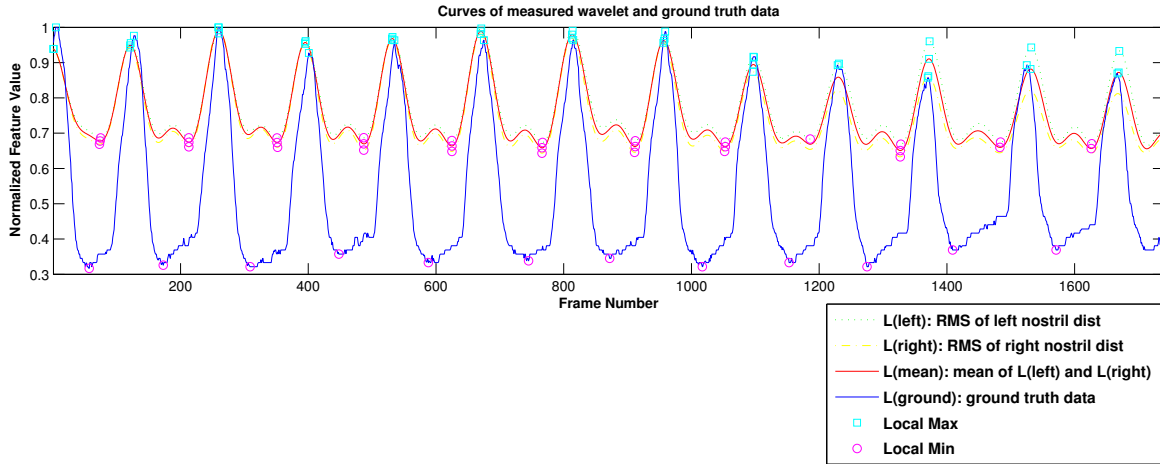
Figure 4.5: Comparison between the breathing waveform output from our designed system (red line) and the baseline breathing curve measured by ADInstruments device (blue line). The waveforms are selected from subject no.16 (session no.1, video no.1) which is the same example subject shown in Fig. 4.3 and Fig. 4.4.

struments device) is shown in Fig. 4.5. And by analyzing this example, for each breathing cycle (i.e., includes consecutive two local bottoms and one local maximum point) these two waveforms can achieve a good match at the time dimension.

In order to get a precious comparison between the resulting output from the designed breathing waveform extraction system and the signal produced by the ADInstruments indicating the same video taking process, the Root Mean Square [30] was also applied on each pair of breathing signals (i.e. one is from the ADInstruments, and the other one is from the developed system). Table 4.1 indicates the results from the Root Mean Square calculation.

4.2 Nostril Motion Classification

After applying the breathing waveform extraction on each video, for every subject on both gallery and probe sessions 5 new data sets are generated and each group of data represents a 60-second long respiratory waveform of 39 sampling rate extracted from its corresponding head-up face video in database DB_HU (i.e., each produced new data set consist of 1 by 2340 dimension). Through the observation and analysis of the collected videos from database DB_HU, one phenomenon has been noticed that the nostril flaring symptom doesn't happen on

Table 4.1: The Root Mean Square results between the produced waveform from the designed system and its corresponding baseline data generated from the ADInstruments.

# of Subject	Section 1	Section 2	Average	# of Subject	Section 1	Section 2	Average
1	0.20	0.18	0.19	2	0.26	0.28	0.27
3	0.13	0.12	0.12	4	0.27	0.30	0.29
5	0.20	0.16	0.18	6	0.21	0.17	0.19
7	0.06	0.12	0.09	8	0.06	0.07	0.06
9	0.19	0.15	0.17	10	0.11	0.13	0.12
11	0.07	0.11	0.09	12	0.13	0.14	0.14
13	0.28	0.29	0.28	14	0.12	0.16	0.14
15	0.24	0.24	0.24	16	0.15	0.15	0.15
17	0.11	0.16	0.14	18	0.12	0.13	0.12
19	0.12	0.13	0.12	20	0.31	0.31	0.31
21	0.24	0.26	0.25	22	0.17	0.20	0.18
23	0.17	0.17	0.17	24	0.09	0.13	0.11
25	0.19	0.19	0.19	26	0.07	0.06	0.07
27	0.28	0.27	0.27	28	0.22	0.19	0.21
29	0.09	0.13	0.11	30	0.04	0.04	0.04

every person, where some subjects showed comparatively observable nostril movement while asking to take a deep breath, however others motion are invisible or non-significant. Therefore, the hypothesis was proposed that the detectable movement of the nostrils can be powerfully used to categorize subjects into groups.

Our proposed pre-processing technique of nostril motion classification technique following the respiratory waveform extraction module is based on the detectable number of both peak points and bottom spots of a given breathing waveform using the Peakdet algorithm [41] with fixed threshold values (i.e., the pre-define values used to detect and determine the local maximum and local minimum points, the thresholds of detectable number of peaks and bottoms to classify movement and non-movement groups).

Considering the signal noise, which mostly exists in real-life signals, accidentally lead to the zero values of its first derivate, the well-known zero-derivate approach will result in false peak or bottom detection. A better solution employed by Peakdet' is to realize that for each local maximum point there must exist lower value points around it, and the algorithm is to discover the highest point X, around which there are points lower than X on both sides [5]. On the other hand, the local bottom point Y locates on the valley that some higher value points align on both sides of Y.
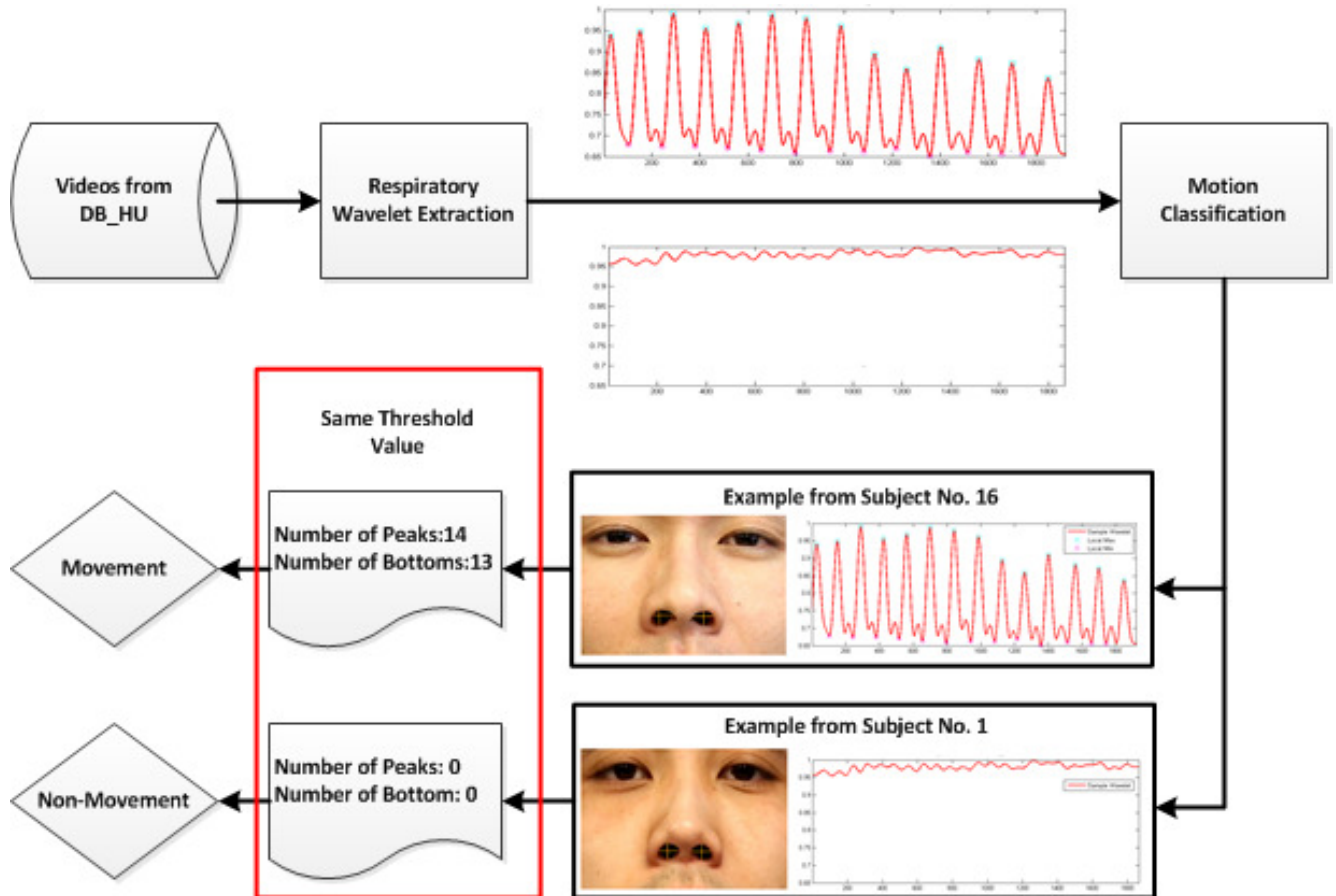
Figure 4.6: Illustration of how our nostril motion classifier that categorizes input subjects into movement and non-movement groups. 2 examples are selected from subject no.16 (session no.1, video no.1) and subject no.1 (session no.1, video no.3). The threshold value for detecting peak and bottom point is 0.15 and the number of local maximums (as well as minimums) used for classification is 4.

Fig. 4.6 illustrates the designed nostril movement classification process, and presents the two examples from both motion group and non-motion group. The defined motion classifier worked on both sessions (i.e., gallery and probe sessions) and categorized subjects from each session into two classes: Movement Group (MG) and Non-Movement Group (NG). It can be biased and mostly introduce some error to determine the subject's group by only analyzing one input curve. Considering for each subject there are five distinct data sets of the same dimensions in each unit (i.e., unit is defined as a group of data for one subject in one session), the motion classification output is referred to the majority-voting results of five input curves within the specific unit. There can be 6 different types of combined results, and the detailed decision making procedure for our motion classifier is listed in Table 4.2.

Table 4.2: The decision making procedure of the motion classifier applied on each unit of 5 videos. And the final decision is referred to the classified group for the corresponding subject in one session.

# of Motion Decision	# of Non-Motion Decision	Final Decision
5	0	Motion
4	1	Motion
3	2	Motion
2	3	Non-Motion
1	4	Non-Motion
0	5	Non-Motion

As a result of the motion classification approach, both the gallery and probe datasets as well as their corresponding subjects are segmented and categorized into two smaller groups. As a pre-processing methodology, it can significantly reduce the computational complexity and increase the matching efficiency during the identity recognition stage, since instead of comparing with the whole population of gallery database it only need to conduct the matching algorithm within a reduced group. Besides comparing with fewer subjects will also decrease the probability of matching the false target.

4.3 Feature Extraction and Clustering

Another pre-processing methodology might be introducing other schema that can further narrow down the number of potential candidates as well as reduce the matching complexity in the recognition procedure. Through observation and analysis of the input respiratory waveforms, certain breathing cycles are presented as the repeated patterns, which is quite similar to the human electrocardiogram (ECG) patterns and features that can successfully identify different subjects.

The following content provides more details about the feature definition and clustering procedure involved in the second pre-processing methodology that introduces another method to further narrow down the group size following the motion classification approach.

4.3.1   Feature Extraction

Inspired by the studies of human ECG identification, we proposed the hypothesis that some crucial features can be extracted from breathing waveforms to help categorize subjects with high feature similarity. Referred to the defined feature vector used for ECG classification [14][37][42][43], 4 dynamic features are selected within each breathing cycle time (i.e., one inspiration and its following expiration) shown in Fig. 4.7. Besides, the extraction of each respiratory cycle is based on the consecutive two detected local bottom points with pre-defined threshold values.

Considering the integrality of a 60-second long respiratory waveform, instead of directly using the four feature values extracted from only one periodic time, our defined feature vector is consisted of the averaged attribute values among detectable integrated breathing cycles. In addition, two more static features are also involved, which are defined as the ratio of horizontal distance and vertical distance for both left nostril and right nostril and extracted from facial images to complement the target feature vector representing one input breathing waveform. The detailed definition about each attribute is defined as follows:
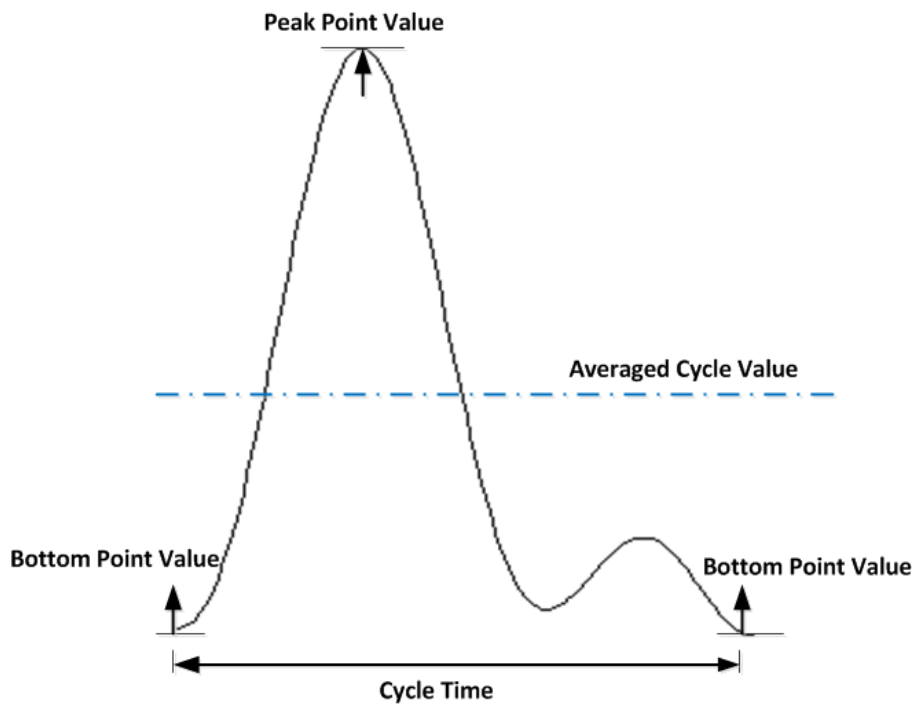
Figure 4.7: Definition of 4 dynamic features for a cycle of breathing waveform. Notice that its a real respiratory cycle extracted from the output of our breathing waveform extraction system for subject No.16.

- **F1:** Averaged ratio value of horizontal distance and vertical distance for left nostril in the first 10 frames: $f_1 = \frac{1}{10} \sum_{k=1}^{10} \frac{W_{left}}{H_{left}}$

- **F2:** Averaged ratio value of horizontal distance and vertical distance for right nostril in the first 10 frames: $f_2 = \frac{1}{10} \sum_{k=1}^{10} \frac{W_{right}}{H_{right}}$

- **F3:** Averaged breathing cycle period including one inspiratory and its expiratory in 60-second.

- **F4:** Averaged detectable local peak value in 60-second.

- **F5:** Averaged detectable local bottom value in 60-second.

- **F6:** Averaged point value for one waveform in 60-second.

4.3.2 Clustering

By using feature extraction as just mentioned, for each subject five distinct feature vectors are generated for both gallery and probe sessions. Since the produced feature sets are of the same dimension (i.e., each feature vector is defined as 1 by 6 dimension), the distance calculation between two feature sets can be applied to measure their similarity.

The method used to further cluster subjects after nostril motion classification is modified K-Nearest Neighbor (or K-NN for short) [31] approach. The goal of designed K-NN model is to discover the top similar subjects measured by feature vector distances. In our study, the K value is set to five. Generally speaking, the cluster treats five feature vectors from every subject in each session as a group, and in terms of each subject or group in the probe (i.e., session 2) the algorithm will find the top 5 subjects or groups from gallery databases (i.e., session 1) with the smallest averaged distance from the given input group from probe. Since for each group five feature vectors are included, the distance measurement between two groups X and Y should contain the comparison of each feature set from X and every feature vector from Y (i.e., totally 25 times of comparisons).

The detailed steps of distance measurement between 2 given groups are defined as follows:

- Group X includes five feature vectors; and group Y includes five feature sets of the same

dimension.

- Each vector v_i (i=1,2,...,5) from group X will compute the Euclidean Distance with all vectors v_j (i.e.,(j=1,2,...,5) from the group Y, and generate 25 results corresponding to each computation.

- Average all 25 distance values produced from step 2.

- Set the averaged value from step 3 to the measured distance between these two given groups X and Y.

Mathematically calculation procedure can be described as:

$$dist(v_i, v_j) = \sum_{k=1}^{6} (f_{ik} - f_{jk})^2 \tag{4.5}$$

$$dist\_value = \frac{1}{25} \sum_{i=1}^{5} \sum_{j=1}^{5} dist(v_i, v_j) \tag{4.6}$$

Where v_i and v_j (i=1,2,...,5 and j=1,2,...,5) are the feature vectors that come from the compared two groups that one comes from gallery and the other one is from probe database; and f_{ik} (k=1,2,...,6) is the corresponding one feature value at position k for specific feature vector v_i; and $dist(v_i, v_j)$ is the measured distance between feature vector v_i and v_j; and $dist\_value$ is the produced distance value between two input groups.

As a results of the second pre-processing methodology introduced above that combines motion classification and feature clustering, the number of potential subjects from gallery database that are identical to the testing subject from probe database is reduced to 5 instead of 30. Consequently it will markedly simplify the computational complexity during the identity recognition stage that only five pairs of images need to compare with.

4.4 Component-Based Face Recognition

During the video recording process, the subjects cannot keep the same head or body position for 60-second long. Besides, involving less facial features from images, measuring the deter-

ministic influence of respiratory waveforms is also our goal throughout the recognition system design. Therefore, instead of using the whole face image the component-based face recognition is employed, which can compensate for head pose changes and involving less facial biometric features as well.

The major steps of the recognition process are described below.

4.4.1 Face Image Selection

Five frames of face images are randomly selected and extracted from each fully frontal face video in database DB_FF as well as the head-up pose video in database DB_HU for each subject in every sessions (i.e., totally, for both gallery and probe image databases there are 150 face images involving 30 subjects). And all the images are in the same size of 1920 by 1080 pixel. Considering the missing information from the images (such as the eyes are closed or the facial expression changes quite a lot), some re-pick manipulations was done at random.

4.4.2 Face Image Geometric Normalization

Since subjects may change their head position or body pose during the 60-second long video shooting process, an image modification scheme was applied to all face images. The approach consisted of three main steps as follows:

- **Pupil detection:** First, the Masayuki Tanaka face parts detection algorithm [5] was applied to all the face image database (i.e., all the images generated from the face image selection procedure), where automated eye detection was performed using a template matching algorithm and the coordinates of pupils were generated automatically.

- **Image rotation:** After localizing the pupils' coordinates, the image rotation scheme was applied to compensate for slight left or right head move in the frontal pose. In order to achieve horizontally rearranged location for the center of pupils, the canonical faces were automatically constructed by applying an affine transformation based on the detected the pupil coordinates. Fig. 4.8 displays the image rotation procedure. In general, the rotation approach is based on the following rules:

– The coordinates of the center point between the two pupils keeps the same value.

– The distance between the two pupils, in other words the distances between the center point to left pupil and right pupil, keeps the same value.

– Since the images before and after the rotation are of the same size. In terms of the pixels that exceeds the boundary after rotation, they are not included in the new image. And for the region in the new image that none of the pixels from the old image will be projected in, they are set as black, i.e., set pixel value as (0, 0, 0). And all the other points keep the same pixel values throughout the rotation, but coordinates may change.
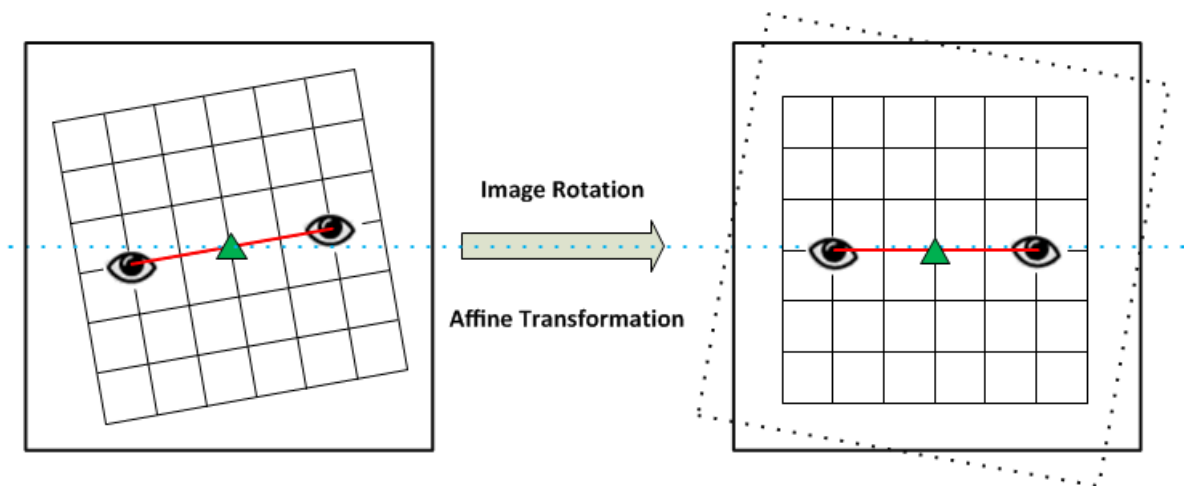
Figure 4.8: Illustration of applying affine transformation to obtain image rotation. Notice that during the rotation process, the center coordinates of the two pupils (i.e., shown as the green triangle point) will keep the same. While the coordinates of the pupils may change during the rotation, the distance between the two pupils will be the same value.

• **Image resizing:** After rotation, all the resulting images have the their two pupils aligned on the same horizontal line. Finally, the images are resized based on different ratio values in order to obtain fixed distance between two pupils (i.e., distance of 180 pixels for every image).

4.4.3 Component-Based Face Image Extraction

Our studies mainly focus on investigating the potential identity distinctness of human respiratory signals, thus the image databases used in the identity recognition procedure are concentrated on nose component where the component-based image databases are built surrounding the nose part.

The input videos are from both the frontal face pose database DB_FF and the head-up pose database DB_HU, and the generated image databases are six different types of combinations of facial components extracted from all the selected and processed face images as mentioned above (i.e., Face Image Selection and Image Reconstruction of Pupil Detection, Image Rotation and Image Resizing approaches). And more details, for each implemented component-based face image is listed in in Table 3.2.

After acquiring the location of eyes as well as the coordinates of pupils by applying the Masayuki Tanaka face parts detection algorithm [5], the segmentation process of facial components was automatically conducted with the pre-defined cutting locations for each combination (i.e., defining the specific width and height of the target image for each type of combination). Since the proportion of facial features may vary among different faces, the resulting facial part segmentations may contain other unnecessary features that may have some influence on the recognition performance, which is not expected in our experiments (such as the cutting of the nose component may include partial canthus structure). Therefore, in terms of each type of component-based face combination image, a specific covering strategy was applied on the segmentation output in order to eliminate the influence of other unnecessary features, and Table 3.2 also presents the specific strategies about the pre-determined covered regions for each type of facial parts combination.

Fig. 4.9 presents more details about each type of segmentation and its corresponding covering strategy, and displays a visual illustration about each type of facial components combination of the original output and the resulting segmentation images with pre-defined covering area. And in terms of the pixels within the specific regions that may contain other unnecessary facial components, their pixel values will be reset as (255, 255, 255) the covers the original structures and contents.
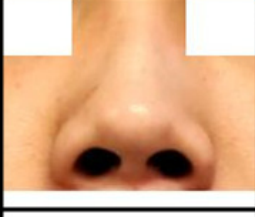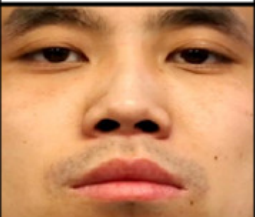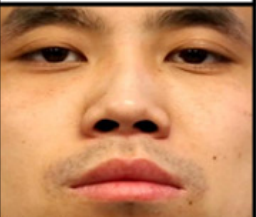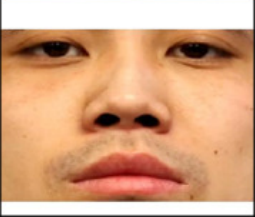
Figure 4.9: An example (subject no. 1) of the comparison between the original component-based face images and the resulting segmentation outputs with pre-defined covered regions for each type of combination. Notice that in order to show 6 different types of facial parts combination in this work, all the displayed images were resized to the same size on this figure instead of using their original size as listed.)

4.4.4 Face Image Photometric Normalization

Normalization of facial data is a crucial technology for human recognition application. The objective of applying face normalization approaches is to eliminate or alleviate the redundant, disturbed and unnecessary information such background, cloth, hair etc. in order to improve the performance of identity recognition.

In this study, 14 different types of face normalization approaches provided in Illumination Normalization techniques for robust Face recognition (INface) toolbox v2.0 [44][45] have been tested on all six kinds of component-based face feature combination. Fig. 4.10 represents the normalized images obtained based on 14 different types of normalization techniques applied (for one given sample image). More detailed information about each normalization approach is introduced as follows:

- **Wavelet Based Normalization Technique (WA):** The WA approach was proposed by Du and Ward [32], that applies the discrete waveform transform to an image and processes the obtained sub-bands. This technique focuses on the detailed coefficient matrices and employs histogram equalization to the approximate transform coefficients. Finally, reconstructs the normalized image using the inverse waveform transform after the manipulation of each individual sub-band.

- **Anisotropic Diffusion Based Normalization Technique (AS):** The AS approach adopts anisotropic smoothing of the input image to evaluate the luminance function, which was introduce by Gross et al. [46] to the face recognition area.

- **Discrete Cosine Transform Based Normalization Technique (DCT):** The DCT technique was introduced by Chen et al. [47] that truncates an appropriate number of DCT coefficients to minimize the illumination variations under different lighting conditions.

- **Gradientfaces Based Normalization Technique (GRA):** The GRA approach proposed by Zhang et al. [48] is to transform image into the gradient domain and use the generated face representation as the illumination invariant version of the target image.

- **Homomorphic Filtering Based Normalization Technique (HOMO):** The HOMO is to transform the input image into the logarithm followed by the frequency domain in order

to enhance the high-frequency components and weaken the low-frequency parts. Finally, apply the inverse Fourier transform to obtain the output image in the spatial domain [49].

- **Isotropic Diffusion Based Normalization Technique (IS):** The IS approach [49], is applied to estimate the luminance function of the input image using isotropic smoothing algorithm which is a simpler variance of the anisotropic diffusion based normalization technique [46].

- **Large-Scale and Small-Scale Features Normalization Technique (LSSF):** The LSSF proposed by Xie et al. [50], firstly computes the reflectance and luminance function of the input image and then further analyzes both generated functions using a second time of normalization. Within the INface toolbox used in our experiment, the SSR technique
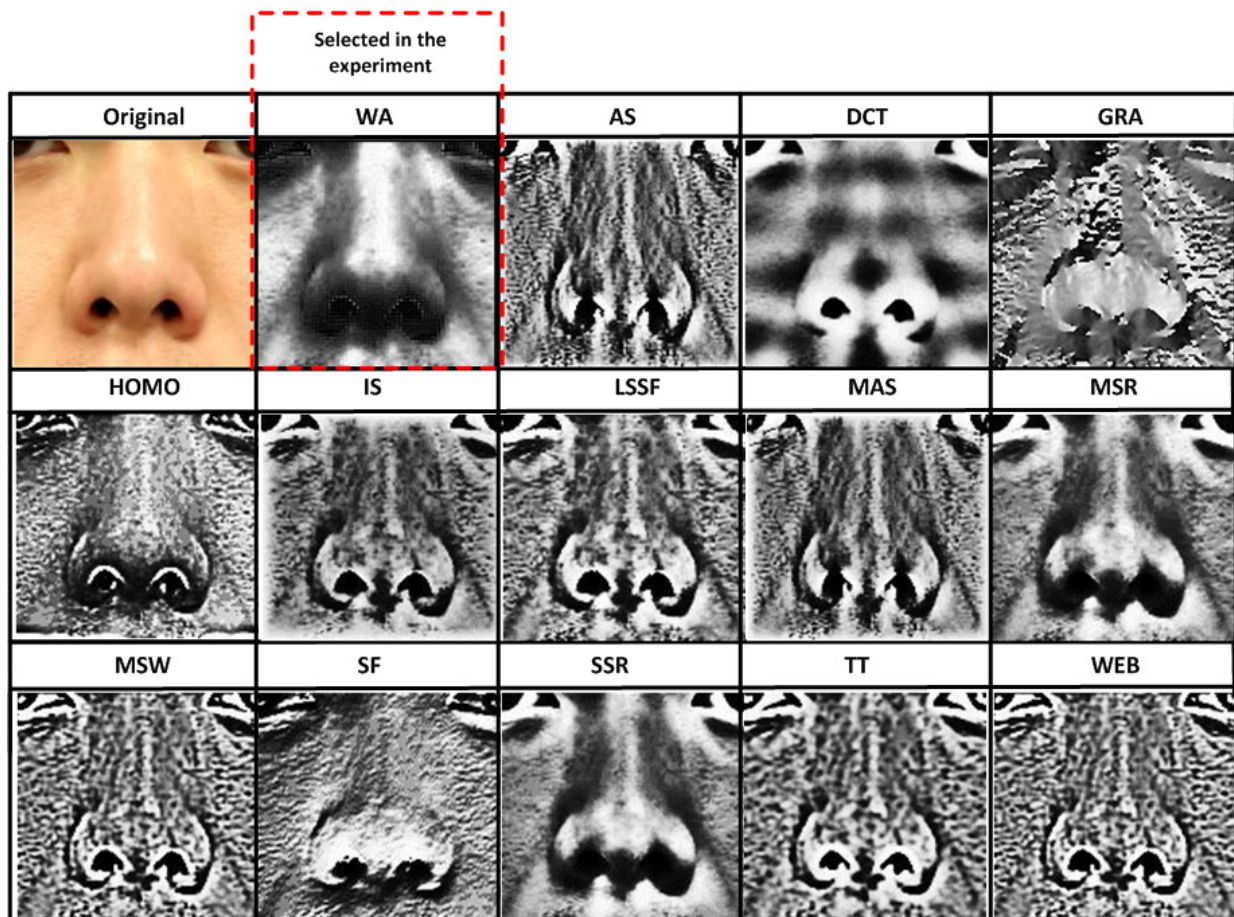
Figure 4.10: Illustration of the resulting images from the 14 types of face normalization approaches for the same example image of only nose component. And the WA technique was proven to provide the best result from identity recognition process and selected to apply to all facial component combinations.

is implemented as the normalization approach in both steps of LSSF technique.

- **Modified Anisotropic Diffusion Normalization Technique (MAS):** The MAS approach included two main modification into the original anisotropic diffusion normalization technique [46]: First, introducing an additional atan function to estimate the local contrast. Second, apply a robust post-processing procedure proposed by Tan et al. [34] in the final stage of this technique.

- **Multi-Scale Retinex (MSR) Algorithm:** The MSR method is to extend the previously designed single-scale center/surround retinex technique to a multi-scale version proposed by Jobson et al. [51].

- **Multi-Scale Weberfaces Normalization Technique (MSW):** The MSW is an extend the single-scale Weberfaces approach proposed by Wange et al. [52]. The technique is used to compute the relative gradient using a modified Weber contrast method for diverse neighborhood sizes and to apply a linear combination of the produced face representation as an illumination invariant version of the target output.

- **Steerable Filter Based Normalization Technique (SF):** The SF approach produces the target normalized image by removing illumination induced appearance variation from the input facial image using steerable filters.

- **Single Scale Retinex (SSR) Algorithm:** The SSR approach was proposed by Jobson et al. [53] on the basis of the retinex theory [54] as the majority of photometric normalization techniques.

- **Tan and Triggs Normalization Technique (TT):** The TAT is to employ a processing chain on the input image by firstly using gamma correction, then applying DoG filtering and finally adopting a robust post-processor to generate the output normalized image [34].

- **Single Scale Weberfaces Normalization Technique (WEB):** The WEB method is used to compute the relative gradient using a modified Weber contrast algorithm and treat the generated face representation as an illumination invariant version of the target image [52].

Based on the generated experimental output, the waveform based normalization technique (WA) provided the best experimental results in the identity recognition procedure, and the WA normalization technique was applied to all component-based face images from databases DB_FF1~DB_FF6 and DB_HU1~DB_HU6. The WA approach was proposed by Du and Ward [32], which applies the discrete waveform transform to an input image and processes the obtained sub-bands. This technique focuses on the detailed coefficient matrices and employs histogram equalization to the approximate transform coefficients. Finally, our proposed approach reconstructs the normalized image using the inverse waveform transform after the manipulation of each individual sub-band.

4.4.5 Human Recognition based on Sub-facial Regions

The standard texture based face recognition methods, Local Binary Pattern (LBP) and Local Ternary Pattern (LTP), were employed to perform the component-based face recognition experiments. The LBP was introduced by Ojala et al [55], is an effective gray-scale invariant texture measure. In the simplest form, the LBP description of a given pixel is computed by thresholding the values of the 3×3 neighborhoods of the target pixel based on the value of the central pixel (i.e., each neighbor pixel greater than or equal to the central pixel value is set as 1, otherwise the new value is assigned 0), and these binary values are arranged to build a binary number. Finally, the resulting binary pattern (i.e., converts the results of the eight neighborhoods to a binary number) is interpreted as a decimal value. An example of this LBP code process is presented in Fig. 4.11.

The LBP operator employed in our experiments is generalized as considering a larger local neighborhood that the sampling points evenly space on a circle, which is centered at the target pixel to be labeled. Specifically the descriptor is defined as $LBP_{P,R}^{U^2}$, where R represents the a circle neighborhood radius and P refers to the number of sampling points located on the circle. The symbol U^2 means the uniform pattern because of the most frequently existence in our experiments (Since Ojala has observed that uniform patterns accounted for around 90% of all the patterns in their image datasets [56]). An LBP operator is named uniform, if there exists at most two bitwise transitions (i.e., changes from 0 to 1 or 1 to 0) when considering

the corresponding binary string as a circular [56]. For instance, the 8-bit strings 00110000 (0 transition) and 01111110 (2 transitions) are uniform patterns, while 01010101 (7 transitions) and 00011010 (4 transitions) are not.
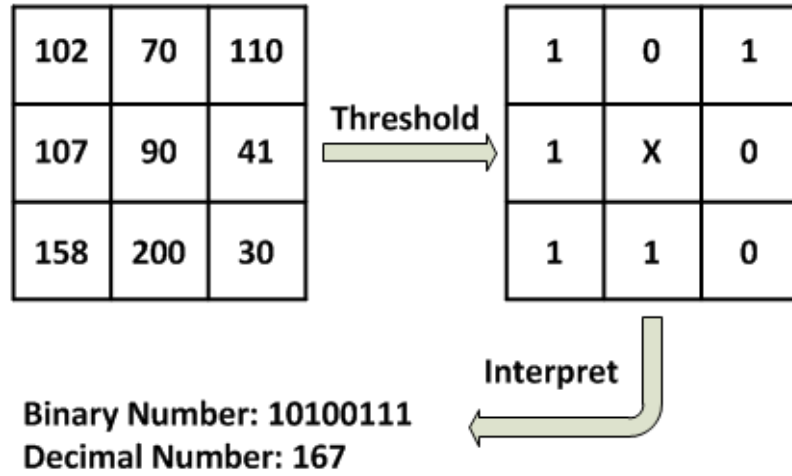
Figure 4.11: The LBP operator thresholds each pixel against its eight neighborhoods, and forms these binary patterns to a binary number (10100111), which is finally converted to a decimal value (167).

Formally, in terms of an input pixel at (x_a, y_a) the resulting LBP value can be interpreted in decimal form as follows [57]:

$$LBP_{P,R}(x_a, y_a) = \sum_{p=0}^{P-1} S(v_p - v_a)2^P \tag{4.7}$$

Where v_a and v_p (i.e., $p = 0, 1, \ldots, P - 1$) are the gray-level pixel values of the central pixel and its P surrounding pixels in the circle neighborhood with a radius R, 2^P is a binomial weight assigned to each $S(v_p - v_c)$, and the function $S(x)$ is mathematically defined as:

$$S(v_p - v_a) = \left\{ \begin{array}{l} 1, \text{ if } v_p - v_a \geq 0; \\ 0, \text{ if } v_p - v_a < 0. \end{array} \right\} \tag{4.8}$$

The LBP operator works as a powerful mean of texture descriptor. However, since the computation of the binary code is based on thresholding the center of the pixel region, the LBP tends to be sensitive to noise in homogeneous image regions [58].

Local ternary pattern (LTP) operator is introduced as an extension of LBP that employs a
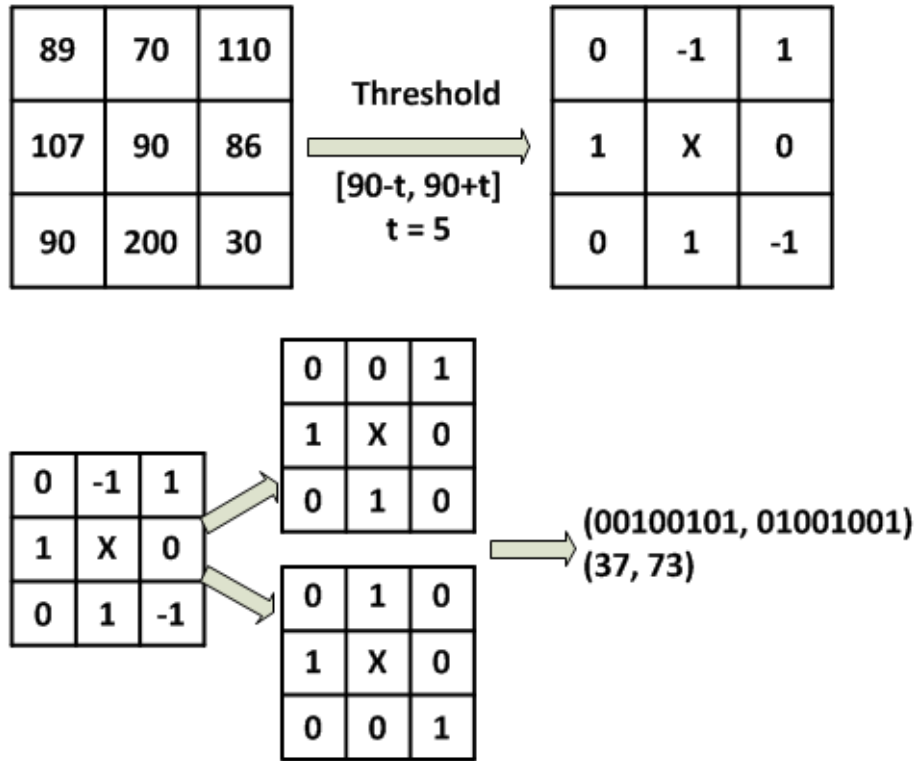
Figure 4.12: The LTP operator thresholds each pixel against its eight neighborhoods and assigns a 3-valued code of 0, 1 and -1. These trinary values are arrange to form a trinary number (0"-1"10"-1"101), and then transferred into two binary number (00100101, 01001001) and two decimal values(37, 73).

3-valued codes [34]: (1). the gray-level pixel values in an interval value of $\pm t$ around central pixel value are quantized to zero; (2) the pixel values above that interval are assigned to +1; and (3) the pixel values below it are set as -1, i.e. the function $S(x)$ is alternatively expressed as:

$$S(v_p - v_a, t) = \begin{cases} 1, & \text{if } v_p - v_a \geq t; \\ 0, & \text{if } |v_p - v_a| \leq t; \\ -1, & \text{if } v_p - v_a \leq -t. \end{cases} \qquad (4.9)$$

The threshold interval value t can be adjusted to generate diverse patterns, which makes the LTP descriptor become more tolerant to noise. A coding scheme is applied to split each trinary number into two separate channels of LBP descriptors, i.e. the positive one and the negative one. Finally the feature vector is formed by concatenating these two histograms of these two decimal values. An example of this LBP code process is presented in Fig. 4.12.

Consequently, the identification performance of our designed system shown in Fig. 4.1 is

evaluated through the Calculative Match Characteristic (CMC) curve, which measures the 1:

m recognition performance and evaluates the ranking capability of the system [58].

Chapter 5

Experimental Results and Discussion

The component-based face recognition experiments worked on the databases of all facial parts combinations (i.e. databases of DB_HU1~DB_HU6 and DB_FF1~DB_FF6) for both WA-normalized and non-normalized component-based face images under the following three designed scenarios, and Fig. 5.1 displays the detailed steps for each experiment.

- **Original Images (OI):** Original face parts images without any pre-processing manipulations. The face recognition algorithm worked on both WA-normalized and non-
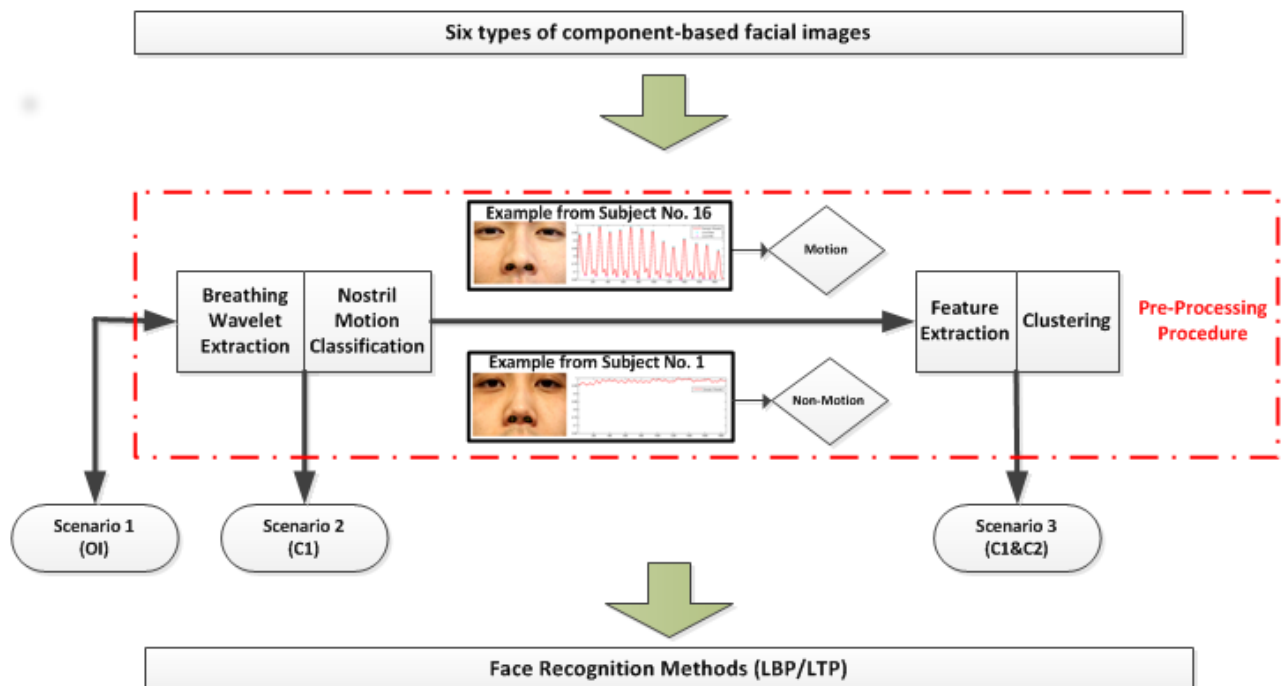
Figure 5.1: Illustration of the process flows for three scenarios applied in our study.

normalized facial parts combination images.

- **Motion Classification (C1):** The pre-processing technique only contains motion classification method. Firstly, subjects are categorized into two groups (i.e. motion and non-motion groups) based on the attributes of the extracted respiratory waveforms, then face recognition algorithm is applied separately on each group of both WA-normalized and original facial parts combination images.

- **Motion Classification and Feature Clustering (C1&C2):** The pre-processing technique contains both motion classification and feature clustering approaches. Firstly, subjects are categorized into two groups. Then, for each subject, the top 5 most similar entities via feature clustering are found, finally face matching method is applied on each cluster of the 5 candidates in both WA-normalized and non-normalized form of facial components combination images.

5.1 Motion Classification Results (C1)

For all waveforms generated from our respiratory waveform extraction system, which worked on every videos from database DB_HU, the motion classifier was initially applied as to categorize input data sets representing waveforms in both sessions from identical subjects into same group with accuracy of 100%. The detailed classification output for each subject can be found in Table 5.1.

Table 5.1: Output of two categorizations (movement and non-movement group) from nostril movement classification for each subject in both sessions of database DB_HU.

Group	# of Subjects	Subjects' ID
Motion Group (MG)	14	6, 7, 8, 11, 12, 14, 16, 17, 18, 19, 22, 23, 24, 29
Non-Motion Group (NG)	16	1, 2, 3, 4, 5, 9, 10, 13, 15, 20, 21, 25, 26, 27, 28, 30

As a result of applying the first pre-processing technique, in the recognition process instead of comparing with 30 subjects in gallery, for each subject it only need to compare with half of

the original size. Specifically speaking, for the subject in the motion group, it has to match with 14 potential subject, while in the other group, the recognition algorithm is required to apply 16 times for each subject.

5.2 Feature Clustering Results (C1&C2)

After categorizing subjects into 2 group, extracted feature vectors for each waveform was used to manipulate further clustering using K Nearest Neighbor (i.e. K is set as 5). As a result, for each subject in probe session the cluster will generate an output consisted of the top 5 most similar subjects in gallery session. The accuracy of obtaining the identical subject in the output cluster is shown in Table 5.2.

Table 5.2: Accuracies of finding the identical subject in the output group after employing feature vector clustering.

Group Name	Accuracy(%)
Motion Group (MG)	100
Non-Motion Group (NG)	93.75
All Subjects	96.67

Consequently, the second pre-processing technique can further improve the recognition efficiency that instead of comparing with about 15 subjects after the motion classification method, it only requires to conduct the comparison with 5 subjects. However, from the results in Table 5.2 we can see that the second pre-processing methodology fails to provide 100% accuracy, which may reduce the system's recognition performance. In the following subsection we can clearly see the comparison between the first and the second pre-processing approaches.

5.3 Experimental Results for Three Scenarios

The experimental results generated from LBP descriptor are comparative to the output produced from using LTP approach. Therefore, in the following content, we just discuss the LBP descriptor's performance.

The CMC performance (LBP approach) for scenarios 1-3 across all databases (DB_HU1~DB_HU6 and DB_FF1~DB_FF6) for 2 types of images (waNorm and noNorm) is illustrated in Fig. 5.2 and Fig. 5.3. In addition the performance measured by rank-1 identification rate of the proposed 3 experiments (i.e., OI, C1 and C1&C2) are presented in Table 5.3.

Table 5.3: Rank 1 identification rates when utilizing CMC curves (LBP) for OI, C1, and C1&C2 scenarios of both WA-normalized (waNorm) and Non-normalized (noNorm) matching.

| | *DB_HU* | | | | | | |
	Scenario	DB_HU1	DB_HU2	DB_HU3	DB_HU4	DB_HU5	DB_HU6
	OI	0.80	0.93	0.89	0.95	0.91	0.92
noNorm	C1	0.84	0.95	0.90	0.95	0.91	0.93
	C1&C2	0.87	0.95	0.89	0.97	0.92	0.94
	OI	0.82	0.92	0.89	0.99	0.92	0.96
waNorm	C1	0.85	0.93	0.89	0.99	0.93	0.96
	C1&C2	0.89	0.94	0.90	0.97	0.95	0.95
	DB_FF						
	Scenario	DB_FF1	DB_FF2	DB_FF3	DB_FF4	DB_FF5	DB_FF6
	OI	0.93	0.96	0.97	0.97	0.97	0.97
noNorm	C1	0.96	0.98	0.97	0.97	0.97	0.97
	C1&C2	0.93	0.95	0.93	0.94	0.93	0.93
	OI	0.89	0.97	0.96	0.99	0.99	1
waNorm	C1	0.95	0.98	0.97	0.99	0.99	1
	C1&C2	0.93	0.95	0.97	0.97	0.96	0.97

The experimental results indicate that human identify recognition process involving only nose component under uncontrolled situation (i.e. database DB_HU1 collected under the condition that the illumination may vary under different collection days and the head pose as well as the body position may change during the 60-second video shooting process) is a very challenging problem, since the component-based face image identification rate at rank-1 for original facial parts image (i.e., scenario OI, baseline scenario) of head-up pose is 80% for non-normalized form and 82% after applying WA normalization technique, when both gallery and probe images were randomly acquired from videos under possibly variant illumination situation and uncontrolled head pose changing while taking videos in diverse days. However, when
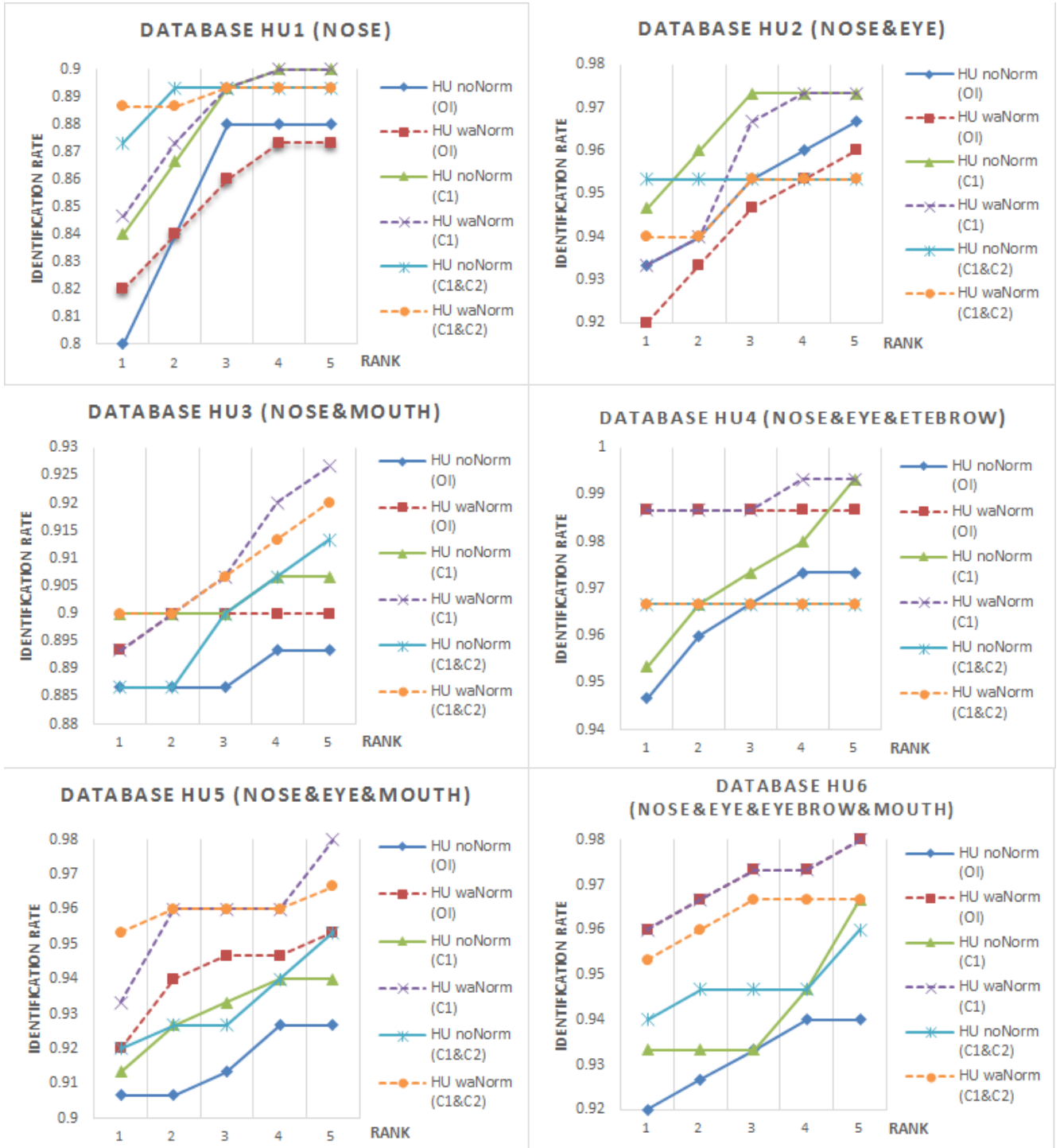
Figure 5.2: CMC curves comparing the performance of 3 types of scenarios (i.e., OI, C1 and C1&C2) for head-up face images of 6 different facial parts combinations (DB_HU1~DB_HU6).

using our proposed pre-processing technique of applying motion classification approach (i.e., scenario C1), the identification rate is increased to 84% for the original facial parts images and 85% with conducting WA-normalization manipulation. Moreover, in terms of the sec-
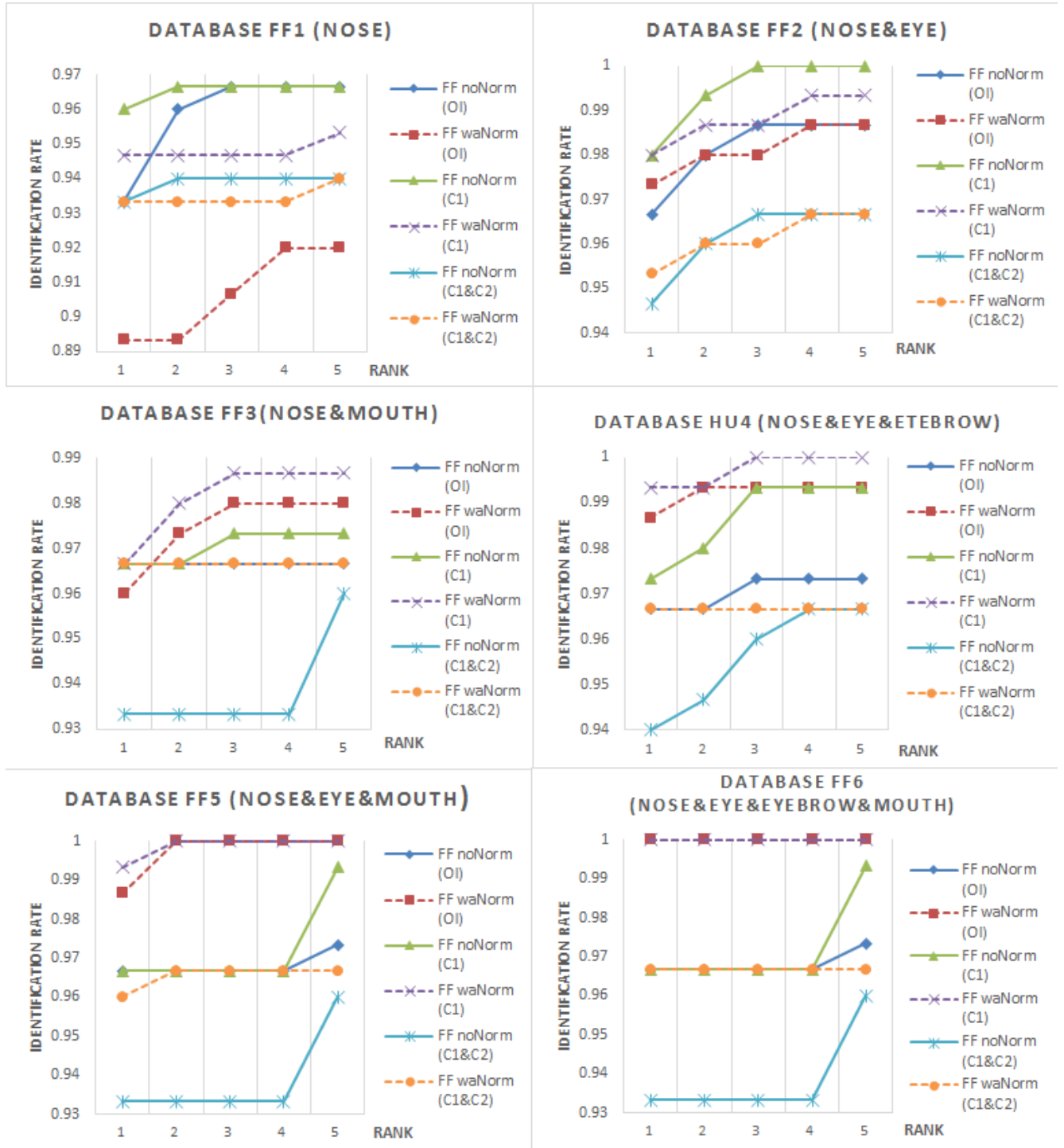
Figure 5.3: CMC curves comparing the performance of 3 types of scenarios (i.e., OI, C1 and C1&C2) for fully frontal face images of 6 different facial parts combinations (DB_FF1∼DB_FF6).

ond pre-processing technique, the feature clustering approach is respectively applied on the two classified categorizations (i.e., scenario C1&C2) produced from motion classification, the identification rate at rank-1 improves about 7%, resulting in 87% accuracy for original input

and 89% accuracy with WA-normalization manipulation compared to baseline scenario.

In the other databases (DB_HU2∼DB_HU6), more facial biometric features (i.e., eye, eyebrow and mouth) are included besides nose component that form other five different facial parts combinations. Although the baseline experimental results are increased to over 90% matching accuracy at uncontrolled condition, our feature clustering approach in combination with motion classification technique (i.e. the second pre-processing technique tested in scenario C1&C2) can still improve the identification rate (at rank-1) by approximately 2%. However, for several experimental outputs (i.e. database DB_HU3 of noNorm case, DB_HU4 and DB_HU6 of waNorm case), the feature clustering approach from scenario 3 unfortunatelly reduce the identification rate compared with the first pre-processing technique that only applys the motion classificaiton approach. It mainly accounts for the clustering errors introduced from the feature clustering method.

Under the semi-controlled condition (DB_FF1∼DB_FF6), when the head pose were fixed to fully frontal position but the illumination status were still dependent on the natural light of the video collection date. Even though the component-based face image identification rates of the baseline scenario (i.e., Scenario OI) for all types of facial-feature combinations in both WA-normalized and original forms are observably increased compared to the baseline situation of uncontrolled condition, our proposed motion classification technique can still provide good performance. Especially for database DB_FF1 only containing the nose information, the advantage of using motion classification is that it performs even better than baseline performance by approximately 3% of non-normalization and 6% of WA-normalized form. However, in this semi-controlled condition (i.e. the head pose is fixed to fully frontal, while the illumination situation may still be variant), the feature clustering technique from the second pre-processing technique failed to perform well. The main problem is that although the motion classification can produce 100% accuracy, the following feature clustering approach still introduces some error (i.e., overall accuracy is 97% with 100% accuracy in motion group and 94% for non-motion group) which will lead to the output identification rate as best as 97%.

Overall, our proposed video-based human respiratory waveform extraction and component-based face image recognition system can achieve significantly increased identification rates

across all datasets applied for both original and WA-normalized facial parts images, and it manages to obtain a markedly high identification accuracy (ranging from 96%-100% when using a single or a combination of facial components), yielding an average of 7% raise, when compared to the baseline scenario. To the best of our knowledge, this is the first time that a biometric system is composed of an important human vital feature (human breathing waveform) that is fused with facial features in such an efficient manner.

Chapter 6

Conclusions and Future Work

In this work our focus was on investigating the problem of human identity recognition in un-controlled environments with the help of respiratory waveforms extracted from videos, when taking advantage of a medical condition known as nasal flaring. Specifically we have studied the human facial components identification tasks under several pre-processing techniques. The breathing waveform produced by our proposed respiratory waveform extraction system is comparative to the ADInstruments device measured output for some portion of subjects showing nostril flaring symptom. The proposed motion classification technique can successfully categorize input waveforms into two groups with 100% accuracy. Notice that the defined accuracy is to classify the waveforms in both data collected sessions from the same subjects into the same group. In addition, our designed feature clustering technique, combined with our proposed motion classification approach, can locate the same subject within the top 5 candidates with an accuracy of 97% (i.e., 100% accuracy for motion group and 94% accuracy for non-motion group), which results in significant reduction on the number of comparisons during the identity recognition procedure.

For the purpose of the this work, 3 different scenarios (i.e., Original Image, Motion Classification and Feature Clustering) were designed and tested on 6 types of databases (DB_HU1∼DB_HU6 and DB_FF1∼DB_FF6) of diverse facial components combinations in both non-normalized and waveform Based (WA) normalized form under semi-controlled and uncontrolled indoor environments. The experimental results indicate that, across all datasets used, the application of mo-

tion classification and feature clustering improves facial components recognition performance. Especially when only investing the nose information, our proposed system can significantly increase the rank-1 identification rate by approximately 7% for both original and WA-normalized datasets. The proposed respiratory waveform recognition technique shows particularly good performance on the baseline scenarios with low accuracy under uncontrolled conditions.

Our future plans are to develop an improved respiratory waveform extraction approach that will successfully detect and enhance the slight movement of nostrils. In addition, the experiment video databases will be acquired using cellphone cameras with diverse resolution under uncontrolled conditions, such as varying luminous condition and unfixed head position. This is expected to result in as good as or even improved identification performance.

Bibliography

[1] Adinstruments. `http://www.adinstruments.com/`.

[2] Wikimedia.org Public Domain images and Wikipedia.org. Human ecg picture. `http://a-fib.com/treatments-for-atrial-fibrillation/diagnostic-testing/the-ekg-signal/`.

[3] Thirimachos Bourlai, Cameron Whitelam, and Ioannis Kakadiaris. Pupil detection under lighting and pose variations in the visible and active infrared bands. In *Information Forensics and Security (WIFS), 2011 IEEE International Workshop on*, pages 1–6. IEEE, 2011.

[4] Neil K. Kaneshiro and David Zieve. Nasal flaring. `http://www.nlm.nih.gov/medlineplus/ency/imagepages/17279.htm`. MedlinePlus.

[5] Masayuki Tanaka. Face parts detection algorithm. `http://www.mathworks.com/matlabcentral/fileexchange/36855-face-parts-detection`, 2012. Updated in 2014.

[6] Anil K Jain, Arun Ross, and Sharath Pankanti. Biometrics: a tool for information security. *Information Forensics and Security, IEEE Transactions on*, 1(2):125–143, 2006.

[7] Anil K Jain, Ruud Bolle, and Sharath Pankanti. *Biometrics: personal identification in networked society*. Springer, 1999.

[8] Davide Maltoni, Dario Maio, Anil K Jain, and Salil Prabhakar. *Handbook of fingerprint recognition*. springer, 2009.

[9] Anil K Jain and Stan Z Li. *Handbook of face recognition*. Springer, 2005.

[10] Raul Sanchez-Reillo, Carmen Sanchez-Avila, and Ana Gonzalez-Marcos. Biometric identification through hand geometry measurements. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10):1168–1171, 2000.

[11] John Daugman. How iris recognition works. *Circuits and Systems for Video Technology, IEEE Transactions on*, 14(1):21–30, 2004.

[12] D Srinivasan, WS Ng, and AC Liew. Neural-network-based signature recognition for harmonic source identification. *Power Delivery, IEEE Transactions on*, 21(1):398–405, 2006.

[13] Simon Liu and Mark Silverman. A practical guide to biometric security technology. *IT Professional*, 3(1):27–32, 2001.

[14] Masaki Kyoso and Akihiko Uchiyama. Development of an ecg identification system. In *Engineering in Medicine and Biology Society, 2001. Proceedings of the 23rd Annual International Conference of the IEEE*, volume 4, pages 3721–3723. IEEE, 2001.

[15] JM Irvine, BK Wiederhold, LW Gavshon, SA Israel, SB McGehee, R Meyer, and MD Wiederhold. Heart rate variability: a new biometric for human identification. In *Proceedings of the International Conference on Artificial Intelligence (IC-AI01)*, pages 1106–1111, 2001.

[16] Steven A Israel, W Todd Scruggs, William J Worek, and John M Irvine. Fusing face and ecg for personal identification. In *Applied Imagery Pattern Recognition Workshop, 2003. Proceedings. 32nd*, pages 226–231. IEEE, 2003.

[17] Steven A Israel, John M Irvine, Andrew Cheng, Mark D Wiederhold, and Brenda K Wiederhold. Ecg to identify individuals. *Pattern recognition*, 38(1):133–142, 2005.

[18] A Travaglini, C Lamberti, J DeBie, and M Ferri. Respiratory signal derived from eight-lead ecg. In *Computers in Cardiology 1998*, pages 65–68. IEEE, 1998.

[19] Tiezhi Zhang, Harry Keller, Matthew J OBrien, Thomas R Mackie, and Bhudatt Paliwal. Application of the spirometer in respiratory gated radiotherapy. *Medical physics*, 30(12):3165–3171, 2003.

[20] Michael K Marks, Michael South, and Bradley G Carter. Measurement of respiratory rate and timing using a nasal thermocouple. *Journal of clinical monitoring*, 11(3):159–164, 1995.

[21] Robert Dean Allison, EL Holmes, and J Nyboer. Volumetric dynamics of respiration as measured by electrical impedance plethysmography. *Journal of applied physiology*, 19(1):166–173, 1964.

[22] Hideo D Kubo and Bruce C Hill. Respiration gated radiotherapy treatment: a technical study. *Physics in medicine and biology*, 41(1):83, 1996.

[23] Bernd Heisele, Purdy Ho, and Tomaso Poggio. Face recognition with support vector machines: Global versus component-based approach. In *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, volume 2, pages 688–694. IEEE, 2001.

[24] Bernd Heisele and Takamasa Koshizen. Components for face recognition. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 153–158. IEEE, 2004.

[25] Johan AK Suykens and Joos Vandewalle. Least squares support vector machine classifiers. *Neural processing letters*, 9(3):293–300, 1999.

[26] Jennifer Huang, Bernd Heisele, and Volker Blanz. Component-based face recognition with 3d morphable models. In *Audio-and Video-Based Biometric Person Authentication*, pages 27–34. Springer, 2003.

[27] Bernd Heisele, Thomas Serre, Massimiliano Pontil, Thomas Vetter, and Tomaso Poggio. Categorization by learning and combining object parts. In *NIPS*, pages 1239–1245, 2001.

[28] Nnamdi Osia and Thirimachos Bourlai. Holistic and partial face recognition in the mwir band using manual and automatic detection of face-based features. In *Homeland Security (HST), 2012 IEEE Conference on Technologies for*, pages 273–279. IEEE, 2012.

[29] Nnamdi Osia and Thirimachos Bourlai. A spectral independent approach for physiological and geometric based face recognition in the visible, middle-wave and long-wave infrared bands. *Image and Vision Computing*, 2014. (2014 in press).

[30] John Francis Kenney and E. S. Keeping. Root mean square. In *Mathematics of statistics*, pages 59–60. Princeton, NJ: Van Nostrand, 3 edition, 1962.

[31] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[32] Shan Du and Rabab Ward. Wavelet-based illumination normalization for face recognition. In *Image Processing, 2005. ICIP 2005. IEEE International Conference on*, volume 2, pages II–954. IEEE, 2005.

[33] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. In *Computer vision-eccv 2004*, pages 469–481. Springer, 2004.

[34] Xiaoyang Tan and Bill Triggs. Enhanced local texture feature sets for face recognition under difficult lighting conditions. *Image Processing, IEEE Transactions on*, 19(6):1635–1650, 2010.

[35] Rudi Hoekema, Gerard JH Uijen, and Adriaan van Oosterom. Geometrical aspects of the interindividual variability of multilead ecg recordings. *Biomedical Engineering, IEEE Transactions on*, 48(5):551–559, 2001.

[36] Biju P Simon and C Eswaran. An ecg classifier designed using modified decision based neural networks. *Computers and Biomedical Research*, 30(4):257–272, 1997.

[37] Lena Biel, Ola Pettersson, Lennart Philipson, and Peter Wide. Ecg analysis: a new approach in human identification. *Instrumentation and Measurement, IEEE Transactions on*, 50(3):808–812, 2001.

[38] Tsu-Wang Shen, WJ Tompkins, and YH Hu. One-lead ecg for identity verification. In *Engineering in Medicine and Biology, 2002. 24th Annual Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint*, volume 1, pages 62–63. IEEE, 2002.

[39] Neil K. Kaneshiro and David Zieve. Nasal flaring. `http://www.nlm.nih.gov/medlineplus/ency/article/003055.htm`, 2012. MedlinePlus.

[40] L. W. Wilkins. Sensory system. In *Lippincott Manual of Nursing Practice Series: Assessment*, pages 265–266. Lippincott Williams & Wilkins, 2007.

[41] E. Billauer. "peakdet"algorithm,. `http://www.billauer.co.il/peakdet.html`.

[42] Yongjin Wang, Foteini Agrafioti, Dimitrios Hatzinakos, and Konstantinos N Plataniotis. Analysis of human electrocardiogram for biometric recognition. *EURASIP journal on Advances in Signal Processing*, 2008:19, 2008.

[43] Foteini Agrafioti, Jiexin Gao, and Dimitrios Hatzinakos. Heart biometrics: theory, methods and applications. *Biometrics: Book*, 3:199–216, 2011.

[44] Vitomir Štruc and Nikola Pavešić. Gabor-based kernel partial-least-squares discrimination features for face recognition. *Informatica*, 20(1):115–138, 2009.

[45] Vitomir Štruc and Nikola Pavešic. Photometric normalization techniques for illumination invariance. *Advances in Face Image Analysis: Techniques and Technologies, IGI Global*, pages 279–300, 2011.

[46] Ralph Gross and Vladimir Brajovic. An image preprocessing algorithm for illumination invariant face recognition. In *Audio-and Video-Based Biometric Person Authentication*, pages 10–18. Springer, 2003.

[47] Weilong Chen, Meng Joo Er, and Shiqian Wu. Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain.

Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on, 36(2):458–466, 2006.

[48] Taiping Zhang, Yuan Yan Tang, Bin Fang, Zhaowei Shang, and Xiaoyu Liu. Face recognition under varying illumination using gradientfaces. *Image Processing, IEEE Transactions on*, 18(11):2599–2606, 2009.

[49] Guillaume Heusch, Fabien Cardinaux, and Sébastien Marcel. Lighting normalization algorithms for face verification. *IDIAP-com 05*, 3, 2005.

[50] Xiaohua Xie, Wei-Shi Zheng, Jianhuang Lai, Pong C Yuen, and Ching Y Suen. Normalization of face illumination based on large-and small-scale features. *Image Processing, IEEE Transactions on*, 20(7):1807–1821, 2011.

[51] Daniel J Jobson, Z-U Rahman, and Glenn A Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on*, 6(7):965–976, 1997.

[52] Biao Wang, Weifeng Li, Wenming Yang, and Qingmin Liao. Illumination normalization based on weber's law with application to face recognition. *Signal Processing Letters, IEEE*, 18(8):462–465, 2011.

[53] Daniel J Jobson, Z-U Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *Image Processing, IEEE Transactions on*, 6(3):451–462, 1997.

[54] Edwin H Land and John McCann. Lightness and retinex theory. *JOSA*, 61(1):1–11, 1971.

[55] Timo Ojala, Matti Pietikäinen, and David Harwood. A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1):51–59, 1996.

[56] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7):971–987, 2002.

[57] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen. Local binary patterns and its application to facial image analysis: a survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 41(6):765–781, 2011.

[58] Nathan D Kalka, Thirimachos Bourlai, Bojan Cukic, and Lawrence Hornak. Cross-spectral face recognition in heterogeneous environments: A case study on matching visible to short-wave infrared imagery. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8. IEEE, 2011.