Development of an Active Shape Model

Using the Discrete Cosine Transform

Kotaro Yasuda

A Thesis

in

The Department

of

Electrical and Computer Engineering

Presented in Partial Fulfillment of the Requirements

for the Degree of Master of Applied Science at

Concordia University

Montreal, Quebec, Canada

April 2014

Abstract

Development of an Active Shape Model
Using the Discrete Cosine Transform

Kotaro Yasuda

Facial recognition systems have been successfully applied in security, law-enforcement and human identification application, for automatically identifying a human in a digital image or a video frame. In a feature-based face recognition system using a set of features extracted from each of the prominent facial components, automatic and accurate localization of facial features is an essential pre-processing step. The active shape model (ASM) is a flexible shape model that was originally proposed to automatically locate a set of landmarks representing the facial features. Various improved versions of this model for facial landmark annotation have been developed for increasing the shape fitting accuracy at the expense of significantly increased computational complexity.

This thesis is concerned with developing a low-complexity active shape model by incorporating the energy compaction property of the discrete cosine transform (DCT). Towards this goal, the proposed ASM, which utilizes a 2-D profile based on the DCT of the local grey-level gradient pattern around a landmark, is first developed. The ASM is then utilized in a scheme of facial landmark annotation for locating facial features of the face in an input image. The proposed ASM provides two distinct advantages: (i) the use of a smaller number of DCT coefficients in building a compressed DCT profile

significantly reduces the computational complexity, and (ii) the process of choosing the low-frequency DCT coefficients filters out the noise contained in the image. Simulations are performed to demonstrate the superiority of the proposed ASM over other improved versions of the original active shape model in terms of the fitting accuracy as well as in terms of the computational complexity. It is shown that the use of the proposed model in the application of facial landmark annotation significantly reduces the execution time without affecting the accuracy of the facial shape fitting.

ACKNOWLEDGEMENT

I would like to express the deepest appreciation to my thesis supervisor, Dr. M. Omair Ahmad, for his guidance and support throughout the development of this thesis. Without his supervision, this thesis would not have been possible. I would also like to thank my family for their constant support for me to pursue my study at Concordia University.

To

my grandfather

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

x_i The x-coordinate of the ith landmark

y_i The y-coordinate of the ith landmark

$\mathbf{x}$ Vector containing the x- and y-coordinates of the n landmarks

$\bar{\mathbf{x}}$ Mean of the vectors containing the x- and y-coordinates of the n landmarks over the N training shape

$\mathbf{S}_s$ Covariance matrix of the deviations from the mean vector

$\mathbf{v}_k$ The kth eigenvector of $\mathbf{S}_s$

λ_k The kth eigenvalue of $\mathbf{S}_s$

$\mathbf{P}$ Matrix whose columns are the t eigenvectors corresponding to the first t largest eigenvalues

$\mathbf{b}$ Vector of weights applied to the t eigenvectors

$\mathbf{x}_s$ Target shape

T Similarity transform that transforms the model shape from the model space to the image space by scaling, rotating and translating the model shape

(x_c, y_c) Center point of a rectangular region containing a face

w Width of a rectangular region containing a face

h Height of a rectangular region containing a face

(t_x, t_y) Translation parameters

s Scaling parameter

θ Rotation parameter

$\mathbf{g_{ij}}$	Grey-level intensity gradient profile for the ith landmark of the jth shape
g_{ijk}	The kth element in $\mathbf{g_{ij}}$
$\hat{\mathbf{g}}_{\mathbf{ij}}$	Normalized grey-level intensity gradient profile for the ith landmark of the jth shape
$\mathbf{g_i}$	Grey-level intensity gradient profile for the ith landmark
$\bar{\mathbf{g}}_{\mathbf{i}}$	Mean grey-level intensity gradient profile for the ith landmark
$\mathbf{S_i}$	Covariance matrix of the deviations from the mean grey-level intensity gradient profile for the ith landmark
$\hat{\mathbf{g}}_{\mathbf{i}}$	Grey-level intensity gradient profile for the ith landmark
$\mathbf{g_{id}}$	Subset of $\hat{\mathbf{g}}_{\mathbf{i}}$ at a candidate position d
$\bar{\mathbf{g}}_{\mathbf{i}}$	Mean grey-level intensity gradient profile for the ith landmark
$f(\mathbf{g_{id}})$	Mahalanobis distance between the mean grey-level intensity gradient profile for the ith landmark and a candidate profile at a position d
$\mathbf{g_{id_{best}}}$	Subset of $\hat{\mathbf{g}}_{\mathbf{i}}$ at a best position d
(t_{xbest}, t_{ybest})	Best translation parameters
s_{best}	Best scaling parameter
θ_{best}	Best rotation parameter
$w(i, j)$	Response of a linear spatial filter
$\mathbf{G}$	Profile Matrix
$G(x, y)$	Filter response at the pixel position (x, y)
g_{ij}	The jth element of $\mathbf{g_i}$
g'_{ij}	Normalized jth element of $\mathbf{g_i}$

g_{ij}''	Normalized and equalized jth element of $\mathbf{g_i}$
$\mathbf{g_i''}$	Normalized and equalized grey-level intensity gradient profile for the ith landmark
$\mathbf{C_G}$	2-D array of the DCT coefficients representing the profile matrix $\mathbf{G}$
$\mathbf{Y_{C_G}}$	Vector containing a zig-zag sequence of $\mathbf{C_G}$
$\mathbf{Y}_{n_c}$	Vector containing the first n_c DCT coefficients representing the low-frequency components of the profile matrix $\mathbf{G}$
y_i	The ith element in $\mathbf{Y}_{n_c}$
y_i'	Normalized ith element in $\mathbf{Y}_{n_c}$
$\mathbf{Y}_{n_c}'$	Vector containing the normalized elements of $\mathbf{Y}_{n_c}$
y_i''	Normalized and equalized ith element in $\mathbf{Y}_{n_c}$
$\mathbf{Y}_{n_c}''$	Vector containing the normalized and equalized elements of $\mathbf{Y}_{n_c}$
$\mathbf{a_i}$	Vector containing the x- and y-coordinates of the ith landmark of the model shape
$\mathbf{b_i}$	Vector containing the x- and y-coordinates of the ith landmark of the target shape
d_i	Euclidean distance between two points given by $\mathbf{a_i}$ and $\mathbf{b_i}$
$E_{average}$	Average normalized fitting error

LIST OF ACRONYMS

ASM	Active shape model
PDM	Point distribution model
LGGM	Local grey-level gradient model
MRASM	Multi-resolution active shape model
STASM	Stacked active shape model
PCA	Principal component analysis
DCT	Discrete cosine transform
GPA	Generalized procrustes analysis
DFT	Discrete Fourier transform
FFT	Fast Fourier transform
MUCT	Milborrow / university of Cape Town
IMM	Informatics and mathematical modelling

CHAPTER 1

Introduction

1.1 Background

In recent advances in biometrics, facial recognition has been one of the promising research areas. The goal of facial recognition is to automatically identify a human in a digital image or a video frame using faces stored in a database. Facial recognition has been successfully applied in security, law-enforcement and human identification applications [1].

In general, a face recognition system consists of three stages, acquisition, feature extraction and classification. In the acquisition stage, a region containing a face is detected from an input image, and some pre-processing is performed on the detected region. In the feature extraction stage, a set of meaningful features is extracted from the detected region. A database of feature set is built in the training phase, using the feature sets extracted from a set of training images. In the classification stage, the distance between the feature set extracted from a test image and each feature set in the database is measured using a distance metric to identify the face in the test image.

Facial recognition systems can be categorized into classes of systems employing holistic approaches [1]-[9] and feature-based approaches [10]-[19]. In the holistic approach, the feature set is extracted from the entire face region in an image, whereas in the feature-based approach, the feature set is extracted from each of the prominent facial

components, such as eyes, nose and mouth. Since these facial components are treated separately, the feature-based approach is more robust to positional variations of the face in the image compared to the holistic approach [20]. This approach, however, relies heavily on an accurate localization of facial features. Thus, automatic and accurate localization of facial features is an essential pre-processing step in a feature-based facial recognition system.

Shape models capable of encompassing facial components could be useful for localizing facial features. In the literature, various approaches that model shape and shape variations have been proposed [21]-[31]. Active shape model (ASM) [31] is a flexible model that has widely been utilized in order to automatically locate a set of landmarks representing a target object in an image [32]-[38]. In an application of ASM in facial landmark localization, the facial features are represented by a set of landmarks, and ASM automatically locates these landmarks by fitting its model shape to a facial shape in an image.

1.2 A Brief Review of Active Shape Model

Active shape model (ASM) was first introduced in 1994 in [31], and has been successfully employed to various computer vision applications including locating organs in medical images, face recognition and hand-written character recognition [31]-[34]. It utilizes a point distribution model (PDM) and a local grey-level gradient model (LGGM) to iteratively fit the model shape generated by ASM to a target object in an image. The point distribution model is used for generating shapes similar to those in the training set, and the LGGM is built using 1-D profiles, each of which represents a local grey-level

gradient pattern along a line centered at a particular landmark of the model shape. The local grey-level gradient pattern is then utilized for searching a new position of each landmark along a line centered at the landmark. The model shape is iteratively reformed for its better fit to the target object in the image. The classical ASM performs well if the landmarks of the initial model shape are placed close to their targets. However, the initial model shape could be placed only roughly whereby the landmarks often get placed away from their targets. Thus resulting in long search lines and consequently making the target search computationally expensive. It could also distract the landmarks by local structures in the image.

In order to overcome the above drawbacks of the classical ASM, a multi-resolution approach of ASM, known as multi-resolution ASM (MRASM), is proposed in [34]. In this scheme, an ASM is first applied to a coarse image to roughly place the model shape near the target object, and then applied to finer images to refine the shape fitting. An image pyramid containing a set of images with different resolutions is used in order to build the LGGM for multiple image resolutions. For each level of the image pyramid, the corresponding LGGM is utilized for searching a new position of each landmark. Unlike the classical ASM, the multi-resolution approach requires only short search lines; thus reducing the risk of distraction of the landmarks by local structures in the image, and decreasing the computational complexity [34]. However, the 1-D profile used in the classical ASM and MRASM does not sufficiently represent the grey-level intensity gradient pattern around each landmark to distinguish the landmarks from one another [39]. Consequently, the search can converge to a local minimum and produces poor fitting results.

For capturing grey-level intensity gradient information around each landmark more accurately, 2-D profiles are introduced in [35]. A 2-D profile is built by sampling grey-level intensities of a square region around each landmark, and computing their intensity gradients in the x and y directions. This profile represents the intensity gradient information in a larger region around the landmark compared to that in the 1-D profile. Stacked active shape model (STASM), introduced in [35], proposes the use of two ASM searches sequentially, a search using the classical ASM with 1-D profiles to roughly place the model shape, and then a search using an ASM with 2-D profiles to refine the shape fitting. This approach yields relatively more accurate shape fitting results. However, the search can still converge to a local minimum [39]. Furthermore, the use of 2-D profiles significantly increases the computational complexity of the search.

In order to further improve the shape fitting accuracy, a model, referred to as an active shape model with a PCA-based LGGM, is proposed in [36]. It employs the principal component analysis (PCA) in the LGGM to model the variations of grey-level intensity gradients in a square region around each landmark. In this scheme, a candidate 2-D profile of a particular landmark is mapped onto a PCA subspace to obtain a set of projection coefficients, and a new profile is constructed by using this set of coefficients. The weighted reconstruction error between the new profile and the candidate profile is computed to determine a new position of the landmark [36]. The use of the PCA contributes to improving the shape fitting accuracy and to reducing the risk of the search converging to a local minimum [39]. However, the computational complexity of ASM with a PCA-based LGGM is much larger than that of the techniques using other versions of ASM, since the computation of the PCA is expensive.

Several other improved versions of ASM have also been reported in the literature. An improved ASM using the edge information is proposed in [37] in order to improve the shape fitting of the landmarks on the face contour by building better local texture models for the landmarks. The authors in [38] employ the color information to detect the center of the mouth and the eyes to improve the shape initialization of ASM. These enhancements contribute to slightly improving the shape fitting accuracy with a small increase in the computational complexity. However, the shape fitting accuracy is much lower than that obtained by the stacked ASM and ASM with a PCA-based LGGM.

1.3 Objective and Scope of the Thesis

As seen from the review of the active shape model and its modified versions, conducted in the previous section, the stacked active shape model as proposed in [35] provides a significant improvement in the shape fitting accuracy by using both the 1-D and 2-D profiles. The active shape model with a PCA-based LGGM developed in [36] increases the accuracy of the shape fitting even further by making use of a subspace offered by the PCA technique. However, these improvements are achieved at the expense of a substantial increase in the computational complexity.

This thesis is aimed at developing a low-complexity active shape model by incorporating the energy compaction property of the discrete cosine transform (DCT). The proposed ASM utilizes a novel 2-D profile of a landmark, which is based on the discrete cosine transform, in order to reduce the computational complexity without affecting the facial shape fitting accuracy. The development of the proposed model puts emphasis on reducing the execution time of a facial shape search while providing a shape

5

fitting accuracy that is better or the same as that provided by the stacked ASM or ASM with a PCA-based LGGM.

1.4 Organization of the Thesis

This thesis is organized as follows.

Chapter 2 presents an overview of the active shape model (ASM) and the discrete cosine transform (DCT) as an essential background material for the development of the work undertaken in this thesis. Some improved versions of ASM to be used as a basis for developing the proposed ASM are also presented.

In Chapter 3, a low-complexity ASM, which utilizes a novel 2-D profile constructed by making use of the energy compaction property of the DCT, is proposed. The 2-D profile consists of a subset of the DCT coefficients of the profile matrix representing the local grey-level gradient pattern around a particular landmark of a facial shape. It is shown that the size of the new profile is much smaller than that of the conventional spatial-domain 2-D profile, since the new profile is built by using only a small subset of DCT coefficients. A detailed building process of the novel profile is presented along with an illustrative example that demonstrates the building process. A theoretical computational complexity analysis of stacked ASM [35] and that of the proposed ASM is also presented in this chapter.

In Chapter 4, a facial shape search scheme using the proposed ASM for localizing a facial shape in an image is presented along with a detailed step-by-step procedure. Experiments are then performed in order to examine the effectiveness of the proposed ASM in an automatic facial landmark annotation of frontal faces. The performance

results of the scheme of the facial landmark annotation, using the proposed and two other ASMs, namely, stacked ASM [35] and ASM with a PCA-based LGGM [36], are provided and compared in terms of the shape fitting accuracy as well as in terms of the computational complexity.

Chapter 5 concludes this thesis by summarizing the main contributions made in this thesis, and suggesting some future work that can be undertaken along the ideas and schemes presented in this thesis.

CHAPTER 2

Background Material

This chapter provides an overview of active shape model (ASM) and discrete cosine transform (DCT) which are essential background materials in the discussions of the development and solution presented in this thesis. In the next section, the ASM is briefly explained to show the essence of the problem addressed in this thesis. Then, an overview of the DCT upon which the solution to the problem in this thesis is built is given. Finally, a concise summary of this chapter is presented.

2.1 Active Shape Model

Active shape model is a flexible shape model widely used in automatically locating a set of landmarks that forms the shape of any known object in an image. In general, the shape generated by the ASM can be freely deformed by adjusting the model parameters to best fit the target object in the image. The model parameters are constrained to generate shapes that are consistent with the shapes in the training set [33]. The active shape model consists of two sub-models called point distribution model and local grey-level gradient model. In this section, the shape of human faces is characterized by using the ASM.

2.1.1. Point Distribution Model

Point distribution model (PDM) is a deformable shape model, which is the foundation of ASM, and it is built by capturing the shape variations in a given set of shapes. The modelling process involves collecting and preparing the training shapes, aligning the collected shapes, and capturing the statistics of the aligned shapes.

A. Collecting and Preparing the Training Shapes

In order to correctly model the shape of human faces, it is first represented by a set of landmarks. Each landmark describes a particular part of the face such as the tip of nose and the eye pupils as shown in Figure 2.1. A set of training images is prepared to collect a range of facial shapes of various individuals. The landmarks are then manually labelled for each training image to generate a set of training shapes.
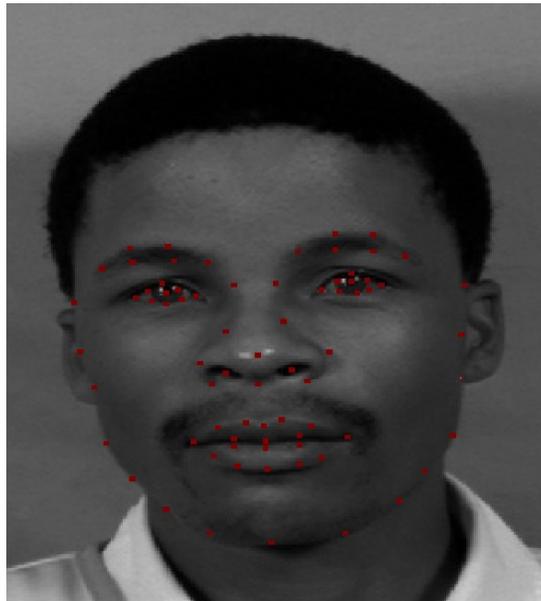
Figure 2.1: The landmarks manually labelled on a facial image.

B. Aligning the Collected Shapes

The training shapes previously prepared all differ in size, orientation and position, since these shapes come from diverse conditions and individuals. Since the PDM is built by analyzing the coordinates of each landmark of the training shapes, corresponding landmarks from different shapes must be aligned to one another with respect to a set of axes [31]. With this in mind, alignment of the shapes is done by using the generalized procrustes analysis (GPA) which scales, rotates and translates each training shape so that the corresponding landmarks in the various images of the training set are as close as possible [42]. The aligned shapes are then statistically analyzed as described next.

C. Capturing the Statistics of the Aligned Shapes

Using the aligned shapes, statistical information of the training shapes is obtained. First, each shape formed by *n* landmarks is represented by a shape vector

$$\mathbf{x} = [x_1, x_2 \ldots x_n, y_1, y_2 \ldots y_n]$$ (2.1)

where x_i and y_i are the x and y coordinates of the *i*th landmark, and n is the number of landmarks. Principal component analysis (PCA) is then performed to analyze the variation of the aligned shapes. The mean of the N training shapes is calculated as

$$\bar{\mathbf{x}} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x_i}$$ (2.2)

where $\mathbf{x_i}$ is the *i*th shape in the set of the N training images. By subtracting the mean shape calculated above from each training shape, the covariance matrix of the deviations from the mean is calculated as follows.

$$\mathbf{S_s} = \frac{1}{N} \sum_{i=1}^{N} (\mathbf{x_i} - \bar{\mathbf{x}})(\mathbf{x_i} - \bar{\mathbf{x}})^T \qquad (2.3)$$

The eigenvectors and eigenvalues of the covariance matrix are subsequently obtained by using the equation,

$$\mathbf{S_s}\mathbf{v_k} = \lambda_k \mathbf{v_k} \qquad (2.4)$$

where $\mathbf{v_k}$ and λ_k are, respectively, the kth eigenvector and eigenvalue of $\mathbf{S_s}$, and $k = 1, 2,\ldots, 2n$. The eigenvectors corresponding to the largest eigenvalues are the principal axes in the 2n-dimensional space, and describe the significant variation of the shapes [31]. The majority of variations can be described by the eigenvalues corresponding to the t largest eigenvalues (t < 2n), since the eigenvectors corresponding to the smaller eigenvalues have very small effects on the shape variation. As a result, any shape $\mathbf{x}$ in the training set can be approximated by the mean shape $\bar{\mathbf{x}}$ and a weighted sum of t eigenvectors. This is mathematically represented as

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \qquad (2.5)$$

where $\mathbf{P} = (\mathbf{v_1}, \mathbf{v_2}, \ldots \mathbf{v_t})$ is a matrix whose columns are the t eigenvectors corresponding to the first t largest eigenvalues, and $\mathbf{b} = [b_1, b_2, \ldots, b_t]^T$ is a vector of weights applied to the eigenvectors. These weights (b_k) are in fact the model parameters which control the shape generated by the PDM; thus, new facial shapes can be generated by varying these parameters. At the same time, these parameters are constrained in order for the model shape to be consistent with those in the training set. Since most of the population lies within three standard deviations of the mean, the model parameters are chosen to lie within the range given by

$$-3\sqrt{\lambda_k} \le b_k \le 3\sqrt{\lambda_k} \qquad (2.6)$$

11

where λ_k is the eigenvalue that corresponds to the kth eigenvector of the t eigenvectors.

To illustrate how the PDM works, a small set of training shapes as shown in Figure 2.2 is used to build the PDM. Figure 2.3 shows the shape generated by the PDM by varying each of the first three modes (the model parameters), b_1, b_2 and b_3, in the range specified by (2.6). As seen from this figure, variation of the model parameters results in the model shapes that are deviated versions of the mean shape.
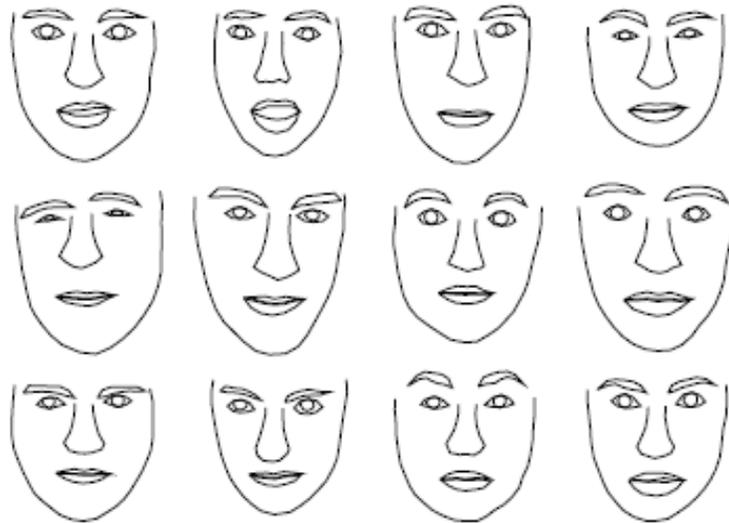
Figure 2.2: Training shapes used to build the PDM [32].

The point distribution model given by (2.5) is utilized to approximate a given target shape $\mathbf{x_s}$ through an appropriate choice of the model parameters $\mathbf{b}$ by minimizing the expression

$$\left| \mathbf{x_s} - T(\overline{\mathbf{x}} + \mathbf{Pb}) \right|^2 \tag{2.7}$$

where T is a similarity transform that transforms the model shape from the model space to the image space by scaling, rotating and translating the model shape. The model

12

parameters $\mathbf{b}_{\text{best}}$ and the transformation parameters $(t_x, t_y, s, \theta)_{best}$ that make the resulting

model shape to best fit the target shape are found. Thus, an approximation of the target

shape can be simply represented by these parameters. Local grey-level gradient model

described next is utilized to move each landmark of the model shape to a new location to

generate a new shape.
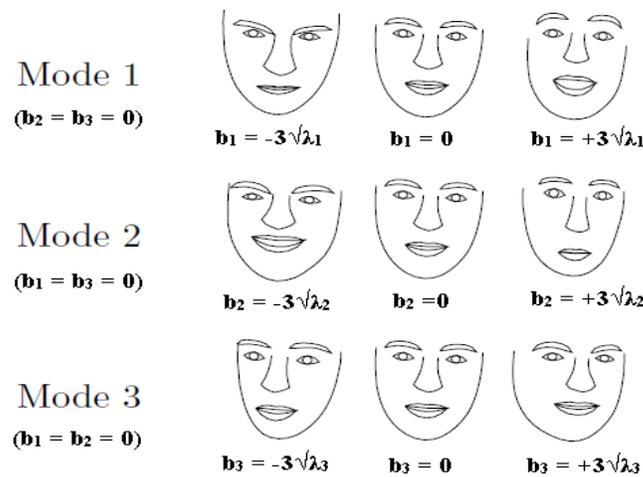
Figure 2.3: Generated shapes by varying each of the first three modes (parameters) within the three standard deviations of the mean shape [32].

2.1.2. Local Grey-Level Gradient Model

As seen previously, the point distribution model can generate a range of facial shapes by

controlling its model and transformation parameters. Local grey-level gradient model

(LGGM) describes the grey-level gradient pattern in the vicinity of each landmark in the

image containing the face in question. Given a model shape generated by the PDM, the

goal of the LGGM is to determine a new position for each landmark of the model shape

to generate a new shape which outlines the face in question better than the model shape. To build the LGGM, a grey-level intensity gradient profile $\mathbf{g_{ij}}$ for each landmark i of each shape j in the training set containing N shapes is first constructed. The classical ASM in [33] employs 1-D profile, which is generated by sampling grey-level intensities of m pixels centered at each landmark. The direction of the profile is chosen to be along a line orthogonal to the shape boundary at each landmark as shown in Figure 2.4.
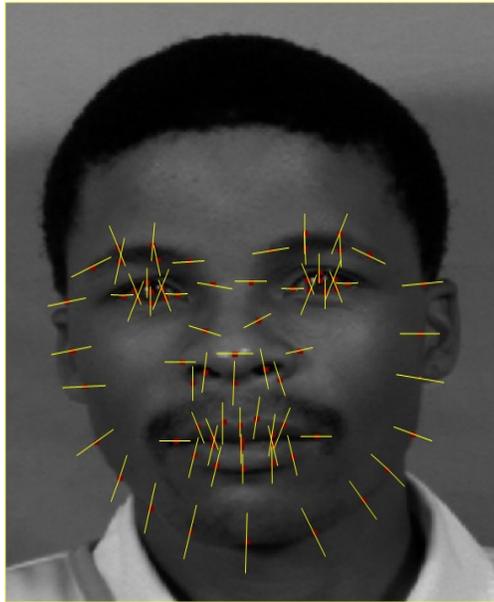
Figure 2.4: Directions of the 1-D grey-level intensity gradient profile used in [33].

In order to construct the profile $\mathbf{g_{ij}}$, each element of the sampled pixel intensities is replaced by its intensity gradient, which is the difference between intensities of the pixel and its neighbouring pixel. In the construction of the profiles, the gradient values rather than the pixel intensities are used for reducing the effect of global intensity changes from image to another image in the training set [35]. Each gradient value is

14

subsequently normalized by dividing by the sum of absolute values of the gradients in a profile as

$$\hat{\mathbf{g}}_{\mathbf{ij}} = \frac{\mathbf{g}_{\mathbf{ij}}}{\sum_{k=1}^{m} |g_{ijk}|} \qquad (2.8)$$

where g_{ijk} is the kth element in the profile $\mathbf{g}_{\mathbf{ij}}$. As a result, the normalized profile is robust to uniform scaling and the addition of a constant [31]. Finally, the LGGM for each landmark is represented by the mean profile given by

$$\bar{\mathbf{g}}_{\mathbf{i}} = \frac{1}{N} \sum_{j=1}^{N} \hat{\mathbf{g}}_{\mathbf{ij}} \qquad (2.9)$$

and the covariance matrix given by

$$\mathbf{S}_{\mathbf{i}} = \frac{1}{N} \sum_{j=1}^{N} (\hat{\mathbf{g}}_{\mathbf{ij}} - \bar{\mathbf{g}}_{\mathbf{i}})(\hat{\mathbf{g}}_{\mathbf{ij}} - \bar{\mathbf{g}}_{\mathbf{i}})^{T}, \ 1 \le i \le n. \qquad (2.10)$$

The local grey-level gradient model is thus used for determining the best position for each landmark using its statistical information given by the mean profile and covariance matrix for each landmark. A comprehensive procedure of finding a shape in an image using the PDM and the LGGM is described next.

2.1.3. Search of a Face in an Image

Searching a face in an image is equivalent to automatically finding the best location of the landmarks in the face. This is done by optimally fitting the landmarks obtained by using the PDM and the LGGM described previously onto the given face. This search involves an iterative refinement of the model shape by moving each landmark of a current model shape to a new position by using the LGGM, and then updating the current

model parameters **b** of the PDM and the transformation parameters (t_x, t_y, s, θ) in order to achieve a better fit of the model shape to the given face. The search method consists of four steps described below.

A. Generating an Initial Model Shape

The search method starts with a detection of the face in the image containing the face. Viola et al. in [43] have introduced a fast detection technique in which the face is detected by locating a rectangular region containing the face. This technique makes use of a set of simple Haar-like features for classifying the various rectangular regions containing the face. Evaluation of these features is accelerated by a fast computation and making use of a so-called integral image. A learning algorithm, known as AdaBoost, is used for selecting a smaller set of important features from the large set of the computed Haar-like features in order to build a set of weak classifiers. These weak classifiers are combined in cascade so as to quickly discard the non-face regions.

The rectangular region is represented by its center point (x_c, y_c), width w, and height h, and includes most of the facial features as shown in Figure 2.5. In order to generate an initial model shape, the model and transformation parameters are determined by using the rectangular region containing the face. The mean shape $\bar{\mathbf{x}}$ of the training set is mapped onto the given image with (x_c, y_c) as the reference point, as shown by a red line in Figure 2.5. The shape is then scaled up or down and rotated by respectively assigning a value to the scaling parameter s and the rotation parameter θ depending on the value of w and h so as to approximately fit it in an appropriate region (say 90%) of

16

the rectangular region, as shown in Figure 2.6 [35]. Therefore, the initial model shape is represented by the model parameters $\mathbf{b_0} = \mathbf{0}$ and the transformation parameters $(t_x, t_y, s, \theta)_0$ where $t_x = x_c$ and $t_y = y_c$, and subsequently used for finding a new shape with the LGGM in the search method.

Figure 2.5: An example of the rectangular region detected by the face detector, and the mean shape mapped onto the image with (x_c, y_c) as the reference point.

B. Refining the Landmark Positions using LGGM

Given the initial model shape represented by $\mathbf{b_0}$ and $(t_x, t_y, s, \theta)_0$, each landmark of the model shape is moved to a suitable position determined by using the LGGM to generate a new shape. The grey-level pixel intensities of a pre-specified number (say $L > m$) of
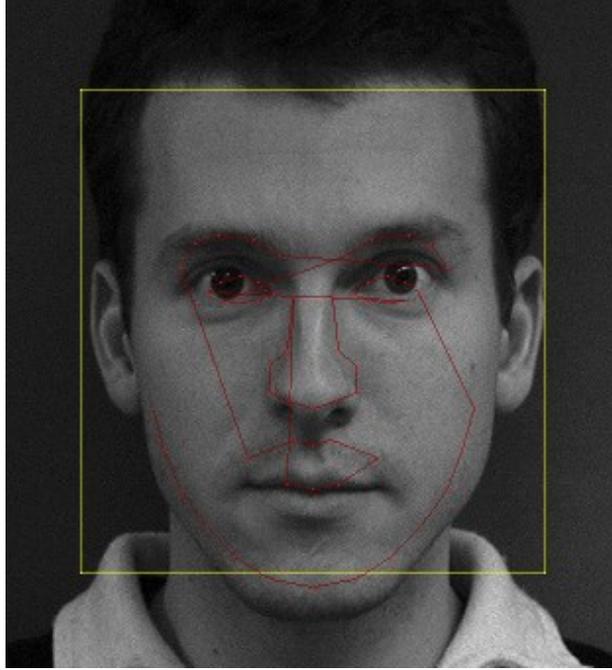
Figure 2.6: An example of the initial model shape approximately fitted in the rectangular region.

pixels centered at a particular landmark i of the current model shape are sampled from the image. The intensity gradients $\hat{\mathbf{g}}_{\mathbf{i}}$ are then calculated from the sampled pixel intensities as described in Section 2.1.2. A subset $\mathbf{g}_{\mathbf{id}}$ of the intensity gradients $\hat{\mathbf{g}}_{\mathbf{i}}$ of length m pixels centered at a candidate position d is normalized using (2.8). The cost of fit of $\mathbf{g}_{\mathbf{id}}$ to the mean profile $\bar{\mathbf{g}}_{\mathbf{i}}$ for the ith landmark of the LGGM is evaluated using the Mahalanobis distance given by

$$f(\mathbf{g}_{\mathbf{id}}) = (\mathbf{g}_{\mathbf{id}} - \bar{\mathbf{g}}_{\mathbf{i}})^{T} \mathbf{S}_{\mathbf{i}}^{-1} (\mathbf{g}_{\mathbf{id}} - \bar{\mathbf{g}}_{\mathbf{i}}) \tag{2.11}$$

where $\mathbf{S}_{\mathbf{i}}$ is the covariance matrix for the ith landmark. The cost of fit is evaluated at L - (m - 1) candidate positions on $\hat{\mathbf{g}}_{\mathbf{i}}$ using (2.11) including that at the landmark in question (i.e. the ith landmark), as shown in Figure 2.7. The position d_{best} at which the sub-profile $\mathbf{g}_{\mathbf{id}_{best}}$ yields the smallest Mahalanobis distance (the minimum cost of fit) is then selected

18

as the new position of the landmark. The above process is repeated for each landmark of the model shape to obtain a new shape $\mathbf{x}_S$ formed by the new positions of the landmarks. The new model and transformation parameters are calculated using the shape $\mathbf{x}_S$ as described next.
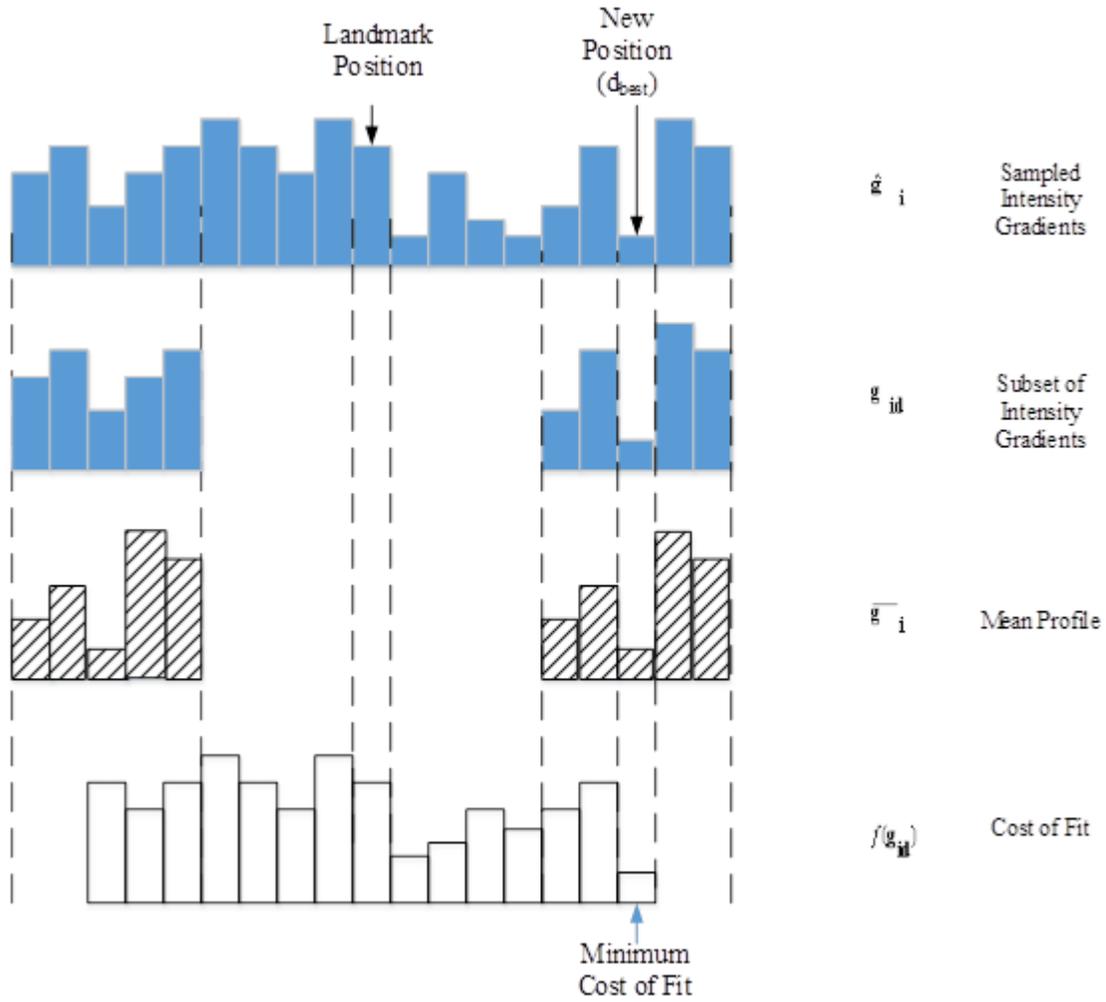
Figure 2.7: Sampled intensity gradient profile $\hat{\mathbf{g}}_i$, the subset $\mathbf{g}_{id}$ extracted from $\hat{\mathbf{g}}_i$, the mean profile $\overline{\mathbf{g}}_i$ from the LGGM aligned to the subset $\mathbf{g}_{id}$, and the cost of fit $f(\mathbf{g}_{id})$ [32].

C. *Finding a Model Shape by Calculating the Model and Transformation Parameters*

Given the new shape $\mathbf{x_S}$, the model parameters $\mathbf{b_{best}}$ and the transformation parameters $(t_x, t_y, s, \theta)_{best}$ are obtained by minimizing the expression given by (2.7) using the similarity transform T and the PDM. An iterative method to minimize the expression is comprehensively presented in [32]. Consequently, the shape $\mathbf{x_S}$ is closely approximated by the model shape given by $T_{(t_x, t_y, s, \theta)_{best}}(\overline{\mathbf{x}} + \mathbf{Pb_{best}})$, while satisfying the constraints of the model parameters defined by (2.6).

D. *Updating the Model and Transformation Parameters*

The model parameters $\mathbf{b}$ and the transformation parameters (t_x, t_y, s, θ) are then updated by assuming the best parameters $\mathbf{b_{best}}$ and $(t_{xbest}, t_{ybest}, s_{best}, \theta_{best})$:

$$t_x \leftarrow t_{xbest} \tag{2.12}$$

$$t_y \leftarrow t_{ybest} \tag{2.13}$$

$$\theta \leftarrow \theta_{best} \tag{2.14}$$

$$s \leftarrow s_{best} \tag{2.15}$$

$$\mathbf{b} \leftarrow \mathbf{b}_{best} \tag{2.16}$$

The updated parameters $\mathbf{b}$ and (t_x, t_y, s, θ) represent a model shape that approximates the shape $\mathbf{x_S}$ while satisfying the shape constraints imposed by (2.6). The model shape is then used as the starting shape in the second iteration of the search method. The steps B, C and D described above are repeated until the model shape converges, that is, the iterative process satisfies a pre-specified terminating condition. For the reason that the

shape generated by the PDM is iteratively modified to fit the target shape, and the generated shape is consistent with the shapes found in the training set, the model is called the active shape model [33].

The iterative shape search method described above for fitting the model shape to the face in the image can be summarized in the following algorithm.

Algorithm 1: Classical ASM

1. Generate the initial model shape by locating a rectangular region containing the face in the image using the Viola Jones face detector, mapping the mean shape onto the center of the region, and applying scaling and rotation to the shape to approximately fit it in the region.

2. Refine the positions of the landmarks of the model shape using the local grey-level gradient model to generate a new shape $\mathbf{x_S}$ given by the landmarks, each of which is determined by minimizing the Mahalanobis distance given by (2.11).

3. Find a new set of the model parameters $\mathbf{b}$, and the transformation parameters (t_x, t_y, s, θ) using the point distribution model and the similarity transformation to best fit the model shape to the shape $\mathbf{x_S}$, enforcing the limits given by (2.6) on the model parameters $\mathbf{b}$ and minimizing the expression given by (2.7).

4. Update the model and transformation parameters to the new parameters obtained in Step 3 (Equations (2.12) – (2.16)).

5. Repeat Steps 2 to 4 until convergence.

2.1.4. Multi-Resolution Active Shape Model

In [34], the classical ASM has been modified by searching the desired face at different levels of resolution using Algorithm 1 described in the previous section. The resulting model can be referred to as multi-resolution ASM (MRASM). It utilizes an image pyramid to build the LGGM for multiple image resolutions. Here, as shown in Figure 2.8, the authors use an image pyramid consisting of a set of images with different resolutions, which are created from an original image using Gaussian smoothing and sub-sampling [34]. Level 0 (base) of the image pyramid is the original image, level 1 is the image with half the number of pixels of the original image in each dimension, and so on. The authors in [34] have used five levels of resolutions. In order to build the multi-resolution LGGM, a mean profile and a covariance matrix of the LGGM described in Section 2.1.2 are calculated for each of the levels of the image pyramid in the training stage. At the testing stage, an image pyramid with 5 levels is first created using the test image. The shape search given by Algorithm 1 is then conducted using the image from each level of the pyramid starting at the coarsest level (level 4) and terminating at the finest level (level 0). In the shape search, the LGGM corresponding to each level of the image pyramid obtained in the training stage is utilized to find a new position of the landmarks of the model shape. As a result, there are large movements of landmarks of the model shape in the coarser levels, while smaller movements are made at the finer levels as shown in Figure 2.9. The multi-resolution active shape model successfully improves the fitting accuracy and the computational speed by using fewer iterations and shorter search lines than those used in the classical ASM [34]. Moreover, the MRASM tends not to be stuck on local image structures due to search at coarser resolution [34].

2.1.5. Active Shape Model with 2-D Profiles

In order to improve the performance of the shape search using the ASM, the authors in [35] have utilized 2-D profiles in addition to 1-D profiles in the LGGM of the classical ASM. While the 1-D profile shown in Figure 2.4 captures the grey-level intensity
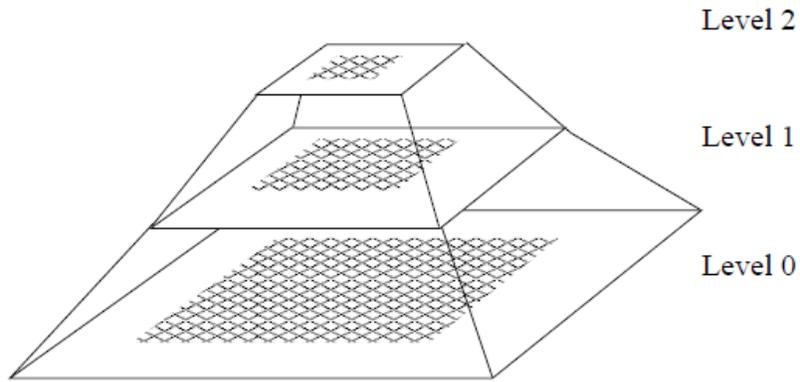
Figure 2.8: Image pyramid with 3 levels and their relationships in terms of the size [32].
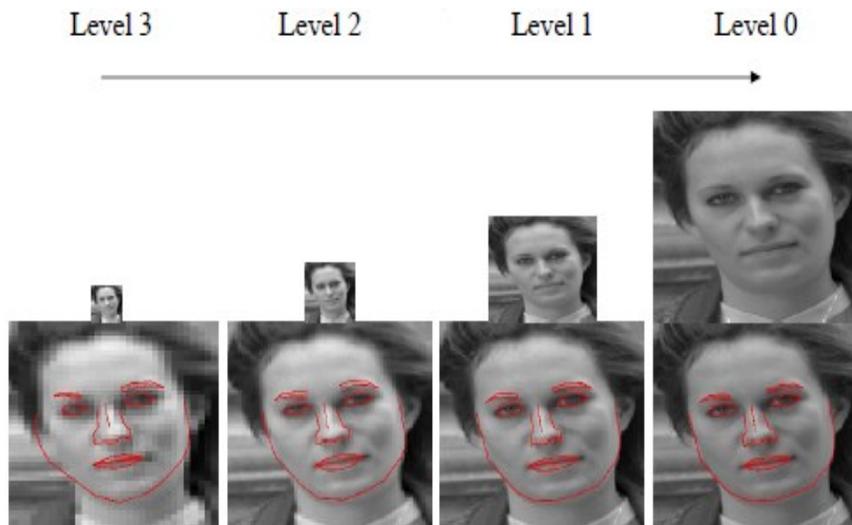
Figure 2.9: Multi-resolution iterative shape search method [36].

gradient pattern along a line orthogonal to the shape boundary at each landmark, the 2-D profile captures more information from a square region around each landmark. Figure 2.10 shows the 2-D profiles of the same landmarks as those shown in Figure 2.4.
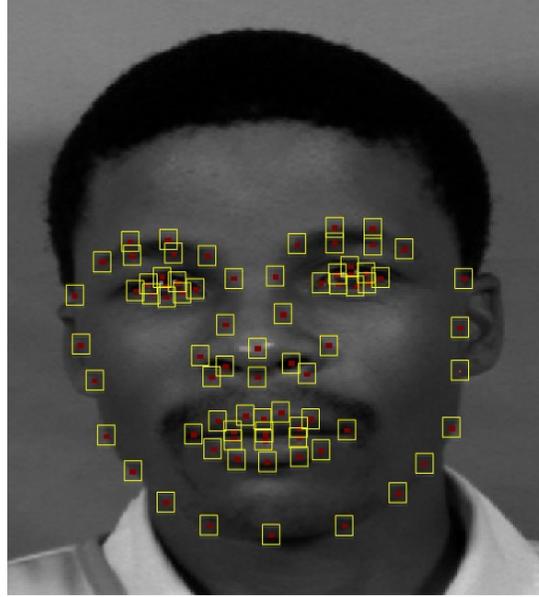
Figure 2.10: The square region around each landmark used to generate the 2-D profile.

In [35], in order to build the 2-D profile, the response of a t x q linear spatial filter $w(i, j)$ is calculated for each pixel in a 13 x 13 square region around a particular landmark i, and the resulting filter responses are stored in a 13 x 13 profile matrix **G** [40]. The filter response $G(x, y)$ at the pixel position (x, y) is given by the sum of the products of the filter coefficients and the corresponding grey-level intensities in the region spanned by the filter mask [41] as

$$G(x, y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} w(i, j)\, i(x+i, y+j) \tag{2.17}$$

where a = (t - 1) / 2 and b = (q - 1) / 2. The elements of the profile matrix **G** are then arranged into a vector by concatenating its rows. The resulting vector $\mathbf{g_i}$ corresponding

to the ith landmark is normalized by dividing each element g_{ij} by the mean of the

absolute values of all the elements as

$$g'_{ij} = \frac{g_{ij}}{\dfrac{1}{n_g}\sum_{k=1}^{n_g}\left|g_{ik}\right|} \qquad (2.18)$$

where n_g (= 169) is the number of the elements in $\mathbf{g_i}$. Each normalized g'_{ij} is then

equalized by applying a sigmoid transform as

$$g''_{ij} = \frac{g'_{ij}}{\left|g'_{ij}\right|+c} \qquad (2.19)$$

where c is the shape constant, which defines the shape of the sigmoid [35]. The resulting

vector $\mathbf{g''_i}$ with elements g''_{ij} is used for representing the 2-D intensity gradient pattern

around a landmark. The use of the so called stacked ASM that uses the 2-D profile in

addition to the 1-D profile significantly improves the fitting accuracy compared to the

classical ASM using only 1-D profiles [35].

2.1.6. Active Shape Model with a PCA-based LGGM

Seshadri et al. in [36] have developed an active shape model with a PCA-based LGGM

employing 2-D profiles to further improve its fitting accuracy. It employs the PCA to

build a subspace which models the variation of pixel appearance in a square region

around each landmark [36]. The mean profile and the covariance matrix are obtained

from the 2-D profile obtained by using a 13 x 13 square region around each landmark in

the same manner as in [35]. Then, the eigenvectors corresponding to the first 97% of the

eigenvalues of the covariance matrix of each landmark are computed. In the testing stage,

the 2-D profile obtained at each candidate position of a 5 x 5 pixel search region around a particular landmark i is projected onto the subspace using the eigenvectors to obtain a set of projection coefficients. The Mahalanobis distance between the 2-D profile at the candidate position and its reconstructed profile obtained by using the projection coefficients and the eigenvectors is computed. The optimal position at which the profile yields the minimum Mahalanobis distance is determined, and is used as the new position for the landmark. This process is repeated for each landmark to generate a new shape. Thus, ASM with a PCA-based LGGM utilizes the PCA to determine a new position for each landmark instead of using the mean profile and the covariance matrix. The use of the PCA subspace contributes to improving the fitting accuracy, and helps to mitigate the problem of local minima [39].

2.2 Discrete Cosine Transform

The discrete cosine transform (DCT) has a powerful energy compaction property [44]. This property of the DCT could be effectively utilized to reduce size of the 2-D profiles used in the ASM, and consequently, decrease the computational complexity of ASM itself. In this section, the fundamental concepts of the 1-D and 2-D DCTs and their useful properties as well as fast methods for their computation are briefly reviewed.

2.2.1.   1-D DCT

The 1-D DCT of a data sequence $x(n)$ of length N is defined by

$$C_x(k) = \alpha(k) \sum_{n=0}^{N-1} x(n) \cos\left(\frac{(2n+1)k\pi}{2N}\right) \qquad (2.20)$$

where k = 0, 1, 2,…, N - 1 and

$$\alpha(k) = \begin{cases} \sqrt{\dfrac{1}{N}} & \text{for } k = 0 \\[2ex] \sqrt{\dfrac{2}{N}} & \text{otherwise} \end{cases}. \qquad (2.21)$$

The 1-D inverse DCT of length N is given by

$$x(n) = \sum_{k=0}^{N-1} \alpha(k) C_x(k) \cos\left(\frac{(2n+1)k\pi}{2N}\right) \qquad (2.22)$$

where n = 0, 1, 2,…, N - 1.

A. Basis Functions

As seen in (2.20), the DCT coefficients are represented by a weighted sum of the functions $\cos\left(\dfrac{(2n+1)k\pi}{2N}\right)$. These functions are called the cosine basis functions, and are orthogonal to one another [44]. Similarly, the input sequence $x(n)$ as given by (2.22) can be seen to be a weighted sum of the cosine basis functions. The plot of the cosine basis functions for N = 8 is shown in Figure 2.11. The first basis function (k = 0) is represented by a zero-frequency constant function, and the DCT coefficient corresponding to k = 0 is known as the DC coefficient. Other basis functions (k > 0) are described by a set of cosine functions with different frequencies, and the DCT coefficients calculated using these functions are called the AC coefficients. The lower (smaller k) and higher (larger k) frequency basis functions represent, respectively, smooth and varying parts of the input sequence.
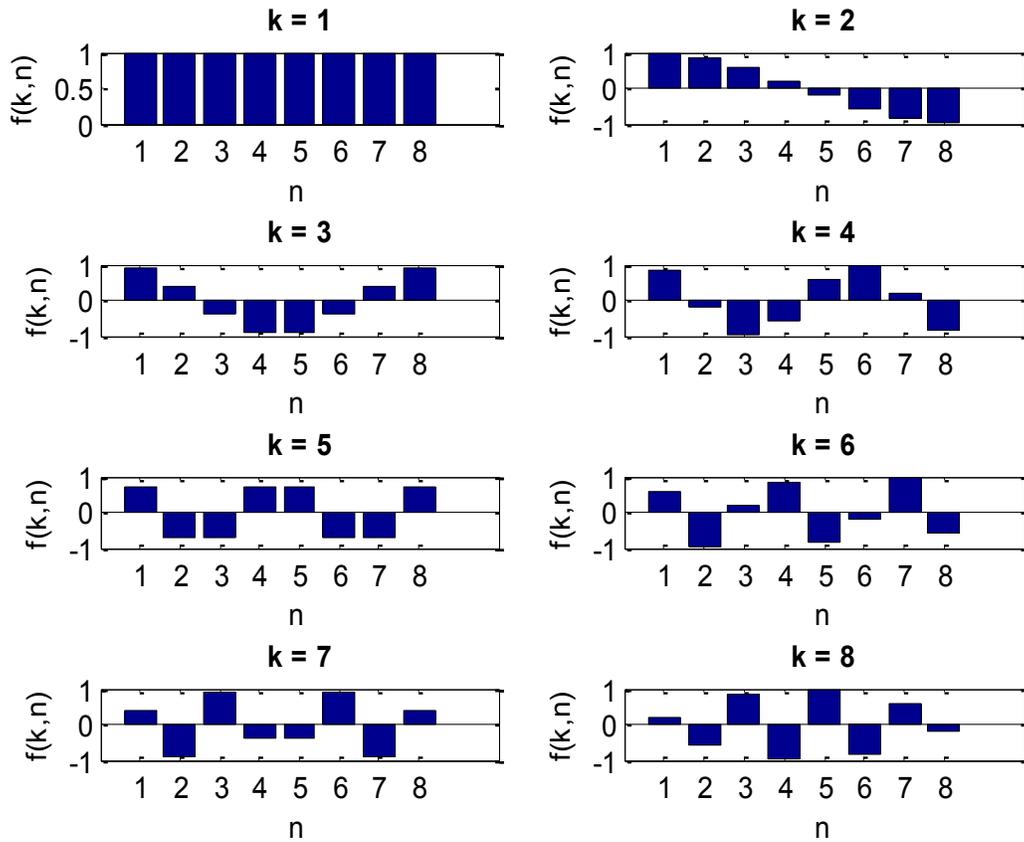
Figure 2.11: A plot of the cosine basis functions for N = 8.

B. *Energy Compaction Property*

The discrete cosine transform is well-known for its excellent energy compaction property, and it packs most of the signal energy of highly correlated input data in a small set of the DCT coefficients [45]. These coefficients represent the lower frequency components of the original signal, and contain most of the signal energy. The other coefficients corresponding to the higher frequency components can be safely discarded without losing

much of the signal energy. Therefore, the original signal can be approximated using only the small set of the DCT coefficients with a small reconstruction error.

C. Fast Computation of DCT

A direct calculation of the N-point DCT requires N multiplications and N-1 additions for each of the N coefficients, and the computational complexity of the calculation is given by $O(N^2)$. However, the 1-D DCT is derived from the 1-D Discrete Fourier Transform (DFT) and it can be calculated with the computational complexity of $O(N \log N)$ using the fast Fourier transform (FFT) [46]. Further acceleration of the calculation is achieved by minimizing the number of multiplications and additions required to compute the N-point DCT. Loeffler et al. in [47] developed a fast DCT algorithm which involves only 11 multiplications and 29 additions to compute the 8-point DCT. Moreover, Arai et al. [48] proposed a scaled DCT algorithm that requires 13 multiplications and 20 additions but defers 8 of these multiplications for the final scaling in quantization step to decrease the number of multiplications in the computation of the 8-point DCT. Therefore, the DCT can be rapidly calculated using these fast DCT implementations.

2.2.2. 2-D DCT

The 2-D DCT of a 2-D image data $x(n_1, n_2)$ is defined by

$$C_x(k_1, k_2) = \alpha(k_1)\alpha(k_2) \sum_{n_1=0}^{N-1} \sum_{n_2=0}^{N-1} x(n_1, n_2) \cos\left(\frac{(2n_1+1)k_1\pi}{2N}\right) \cos\left(\frac{(2n_2+1)k_2\pi}{2N}\right) \qquad (2.23)$$

where k_1, k_2 = 0, 1, 2,..., N - 1, and $\alpha(\cdot)$ is defined in (2.21).

The 2-D inverse DCT of $C_x(k_1, k_2)$ is then given by

$$x(n_1, n_2) = \sum_{k_1=0}^{N-1} \sum_{k_2=0}^{N-1} \alpha(k_1) \alpha(k_2) C_x(k_1, k_2) \cos\left(\frac{(2n_1+1)k_1\pi}{2N}\right) \cos\left(\frac{(2n_2+1)k_2\pi}{2N}\right) \quad (2.24)$$

where n_1, $n_2 = 0, 1, 2, \ldots, N - 1$.

A. Basis Images

The cosine functions $\cos\left(\dfrac{(2n_1+1)k_1\pi}{2N}\right) \cos\left(\dfrac{(2n_2+1)k_2\pi}{2N}\right)$ in (2.23) and (2.24) form a set

of 2-D orthogonal functions called the basis images. The 2-D DCT coefficients are thus represented by a weighted sum of the basis images. Similarly, the original image $x(n_1, n_2)$ is a weighted sum of the basis images, as seen in (2.24). Figure 2.12 is a visualization of the basis images for N = 8. As seen in the figure, when u and v are both zero (top left), the basis image is a constant 8 x 8 image (DC image) and represents the zero frequency components of the image. The basis images with higher values of k_1 and k_2 represent the high frequency components of the image.

B. Row-Column Decomposition

The direct computation of the 2-D DCT given by (2.23) has a computational complexity of $O(N^4)$. However, the 2-D DCT can be decomposed into a series of 1-D DCTs as

$$C_x(k_1, k_2) = \alpha(k_1) \sum_{n_1=0}^{N-1} \left(\alpha(k_2) \sum_{n_2=0}^{N-1} x(n_1, n_2) \cos\left(\frac{(2n_2+1)k_2\pi}{2N}\right) \right) \cos\left(\frac{(2n_1+1)k_1\pi}{2N}\right) \quad (2.25)$$

$$= \alpha(k_1) \sum_{n_1=0}^{N-1} C_x(n_1, k_2) \cos\left(\frac{(2n_1+1)k_1\pi}{2N}\right)$$
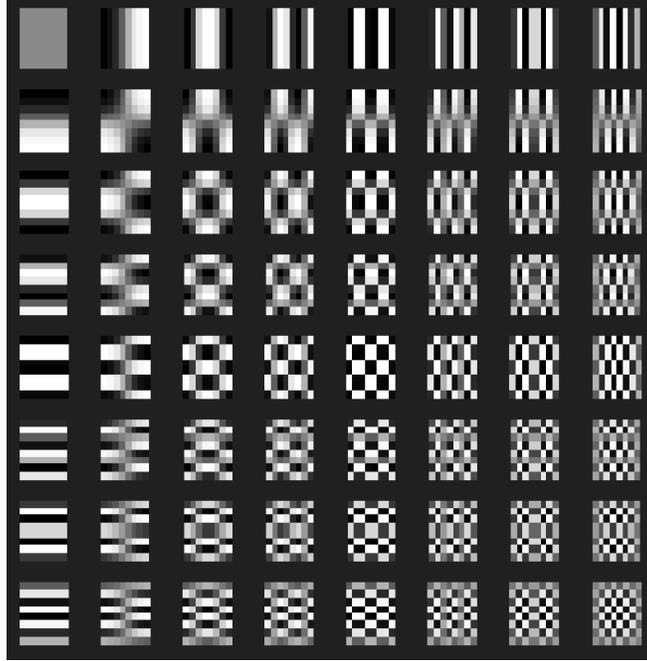
Figure 2.12: A plot of the basis images of the 2-D DCT for N = 8 [44].

where

$$C_x(n_1, k_2) = \alpha(k_2) \sum_{n_2=0}^{N-1} x(n_1, n_2) \cos\left(\frac{(2n_2 + 1)k_2\pi}{2N}\right) \tag{2.26}$$

and k_1, k_2 = 0, 1, 2,…, N - 1. In the row-column decomposition method, the 1-D DCT of

the image data $x(n_1, n_2)$ for each row is first calculated, and it is stored in $C_x(n_1, k_2)$. The

1-D DCT of $C_x(n_1, k_2)$ for each column is then computed to obtain the final 2-D DCT

transform coefficients $C_x(k_1, k_2)$ as illustrated in Figure 2.13. As a result, the row-column

decomposed 2-D DCT computation described above reduces the computational

complexity from $O(N^4)$ to $O(N^3)$.
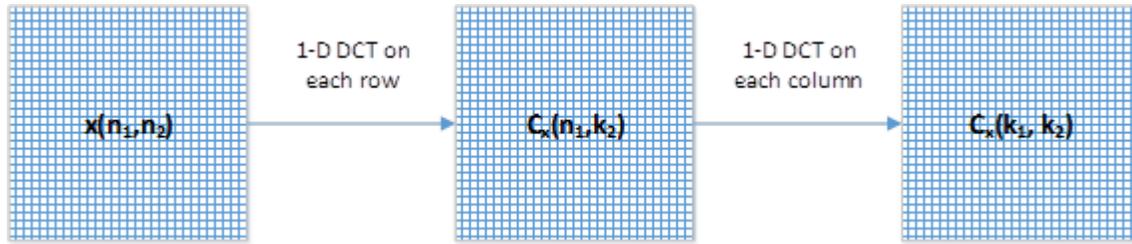
Figure 2.13: Row-Column Decomposition of the 2-D DCT.

C. Fast Computation of 2-D DCT

As described previously, the 2-D DCT is efficiently calculated in $O(N^3)$ using the row-column decomposition method. The computation complexity can be further reduced to $O(N^2 \log N)$ by using the fast 1-D DCT based on the FFT described in [46]. Use of the fast 1-D DCT computation techniques such as these described in [47] or [48] can additionally reduce the number of multiplications and additions to accelerate the computation of the 2-D DCT.

2.3 Summary

In this chapter, some background material necessary for the development of the work undertaken in this thesis has been presented. Active shape model (ASM), which is a flexible model utilized to iteratively fit the model shape to a target face in an image, has been presented in detail. It consists of point distribution model (PDM) that can generate any shape similar to the training set of shapes by controlling the model parameters, and local grey-level gradient model (LGGM) that describes the grey-level intensity gradient pattern around each landmark. In order to search a face in an image, the model shape given by the ASM is iteratively reformed in order to better fit the face. Three improved

versions of ASM, namely, the multi-resolution ASM, ASM with 2-D profiles and ASM with a PCA-based LGGM, have also been discussed. A multi-resolution ASM employs an image pyramid consisting of a set of images with different resolutions to accelerate the shape fitting. An ASM with 2-D profiles utilizes a 2-D profile obtained from a square region around each landmark in order to better represent the grey-level intensity gradient pattern around the landmark. An ASM with a PCA-based LGGM makes use of a PCA subspace to determine a new position for each landmark. Since the energy compaction property of the discrete cosine transform (DCT) can be utilized to reduce the computational complexity of ASM, finally in this chapter, this transform, some of its properties and a method for its fast computation have been reviewed.

CHAPTER 3

Proposed Active Shape Model Using Discrete Cosine Transform

3.1 Introduction

As described in Chapter 2, the active shape model (ASM) using the 2-D profiles [35] and ASM with a PCA-based local grey-level gradient model (LGGM) [36] significantly improve the shape fitting accuracy. However, these improvements are achieved at the expense of substantially increased computational complexity. It was also pointed out that the discrete cosine transform (DCT) has a powerful energy compaction property. As a result of this property, the DCT has been successfully utilized in a variety of pattern recognition applications for the purpose of dimensionality reduction [49]-[51]. In order to reduce the size of the 2-D profiles, and subsequently decrease the computational complexity of ASM itself, we propose a low-complexity ASM that utilizes a novel profile constructed by making use of the energy compaction property of the DCT [52]. In Section 3.2, the process of building the novel profile and its integration into an ASM is described in detail. In Section 3.3, an illustrative example of the building process of this profile is provided. In Section 3.4, a computational complexity analysis of stacked ASM [35] and that of the proposed ASM employing the new profile is carried out.

3.2 Proposed Active Shape Model using a DCT-Based Gradient Profile

In this section, a low-complexity active shape model (ASM) is developed by introducing a 2-D profile of a landmark of a facial shape based on the discrete cosine transform (DCT) of the local intensity gradients of the pixels around it. Because of the energy compaction property of the DCT, the size of the 2-D profile can be made much smaller than that of the conventional spatial-domain 2-D profile. The use of the compressed DCT profile for the landmark can, therefore, be expected to reduce the computational complexity of the resulting ASM. The process of building the compressed profile consists of five steps described in the following sub-sections.

3.2.1. Sampling

Computation of a 2-D profile of a landmark requires information not only on the pixel representing the landmark in question but also that on pixels in its neighbourhood. The building process of such a profile for a landmark starts with sampling grey-level intensities from an m x m square region around a particular landmark. Figure 3.1 illustrates an 8 x 8 square region around a landmark annotated on the upper lip of an image, from which the grey-level intensities are sampled for computing the profile matrix of the landmark.
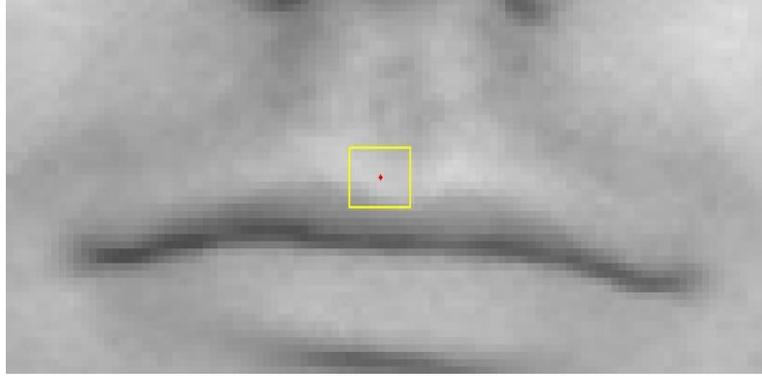
Figure 3.1: An illustration of an 8 x 8 square region (shown in yellow color) around a landmark (shown as a red dot) annotated on the upper lip of an image, used for computing a 2-D profile for the landmark.

3.2.2. Computation of the Profile Matrix

Using the grey-level intensities sampled from an m x m region around a landmark, a profile matrix of the landmark is computed by calculating the response of a t x q linear spatial filter $w(i, j)$ at each pixel of the square region [40]. Figure 3.2 illustrates a 3 x 3 spatial filter mask utilized for computing the profile matrix of the landmark. The filter response $G(x, y)$ at the pixel position (x, y) is given by the sum of products of the filter coefficients and the corresponding grey-level intensities in the region spanned by the filter mask [41], that is,

$$G(x, y) = \sum_{i=-a}^{a} \sum_{j=-b}^{b} w(i, j) i(x + i, y + j) \tag{3.1}$$

where a = (t – 1) / 2 and b = (q – 1) / 2. For example, using a 3 x 3 grey-level intensity region centered at the pixel position (x, y) shown in Figure 3.3, the response of the 3 x 3 spatial filter at the pixel position (x, y) of the region is computed as

w(-1, -1)	w(-1, 0)	w(-1, 1)
w(0, -1)	w(0, 0)	w(0, 1)
w(1, -1)	w(1, 0)	w(1, 1)

j

i
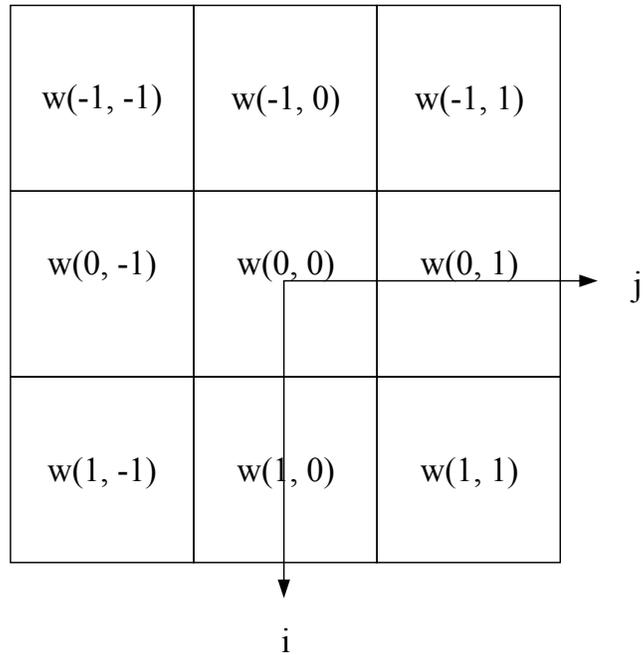
Figure 3.2: An illustration of a 3 x 3 spatial filter mask utilized for computing the profile matrix of a landmark.

i(x-1, y-1)	i(x-1, y)	i(x-1, y+1)
i(x, y-1)	i(x, y)	i(x, y+1)
i(x+1, y-1)	i(x+1, y)	i(x+1, y+1)

y

x
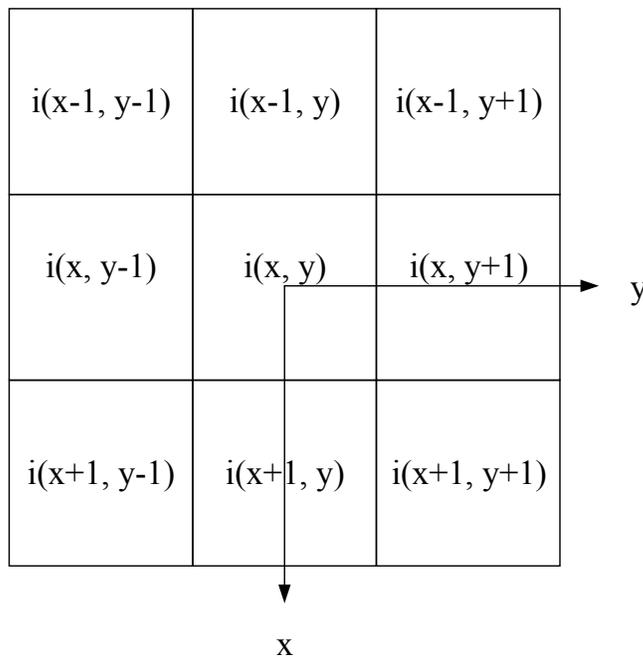
Figure 3.3: An illustration of a 3 x 3 grey-level intensity region centered at the pixel position (x, y), utilized for computing the response of a 3 x 3 spatial filter.

$$G(x,y) = w(-1,-1)\, i(x-1, y-1) + w(-1,0)\, i(x-1, y) + w(-1,1)\, i(x-1, y+1)$$
$$+\, w(0,-1)\, i(x, y-1) + w(0,0)\, i(x, y) + w(0,1)\, i(x, y+1) \qquad (3.2)$$
$$+\, w(1,-1)\, i(x+1, y-1) + w(1,0)\, i(x+1, y) + w(1,1)\, i(x+1, y+1)$$

Figure 3.4 shows examples of some of the 3 x 3 filter masks that can be utilized for computing the profile matrix of a landmark [40]. It is obvious that the use of the mask shown in Figure 3.4 (a) does not make use of any neighbourhood information around the landmark. The mask shown in Figure 3.4 (b) and (c) utilize the information on the forward intensity change only in one direction (i.e. horizontal or vertical). The mask shown in Figure 3.4 (d) would capture the forward intensity variations along both the x- and y-directions. The mask shown in Figure 3.4 (e) would capture the forward and backward intensity variations along the x- and y-directions, whereas the mask of Figure 3.4 (f) would, in addition, capture the diagonal intensity variations. Suitability of a particular mask is application dependent. Choice of a filter mask would be examined in Chapter 4 in the context of an application of the proposed scheme in facial landmark annotation.

In order to compute the filter response at a pixel position, we need the grey-level intensity at this position and that at the positions around it. Therefore, the computation of the filter response at the pixel positions in the first row (column) and the mth row (column) requires the grey-level intensities at the pixel positions outside of the m x m region around the landmark. Figure 3.5 illustrates an example of an 8 x 8 region around a landmark. It is clear from this figure that in order to compute the filter response at pixel positions with x = 0 or x = m - 1 or at pixel positions with y = 0 or y = m - 1, we need the grey-level intensities at the pixel positions with x = -1, x = m, y = -1 or y = m. Consequently, additional 4m + 4 pixel positions outside of the m x m square region are

required in order to compute the m x m array of the spatial filter response. The profile matrix **G** of a landmark is simply an m x m array of the filter response at the pixel positions in the m x m region corresponding to the landmark.

0	0	0
0	1	0
0	0	0

0	0	0
0	-1	1
0	0	0

0	0	0
0	-1	0
0	1	0

(a)   (b)   (c)

0	0	0
0	-2	1
0	1	0

0	1	0
1	-4	1
0	1	0

1	1	1
1	-8	1
1	1	1

(d)   (e)   (f)
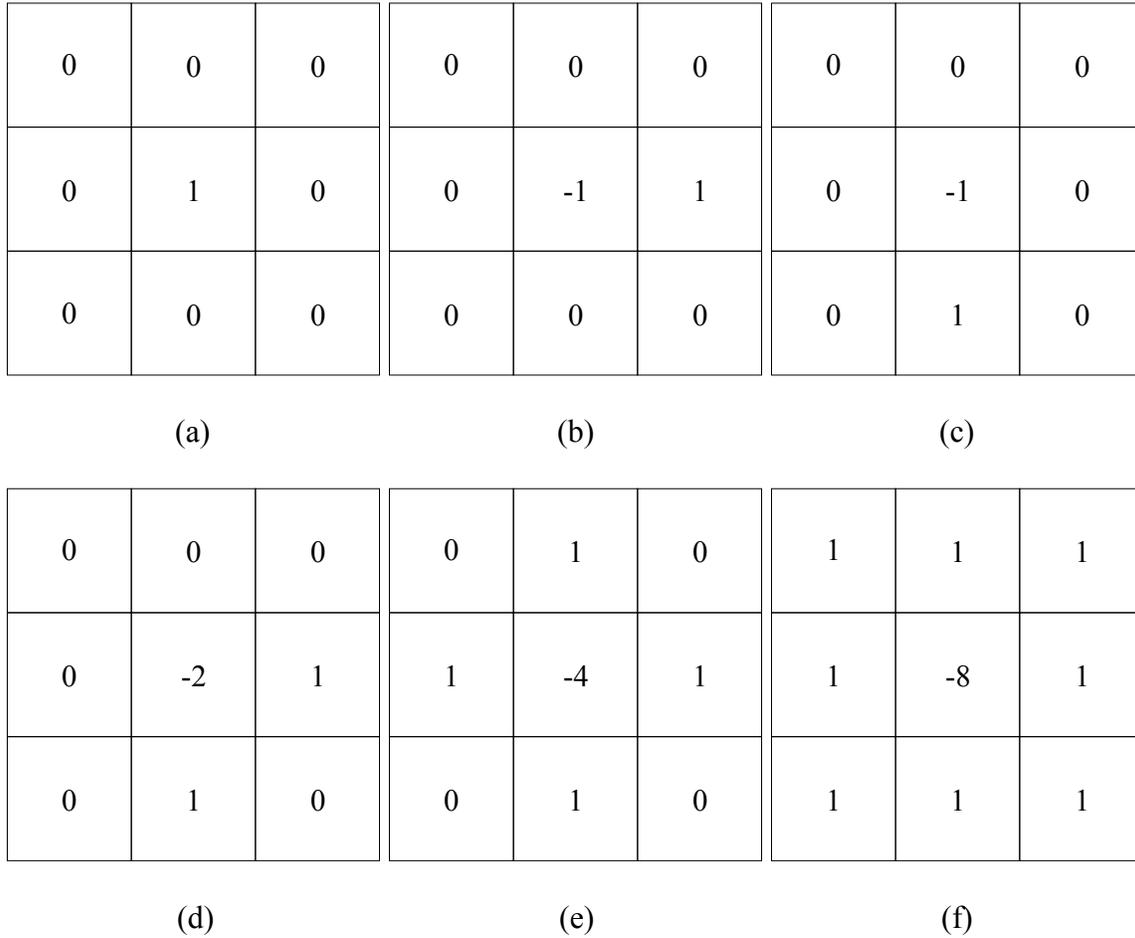
Figure 3.4: Examples of some of the 3 x 3 filter masks that can be used for computing the profile matrix of a landmark. (a) A mask that does not make use of neighbourhood information around the landmark. (b) A mask that utilizes the information on the forward intensity change in the x-direction. (c) A mask that utilizes the information on the forward intensity change in the y-direction. (d) A mask that captures the forward intensity variations along both the x- and y-directions. (e) A mask that captures the forward and backward intensity variations along the x- and y-directions. (f) A mask that captures the forward and backward intensity variations along the diagonals as well as the x- and y-directions [40].

i(-1, -1)	i(-1, 0)	i(-1, 1)	i(-1, 2)	i(-1, 3)	i(-1, 4)	i(-1, 5)	i(-1, 6)	i(-1, 7)	i(-1, 8)
i(0, -1)	i(0, 0)	i(0, 1)	i(0, 2)	i(0, 3)	i(0, 4)	i(0, 5)	i(0, 6)	i(0, 7)	i(0, 8)
i(1, -1)	i(1, 0)	i(1, 1)	i(1, 2)	i(1, 3)	i(1, 4)	i(1, 5)	i(1, 6)	i(1, 7)	i(1, 8)
i(2, -1)	i(2, 0)	i(2, 1)	i(2, 2)	i(2, 3)	i(2, 4)	i(2, 5)	i[(2, 6)	i(2, 7)	i(2, 8)
i(3, -1)	i(3, 0)	i(3, 1)	i(3, 2)	i(3, 3)	i(3, 4)	i(3, 5)	i(3, 6)	i(3, 7)	i(3, 8)
i(4, -1)	i(4, 0)	i(4, 1)	i(4, 2)	i(4, 3)	i(4, 4)	i(4, 5)	i(4, 6)	i(4, 7)	i(4, 8)
i(5, -1)	i(5, 0)	i(5, 1)	i(5, 2)	i(5, 3)	i(5, 4)	i(5, 5)	i(5, 6)	i(5, 7)	i(5, 8)
i(6, -1)	i(6, 0)	i(6, 1)	i(6, 2)	i(6, 3)	i(6, 4)	i(6, 5)	i(6, 6)	i(6, 7)	i(6, 8)
i(7, -1)	i(7, 0)	i(7, 1)	i(7, 2)	i(7, 3)	i(7, 4)	i(7, 5)	i(7, 6)	i(7, 7)	i(7, 8)
i(8, -1)	i(8, 0)	i(8, 1)	i(8, 2)	i(8, 3)	i(8, 4)	i(8, 5)	i(8, 6)	i(8, 7)	i(8, 8)

y

x

Figure 3.5: An illustration of additional pixel intensity values required for the computation of an 8 x 8 profile matrix **G** corresponding to a given landmark.

3.2.3. Computation of the DCT Coefficients

The profile matrix of the landmark in the spatial domain is transformed to the frequency domain by using the 2-D discrete cosine transform (DCT). Specifically, the 2-D DCT is applied to the m x m profile matrix **G** of the landmark to obtain an m x m array of the 2-

D DCT coefficients. Using the definition of the 2-D DCT described in Section 2.2.2, the 2-D DCT of the m x m profile matrix **G** is obtained as

$$C_G(k_1,k_2) = \alpha(k_1)\,\alpha(k_2)\sum_{x=0}^{m-1}\sum_{y=0}^{m-1}G(x,y)\cos\left(\frac{(2x+1)k_1\pi}{2m}\right)\cos\left(\frac{(2y+1)k_2\pi}{2m}\right) \qquad (3.3)$$

where

$$\alpha(k) = \begin{cases} \sqrt{\dfrac{1}{m}} & \text{for k} = 0 \\ \sqrt{\dfrac{2}{m}} & \text{otherwise} \end{cases}$$

and k_1, k_2 = 0, 1, 2,..., m − 1. The resulting m x m array $\mathbf{C_G}$ of the DCT coefficients $C_G(k_1, k_2)$ consists of a DC coefficient representing the zero frequency component and the AC coefficients representing the low- and high-frequency components of the profile matrix. Representation of the elements of the profile matrix of the landmark in the frequency-domain would allow the DCT coefficients to be arranged in a sequence with a decreasing order of their importance, and thus, facilitate the selection of a small subset of the DCT coefficients representing the profile matrix.

3.2.4. Selection of the DCT Coefficients

The 2-D discrete cosine transform efficiently packs most of the energy of the profile matrix into a small subset of the DCT coefficients representing the low-frequency components [44]. The DCT coefficients representing the high-frequency components can, therefore, be safely discarded without losing much information on the profile matrix. Making use of this energy compaction property of the DCT, a set of n_c significant coefficients is selected from the m x m array of the DCT coefficients. The m x m array of

41

the 2-D DCT coefficients are zig-zag scanned in order to select the low-frequency coefficients before selecting the high-frequency coefficients [53]. Figure 3.6 illustrates the zig-zag scanning of an 8 x 8 array of the DCT coefficients. The scanning process starts with the DC coefficient $C_G(0, 0)$ representing the average of the elements of the profile matrix, and continues with the AC coefficients representing increasingly higher-frequency components of the profile matrix. The resulting zig-zag sequence of the m x m DCT coefficients can be represented by a vector given by

$$\mathbf{Y}_{\mathbf{C_G}} = [C_G(0,0), C_G(0,1), C_G(1,0), C_G(2,0), C_G(1,1), C_G(0,2), \ldots, \\ C_G(m-2,m-1), C_G(m-1,m-2), C_G(m-1,m-1)]^T \tag{3.4}$$

The first n_c DCT coefficients representing the low-frequency components of the profile matrix are selected from the vector $\mathbf{Y}_{\mathbf{C_G}}$, giving

$$\mathbf{Y}_{n_c} = [y_0, y_1, y_2, \ldots, y_{n_c-1}]^T \tag{3.5}$$

3.2.5. Normalization and Equalization of the Selected DCT Coefficients

The n_c DCT coefficients are normalized by dividing each coefficient by the mean of the absolute values of the DCT coefficients as

$$y'_i = \frac{y_i}{\dfrac{1}{n_c}\displaystyle\sum_{j=0}^{n_c-1}|y_j|} \tag{3.6}$$

where y_i is the ith element in $\mathbf{Y}_{n_c}$ and stored in

$$\mathbf{Y}'_{n_c} = [y'_0, y'_1, y'_2, \ldots, y'_{n_c-2}, y'_{n_c-1}]^T \tag{3.7}$$

42
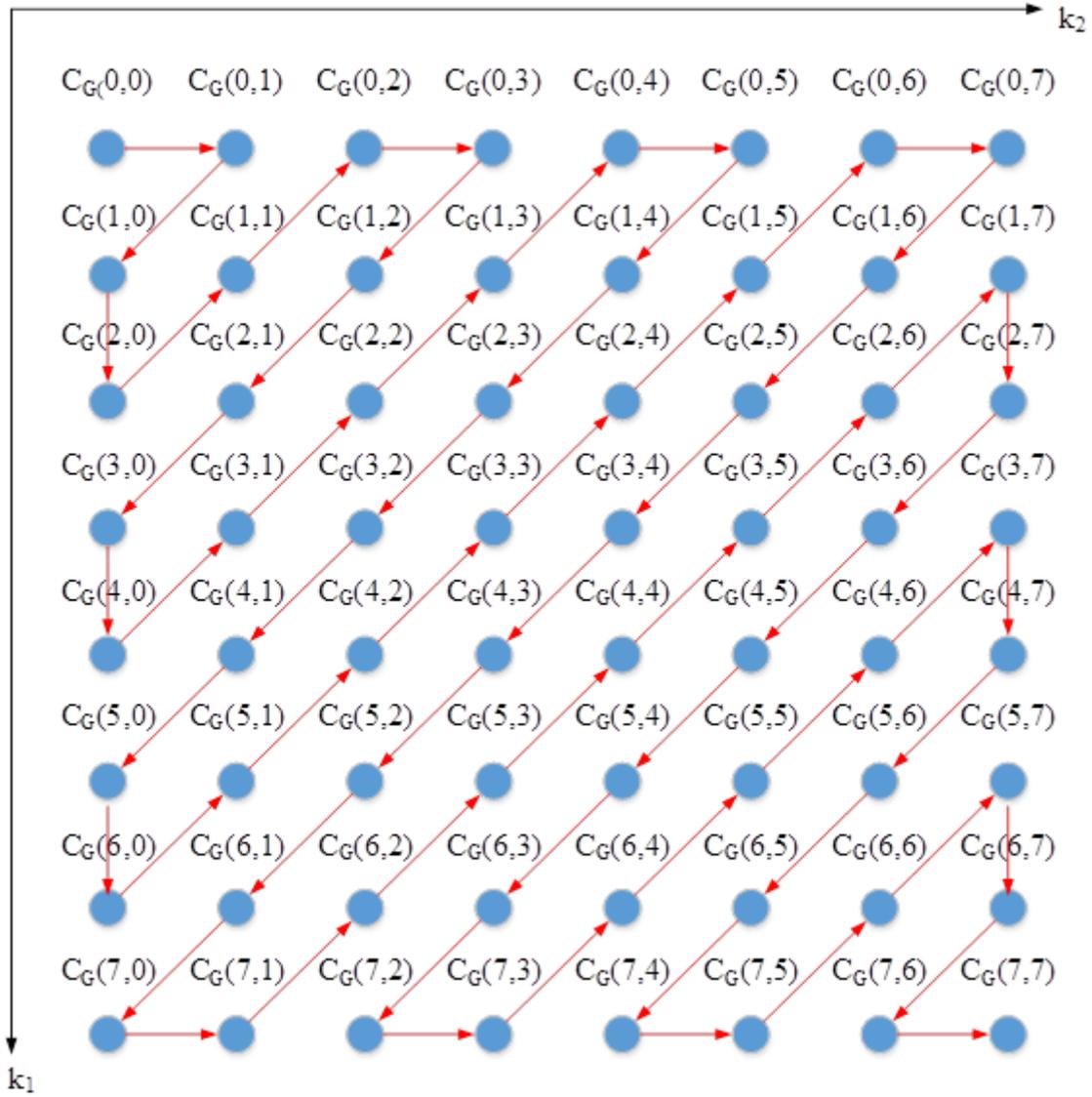
Figure 3.6: An illustration of a zig-zag scanning of an 8 x 8 array $\mathbf{C_G}$ of the DCT coefficients.

As a result of the operation, the normalized DCT coefficients are more robust to variations in brightness and contrast over the images in the training set [40]. In order to reduce the effect of outliers in the normalized 2-D DCT coefficients, each coefficient is then equalized by applying a sigmoid transform as

$$y_i'' = \frac{y_i'}{|y_i'| + c} \tag{3.8}$$

where c is a shape constant [40]. The resulting n_c DCT coefficients are finally used as a compressed DCT profile of the landmark, denoted by

$$\mathbf{Y}''_{n_c} = [y''_0, y''_1, y''_2, \ldots, y''_{n_c-2}, y''_{n_c-1}]^T \qquad (3.9)$$

3.3 An Illustrative Example

In this section, an example is considered in order to illustrate the building process of the compressed DCT profile of a landmark. The example utilizes an image shown in Figure 3.7 for building a compressed DCT profile of a landmark (shown as a red dot) annotated on the face contour of the image. Each step of the process of building the compressed DCT profile described in Section 3.2 is applied to this landmark, and the results are numerically presented.

3.3.1. Sampling

The first step of building the compressed DCT profile is sampling the grey-level intensities from the 10 x 10 region around the landmark in question. The 8 x 8 region is shown by a yellow square in Figure 3.7, and the values of the grey-level intensities within the 10 x 10 region are given in Figure 3.8.

3.3.2. Computation of the Profile Matrix

A profile matrix of the landmark is obtained by computing the response of a 3 x 3 linear

44
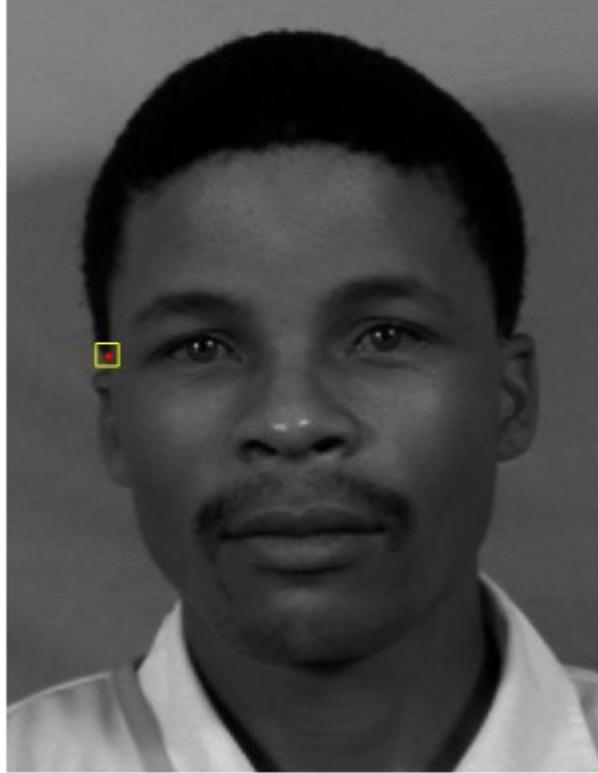
Figure 3.7: An image utilized for building a 2-D profile of a landmark (shown as a red dot) annotated on the face contour and the 8 x 8 square (shown in yellow color) around it [54].

spatial filter $w_{3x3}(i, j)$ at each pixel position (x, y) in the 8 x 8 region. In this example, the filter mask that captures the forward intensity variations both along the x- and y-directions, given in Figure 3.4 (d) and reproduced for convenience in Figure 3.9 (a), is utilized for computing the filter response. For example, the filter response at the pixel position (0, 7) located at the top right corner of the 8 x 8 region in Figure 3.8 is obtained by computing the sum of products of the filter coefficients of the spatial filter $w_{3x3}(i, j)$ and the grey-level intensities from a 3 x 3 region centered at the pixel position (0, 7).

67	47	28	19	24	38	54	55	56	57
73	59	41	27	19	41	54	55	56	56
73	63	44	36	35	43	55	56	56	55
72	63	50	45	40	44	54	56	57	55
73	66	64	58	49	47	55	56	58	57
77	75	72	68	55	52	54	55	55	58
78	78	78	73	61	56	57	58	57	58
77	74	73	69	61	57	58	59	58	56
70	63	57	53	50	52	57	59	57	53
53	47	42	39	41	47	55	45	58	54

y

x

Figure 3.8: An illustration of the grey-level pixel intensity values required for building an 8 x 8 profile matrix for the landmark shown in Figure 3.7.

Figure 3.9 (b) shows the grey-level pixel intensity values of the 3 x 3 region centered at the pixel position (0, 7) in Figure 3.8, utilised for computing the profile matrix of the landmark.

0	0	0		55	56	57
0	-2	1		55	56	56
0	1	0		56	56	55

| (a) | (b) |
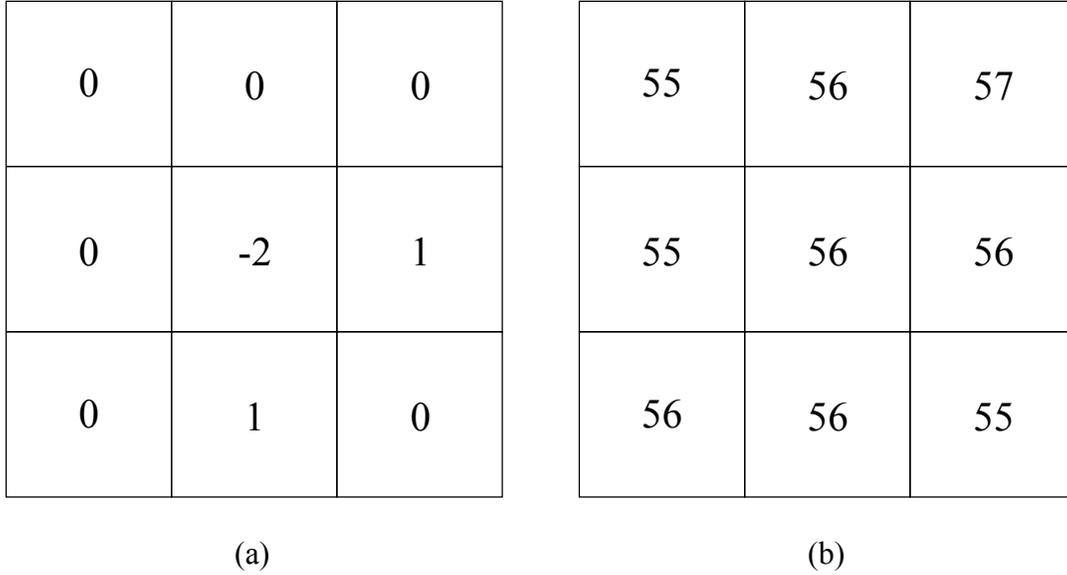
Figure 3.9: An illustration of (a) a 3 x 3 filter mask that captures the forward intensity variations both along the x- and y-directions, and (b) a 3 x 3 grey-level intensity region centered at the pixel position (0, 7), utilized for computing the filter response in this example.

The filter response at the pixel position (0, 7) is computed by using (3.1) as

$$
\begin{aligned}
G(0,7) &= \sum_{i=-1}^{1}\sum_{j=-1}^{1} w(i,j)\, i(0+i, 7+j) \\
&= (0)(55) + (0)(56) + (0)(57) + (0)(55) + (-2)(56) + (1)(56) + (0)(56) \\
&\quad + (1)(56) + (0)(55) \\
&= -112 + 56 + 56 \\
&= 0
\end{aligned}
\tag{3.9}
$$

The resulting 8 x 8 array of the filter response (profile matrix) is shown in Figure 3.10.

3.3.3. Computation of the DCT Coefficients

By using (3.3), the 2-D discrete cosine transform is applied to the 8 x 8 profile matrix of the landmark in order to obtain an 8 x 8 array of the 2-D DCT coefficients representing

y

-14	-11	11	18	15	2	2	0
-19	-2	8	13	13	0	0	0
-10	9	8	13	13	3	1	-1
7	2	1	4	13	0	1	-4
0	2	-8	3	6	4	3	5
-4	-10	-16	-5	2	2	0	2
-12	-20	-24	-15	-4	0	-2	-3
-22	-19	-17	-7	0	-1	0	-3

x

Figure 3.10: The 8 x 8 array of the filter response (profile matrix) of the landmark in the image of Figure 3.7.

the profile matrix. Figure 3.11 shows the resulting DCT coefficients representing the profile matrix of the landmark. The DC coefficient for example considered can be calculated as

$$C_G(0,0) = \alpha(0)\,\alpha(0) \sum_{x=0}^{7} \sum_{y=0}^{7} G(x,y)$$

$$= \left(\sqrt{\frac{1}{8}}\right)\left(\sqrt{\frac{1}{8}}\right)(-14-11+11+\ldots-1+0+3) \quad\quad (3.10)$$

$$= \frac{-67}{8} = -8.375$$

The entire 2-D array of the computed DCT coefficients corresponding to the profile matrix of Figure 3.10 is shown in Figure 3.11.

k_2

-8.3750	-25.8623	-27.2813	3.0615	9.3750	-4.1307	-8.0475	1.8760
36.3147	15.9540	-24.6278	-21.1504	-2.2349	0.2212	-0.8067	4.2932
-18.8883	-16.9922	-13.8791	-8.1846	-1.4699	4.5850	3.8902	2.1556
-1.8384	-5.4606	5.9248	4.2275	5.8891	3.9707	4.1662	-1.4588
6.6250	2.4535	-1.4535	0.5930	2.8750	3.9785	0.7373	2.8040
-3.5243	3.9735	3.1057	7.3797	2.4681	3.3280	3.0976	3.9463
6.3355	1.1238	-3.3598	-2.8130	-0.6088	0.1347	-0.6209	-1.5135
2.0327	-2.2978	0.4535	-2.1281	-2.7695	-1.6652	-1.2404	-4.5095

k_1

Figure 3.11: The array of the DCT coefficients of the profile matrix of Figure 3.10.

3.3.4. Selection of the DCT Coefficients

A set of 25 DCT coefficients representing the low-frequency components of the profile matrix is selected from the 8 x 8 array of the 2-D DCT coefficients given in Figure 3.11. For this selection, a zig-zag scanning of the DCT coefficients is performed as shown in Figure 3.12, since this scanning as explained in Section 3.2.4 takes into account the energy compaction property of the DCT coefficients.
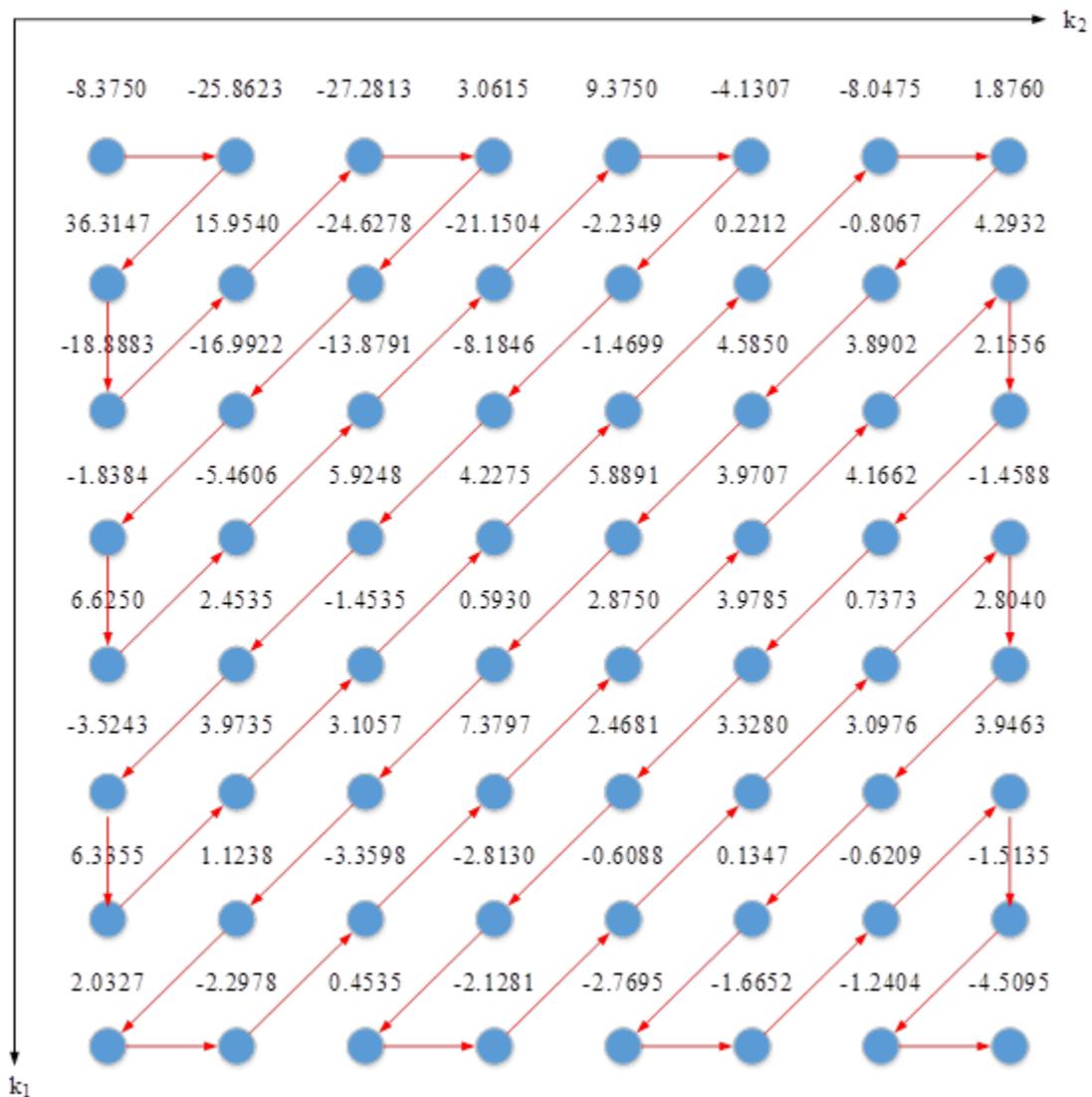
Figure 3.12: A zig-zag scanning of the 8 x 8 array of the DCT coefficients of the profile matrix corresponding to the landmark of Figure 3.7.

It is seen from this figure that through the zig-zag scanning, the low-frequency components of the profile matrix get, in general, selected before the others. The first 25 DCT coefficients from the resulting zig-zag scanning of the 2-D array of the DCT coefficients are selected and stored in a vector (1-D array) $\mathbf{Y}_{25}$ given by

$$\mathbf{Y}_{25} = [\,-8.3750, -25.8623, 36.3147, -18.8883, 15.9540, -27.2813, 3.0615, -24.6278,$$
$$-16.9922, -1.8384, 6.6250, -5.4606, -13.8791, -21.1504, 9.3750, -4.1307, -2.2349,$$
$$-8.1846, 5.9248, 2.4535, -3.5243, 6.3355, 3.9735, -1.4535, 4.2275\,]^{T}$$

$$(3.11)$$

3.3.5. Normalization and Equalization of the Selected DCT Coefficients

The selected 25 DCT coefficients are normalized by using (3.6) and subsequently equalized by using (3.8). For example, the first coefficient in $\mathbf{Y}_{25}$ is divided by the mean of the absolute values of the DCT coefficients as given in (3.6) to obtain the normalized DCT coefficient as

$$
\begin{aligned}
y_0' &= \frac{y_0}{\dfrac{1}{25}\displaystyle\sum_{j=0}^{24}\left|y_j\right|} \\
&= \frac{-8.3750}{\left(\dfrac{1}{25}\right)(278.1283)} \\
&= -0.7528
\end{aligned}
\tag{3.12}
$$

The entire array of the normalized DCT coefficients is stored in $\mathbf{Y}_{25}'$ as

$$\mathbf{Y}_{25}' = [\,-0.7528, -2.3247, 3.2642, -1.6978, 1.4341, -2.4522, 0.2752, -2.2137,$$
$$-1.5274, -0.1652, 0.5955, -0.4908, -1.2475, -1.9011, 0.8427, -0.3713, -0.2009,$$
$$-0.7357, 0.5326, 0.2205, -0.3168, 0.5695, 0.3572, -0.1306, 0.3800\,]^{T}$$

$$(3.13)$$

51

The sigmoid transform of (3.8) with a shape constant of 100 (c = 100) is then applied to the normalized DCT coefficient. For instance, the fifth normalized coefficient of $\mathbf{Y}'_{25}$ is equalized as

$$
\begin{aligned}
y''_4 &= \frac{y'_4}{|y'_4| + 100} \\
&= \frac{1.4341}{|1.4341| + 100} \\
&= 0.0141
\end{aligned}
\tag{3.14}
$$

The entire array of the DCT coefficients after normalization and equalization of each element is given by

$$
\begin{aligned}
\mathbf{Y}''_{25} = [\,&-0.0075, -0.0227, 0.0316, -0.0167, 0.0141, -0.0239, 0.0027, -0.0217, \\
&-0.0150, -0.0016, 0.0059, -0.0049, -0.0123, -0.0187, 0.0084, -0.0037, \\
&-0.0020, -0.0073, 0.0053, 0.0022, -0.0032, 0.0057, 0.0036, -0.0013, 0.0038\,]^T
\end{aligned}
$$

$$
\tag{3.15}
$$

The array $\mathbf{Y}''_{25}$ of the resulting DCT coefficients is used as a compressed DCT profile of the landmark on the face contour of the image shown in Figure 3.7.

3.4 Computational Complexity Analysis

In this section, computational complexity analyses of stacked active shape model (ASM) [35] and the proposed ASM using a DCT-based gradient profile are carried out. These versions of active shape model are utilized in Algorithm 1 for searching a facial shape in an image presented in Chapter 2. The stacked ASM utilizes both the 1-D and 2-D profiles while ASM using a DCT-based gradient profile employs the compressed 2-D DCT profile for searching a suitable location of a landmark. Thus, difference between the two

models arises from the process of building the profile(s) of a landmark used in searching a facial shape in an image.

As described in Section 2.1, in the stacked ASM, both 1-D and 2-D profiles need to be built in the spatial domain. The 1-D profile of a landmark is built by sampling the grey-level intensities of m pixels centered at the landmark, computing the intensity gradient at each pixel position, and normalizing each intensity gradient. Computation of the intensity gradients of the sampled grey-level intensities requires $24m + 35$ addition or multiplication operations. Normalization of the resulting intensity gradients involves $2m + 2$ addition or multiplication operations. The total number of arithmetic operations in building the 1-D profile is $26m + 37$. Therefore, the complexity of building the 1-D profile for a landmark is $O(m)$. On the other hand, building a 2-D profile of a landmark requires sampling the grey-level intensities from an m x m region around the landmark, computing an m x m array of the response of a linear spatial filter at each pixel position in the region (profile matrix), and normalizing and equalizing each element of the profile matrix. In order to compute the profile matrix of the landmark, a 3 x 3 filter mask is correlated with the grey-level intensities in a 3 x 3 region centered at each pixel position in the m x m region. Each computation of the filter response requires 8 additions and 9 multiplications, and the computation of the profile matrix thus requires $17m^2$ addition or multiplication operations. The normalization and equalization of the profile matrix require in total $4m^2 + 3$ addition or multiplication operations. Thus, the total number of arithmetic operations in building a 2-D landmark profile for the stacked ASM is $21m^2 + 3$. Thus, the complexity of building this 2-D profile is $O(m^2)$.

The process of building the compressed DCT profile of a landmark consists of sampling the grey-level intensities from an m x m region around the landmark, computing an m x m profile matrix, computing its DCT coefficients, selecting the n_c significant DCT coefficients, and normalizing and equalizing each element of the selected DCT coefficients. As seen in the process of building the 2-D profile for the stacked ASM, computation of the profile matrix requires $17m^2$ addition or multiplication operations. The DCT coefficients representing the profile matrix can be computed by using a fast 1-D FFT algorithm with row-column decomposition. For example, if we use the 1-D FFT algorithm described in [55], then computing the m x m array of the 2-D DCT coefficients requires $4m^2 \log_2(m) - 2m^2 + 2m$ addition or multiplication operations. The normalization and equalization of a set of n_c ($<= m^2$) significant coefficients selected from the m x m array of the DCT coefficients require $4n_c + 3$ additions or multiplications. Thus, the total number of arithmetic operations in building the compressed 2-D DCT profile of a landmark is $4m^2 \log_2(m) + 15m^2 + 2m + 4n_c + 3$. Therefore, the complexity of building the 2-D profile in the proposed active shape model is $O(m^2 \log_2 m)$.

Table 3.1 summarizes the number of arithmetic operations of each step as well as the total number of operations involved in building the profiles of a landmark for the stacked and proposed active shape models. The last row of this table provides the overall complexity of building the profiles utilized for the two models. It is seen from this table that the overall computational complexity of the proposed scheme in building the profile for a landmark is larger than that in the stacked ASM by a factor of log *m*. However, since the value of m is not too large (m = 13 is used for computation of the 2-D profile, and m = 8 is used for computation of the compressed DCT profile), the difference

between the computational complexities of the profiles of two schemes are not that significant. It is also to be noted that, for the proposed ASM, only one compressed DCT profile needs to be built, whereas for the stacked ASM, both 1-D and 2-D profiles are required. In addition, the size of the compressed DCT profile (i.e. n_c) is much smaller than m^2, the size of the 2-D profile of the stacked ASM. Thus, in the proposed scheme, the use of only one compressed profile compared to the use of one 1-D profile and another larger 2-D profile, can be expected to reduce the total number of operations, and therefore the overall computation time for searching a facial shape in an image using the proposed ASM.

Table 3.1: Computational complexity of each step of the process of building the profiles of a landmark for the stacked and proposed ASMs.

Steps	Number of arithmetic operations		Compressed 2-D DCT Profile in the proposed ASM
	Stacked ASM		
	1-D Profile	2-D Profile	
Computation of intensity gradients /profile matrix	$24m+35$	$17m^2$	$17m^2$
Calculation of DCT coefficients	n/a	n/a	$4m^2\log_2(m)-2m^2+2m$
Normalization /Equalization	$2m+2$	$4m^2+3$	$4n_c+3$
Total number of arithmetic operations	$21m^2+26m+40$		$4m^2\log_2(m)+15m^2+2m + 4n_c+3$
Overall computational complexity	$O(m^2)$		$O(m^2\log m)$

3.5 Summary

In this chapter, we have proposed a low-complexity active shape model (ASM), which employs a novel 2-D profile of a landmark of a facial shape based on the discrete cosine transform (DCT). The process of building such a profile consists of sampling, computing a profile matrix and its DCT coefficients, and selecting, normalizing and equalizing the DCT coefficients. The grey-level intensities are first sampled from an m x m region around a landmark. A profile matrix of the landmark is obtained next by computing the response of a linear spatial filter at each pixel position in the sampled region. The filter response is given by the sum of products of the coefficients of a filter mask and the corresponding grey-level intensities in the region spanned by the filter mask. Next, a 2-D array of the DCT coefficients is obtained by computing 2-D DCT of the profile matrix. A subset of the DCT coefficients, representing the low-frequency components of the profile matrix, is selected through a zig-zag scanning of the 2-D array of the DCT coefficients. The selected DCT coefficients are normalized by dividing each element by the mean of the absolute values of the DCT coefficients, which, in turn, are equalized by applying a sigmoid transform to each element. The resulting DCT coefficients are finally used as a compressed DCT profile representing the landmark. A numerical example has been provided in order to illustrate the process of building the compressed DCT profile of a landmark by applying each step of the building process to the landmark.

A computational complexity analysis of stacked ASM [35] and that of the proposed ASM has also been carried out. The difference between the two models results from the process of building the profile(s) of a landmark used for searching a facial shape in an image. The overall computational complexity of the proposed ASM is theoretically

higher than that of the stacked ASM by a factor of log m. However, in practice the computation time of searching a face using the proposed ASM can be expected to be lower than that using the stacked ASM because of the following reasons. The value of m is generally not very large. The proposed ASM requires building only one compressed DCT profile in comparison to the stacked ASM, which requires building a 1-D profile as well as a full-size 2-D profile.

CHAPTER 4

Experimental Results of Applying the Proposed Active Shape Model in a Facial Annotation Application

In Chapter 3, we have proposed a low-complexity active shape model (ASM) that utilizes a 2-D profile of a landmark of a facial shape based on the discrete cosine transform (DCT). Due to the energy compaction property of the DCT, the size of the compressed DCT profile could be made much smaller than that of the conventional spatial-domain 2-D profile. Searching a facial shape in an image using the proposed ASM is thus expected to be faster than that using stacked ASM [35]. In this chapter, we study the effectiveness of the proposed ASM in an application of automatic facial landmark annotation of frontal faces using a facial shape search method introduced in Chapter 2. In order to build the point distribution model (PDM) and the local grey-level gradient model (LGGM) parts of the proposed ASM and to perform the facial shape search on a variety of images, training and test sets of samples are chosen from three databases. The effectiveness of the proposed ASM is analyzed by using an evaluation method to determine the fitting accuracy of the model shape, as well as by measuring the computation time of the facial shape search.

In Section 4.1, the facial shape search method, using the proposed ASM, for finding the best location of the landmarks of a face in an image is presented. Three

databases used for training and test of the proposed ASM in the face annotation application are presented in Section 4.2. An evaluation method utilized for computing the fitting accuracy of the proposed ASM is described in Section 4.3. In Section 4.4, an empirical study is conducted for determining the number of DCT coefficients and for selecting a suitable filter mask. The experimental results of an automatic facial landmark annotation of frontal faces using the proposed and two other active shape models are presented and compared.

4.1 Facial Annotation using the Proposed Active Shape Model

Active shape model (ASM) can be utilized for searching a facial shape in an image by automatically finding the suitable locations of landmarks of a face. It consists of two sub-models, point distribution model (PDM) and local grey-level gradient model (LGGM). The point distribution model can generate any shape similar to the shapes in the training set by controlling the model parameters given by a vector $\mathbf{b}$, which is a vector applied to the eigenvectors obtained by the principal component analysis of the training shapes. In order to transform the shape in the model space to the image space, a similarity transform T is used for translating, scaling, and rotating the shape by setting the transformation parameters (t_x, t_y, s, θ), where (t_x, t_y), s and θ are, respectively, the translation, scaling and rotation parameters. The point distribution model along with the similarity transform is utilized for approximating a given target shape through an appropriate choice of the model parameters $\mathbf{b}$ and the transformation parameters (t_x, t_y, s, θ). The local grey-level gradient model, which employs a 1-D profile to capture the grey-level intensity gradient

59

information in the vicinity of each landmark, is used for determining a suitable location for each landmark of the model shape. In order to closely fit the model shape to a given face, the shape search iteratively improves the fit of the model shape to the given face by moving each landmark of the current model shape to a new position using the LGGM to generate a new shape $\mathbf{x_S}$, and then updating the model parameters $\mathbf{b}$ and the transformation parameters (t_x, t_y, s, θ) with the new sets of parameters that closely approximate $\mathbf{x_S}$.

In our scheme of face annotation using ASM, we employ the compressed DCT profile for building the LGGM. This 2-D profile captures more information around each landmark of the model shape than the 1-D profile, and the size of the 2-D profile is much smaller than that of the conventional spatial-domain 2-D profile. The shape search method using the proposed ASM is thus expected to be faster than that using the stacked ASM [35]. A detailed description of various steps of the shape search method employing the proposed ASM is given in the following sub-sections.

4.1.1. Generating an Initial Model Shape

The search method begins with localizing a rectangular region in a given image containing a face using the Viola Jones face detector [43]. The rectangular region, specified by its center point (x_c, y_c), width w, and height h, is used for determining the model parameters $\mathbf{b}$ and the transformation parameters (t_x, t_y, s, θ) to generate an initial model shape. The mean shape $\bar{\mathbf{x}}$ of the training set is mapped onto the image using the center of the region (x_c, y_c) as the reference point. The resulting shape is then scaled and

rotated by assigning a value to the scaling parameter s and the rotation parameter θ depending on the value of w and h, so as to approximately fit it in the rectangular region. The initial model shape represented by the model parameter $\mathbf{b_0} = \mathbf{0}$ and the transformation parameters $(t_x, t_y, s, \theta)_0$ is utilized as an initial shape in LGGM for generating a refined shape.

4.1.2. Refining the Landmark Positions using LGGM

Each landmark of the shape, represented by the model parameters $\mathbf{b}$ and the transformation parameters (t_x, t_y, s, θ), is moved to a suitable position determined by using the local grey-level gradient model (LGGM) for the landmark. A new shape $\mathbf{x_S}$, which describes the facial shape in the image better than the current model shape does, is formed by new positions of the landmarks.

In the proposed ASM, the LGGM is built by using the compressed DCT profile described in Section 3.2. The process of building the profile for the ith landmark starts with sampling the grey-level intensities from an 8 x 8 region around the landmark. A profile matrix $\mathbf{G_i}$ of the landmark is obtained by computing the response of a linear spatial filter, which captures the forward intensity variations along both in the x- and y-directions, corresponding to the pixels in a 3 x 3 mask, at each pixel position in the square region. A 2-D array of the DCT coefficients $\mathbf{C_{G_i}}$ is obtained by computing 2-D DCT of $\mathbf{G_i}$ using a fast 1-D FFT algorithm given in [47] with row-column decomposition. A subset of n_c DCT coefficients representing the low-frequency components of $\mathbf{G_i}$ is selected by using the zig-zag scanning of $\mathbf{C_{G_i}}$ and stored in an array

$\mathbf{Y}_{n_c i}$ for each landmark i. The coefficients in $\mathbf{Y}_{n_c i}$ are normalized by dividing each element by the mean of the absolute values of the DCT coefficients contained in $\mathbf{Y}_{n_c i}$. These coefficients are then equalized by applying a sigmoid transform to each element. An array of the resulting DCT coefficients $\mathbf{Y}_i''$ is used as the compressed DCT profile of the ith landmark. In order to determine a suitable position for the ith landmark, the cost of fit of $\mathbf{Y}_i''$ to the mean profile $\overline{\mathbf{Y}_i''}$ of the LGGM is evaluated at each candidate position in a 5 x 5 region around the landmark using the Mahalanobis distance given by

$$f(\mathbf{Y}_i'') = (\mathbf{Y}_i'' - \overline{\mathbf{Y}_i''})^T \mathbf{S}_i^{-1} (\mathbf{Y}_i'' - \overline{\mathbf{Y}_i''}) \qquad (4.1)$$

where $\mathbf{S}_i$ is the covariance matrix for the ith landmark. The position d_{best} at which the profile $\mathbf{Y}_i''$ yields the smallest Mahalanobis distance (the minimum cost of fit) is then selected as the new position of the landmark. The process is repeated for each landmark of the model shape in order to obtain a new shape $\mathbf{x}_S$.

4.1.3. Finding a Model Shape by Calculating the Model and Transformation Parameters

The model parameters $\mathbf{b}_{best}$ and the transformation parameters $(t_{xbest}, t_{ybest}, s_{best}, \theta_{best})$ that best approximate the new shape $\mathbf{x}_S$ obtained in the previous sub-section are determined by minimizing the expression given by (2.7) using the similarity transform T and the point distribution model. In order for the model shape to be consistent with those in the training set, the model parameters $\mathbf{b}_{best}$ are constrained according to (2.6).

4.1.4. Updating the Model and Transformation Parameters

The model parameters $\mathbf{b}$ and the transformation parameters (t_x, t_y, s, θ) are then updated to the previously determined model and transformation parameters $\mathbf{b}_{best}$ and $(t_{xbest}, t_{ybest}, s_{best}, \theta_{best})$ which closely approximate the new shape $\mathbf{x}_S$ as

$$t_x \leftarrow t_{xbest} \tag{4.2}$$

$$t_y \leftarrow t_{ybest} \tag{4.3}$$

$$\theta \leftarrow \theta_{best} \tag{4.4}$$

$$s \leftarrow s_{best} \tag{4.5}$$

$$\mathbf{b} \leftarrow \mathbf{b}_{best} \tag{4.6}$$

The updated model shape is then used as the starting shape in the second iteration of the shape search method. The steps of sub-sections 4.1.2, 4.1.3, 4.1.4 described above are repeated until the iterative process satisfies a pre-specified terminating condition.

4.2 Databases

In order to perform the facial annotation using the scheme described in the previous section, we need a training set and a test set of images and landmarks of the facial shapes. The fitting error between the landmarks of the model shape and the manually labelled landmarks from the test set is computed for evaluating the effectiveness of the proposed ASM. In order to build the training and test sets, various samples are chosen from three databases, namely, the Milborrow / University of Cape Town (MUCT) database, the

BioID database and the IMM database, containing both images and the landmarks of the corresponding facial shapes.

The Milborrow / University of Cape Town (MUCT) database presented in [54] contains 3755 samples of 276 different individuals, each of which consists of a facial image with 480 x 640 pixels and 76 manually labelled landmarks. Figure 4.1 shows a few examples of the samples from this database consisting of a facial image and the 76 landmarks representing different parts of the face in an image.
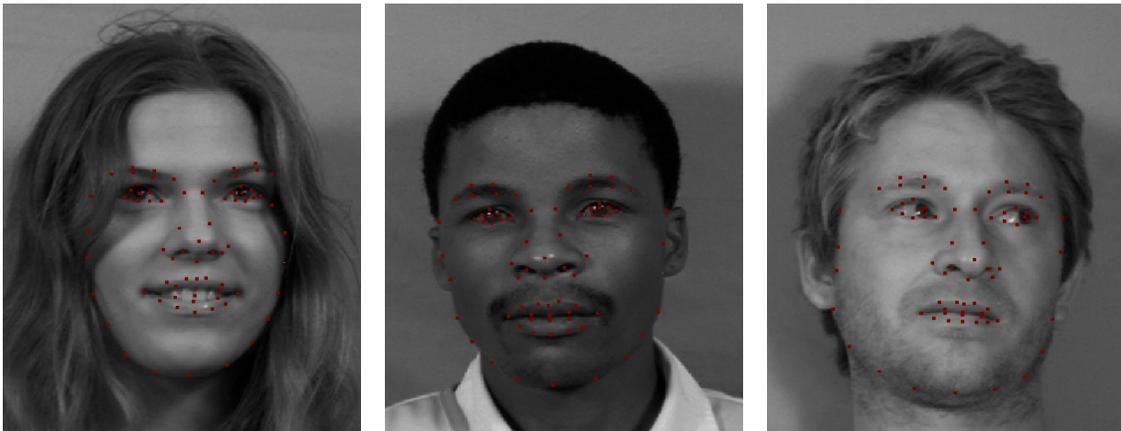
Figure 4.1: Examples of the samples from the MUCT database consisting of a facial image and 76 manually annotated landmarks representing particular parts of a face in the image [54].

In order to generate the samples, each individual is photographed from five different positions of the camera, as shown in Figure 4.2. Since we are interested in an application of automatic facial landmark annotation of frontal faces, only the images photographed from the frontal camera positions (the camera positions a, d and e) are utilized. Figure 4.3 shows examples of the images photographed from the camera positions a, d and e. Each individual is also photographed under three lighting conditions, q, r and s. Figure 4.4

shows examples of the images photographed under the lighting conditions q, r and s from the fontal position a.
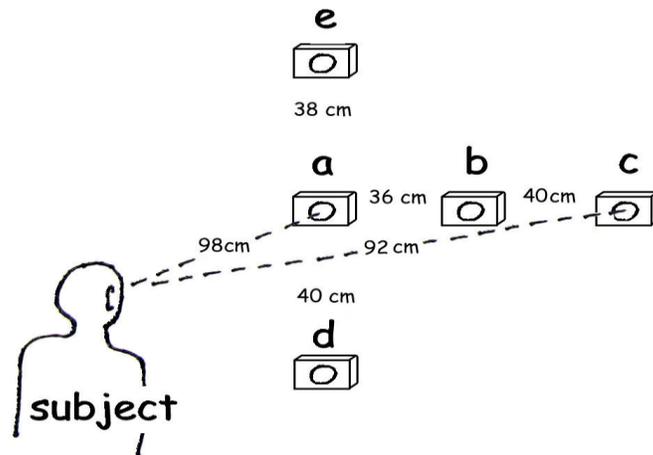
Figure 4.2: The five different positions of the camera [54].

| (a) | (b) | (c) |

Figure 4.3: Examples of the images of an individual photographed from the frontal camera positions under the same lighting condition [54]. (a) An image photographed from the camera position a. (b) An image photographed from the camera position d. (c) An image photographed from the camera position e.

(a)   (b)   (c)

Figure 4.4: Examples of the images of an individual photographed under three different lighting conditions [54]. (a) An image photographed under the lighting condition q. (b) An image photographed under the lighting condition r. (c) An image photographed under the lighting condition s.

The BioID database introduced in [55] consists of 1521 samples of 23 different individuals, each of which contains a facial image with 384 x 286 pixels and 20 manually annotated landmarks. Figure 4.5 shows a few examples of the samples from this database.

Figure 4.5: Some examples of the samples from the BioID database consisting of a facial image and 20 manually annotated landmarks [56].

The informatics and mathematical modelling (IMM) database presented in [57] contains 240 samples of 40 different individuals, each of which consists of a facial image with 640 x 480 pixels and 58 manually annotated landmarks. A few examples of the samples from this database are shown in Figure 4.6.
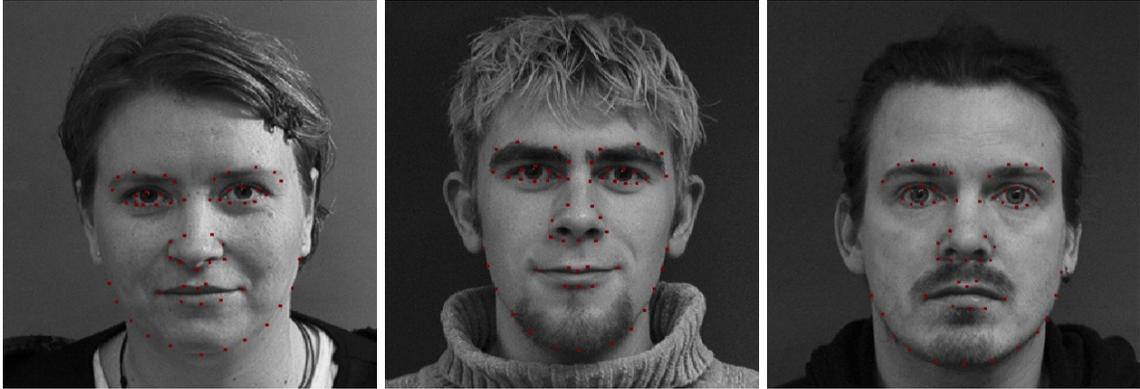
Figure 4.6: Some examples of the samples from the IMM database consisting of a facial image and 58 manually annotated landmarks [57].

In order to double the number of samples in each database, a set of mirrored samples is obtained by horizontally flipping each image in a database using the photo editing software, and by negating the x-coordinate of each landmark and changing the order of the landmarks of each shape in the samples. Figure 4.7 shows an example of the mirrored pairs of the samples from the MUCT database. As a result of generating the mirrored samples, the number of samples in the MUCT, BioID and IMM databases are, respectively, increased to 7510, 1521 and 480 samples. Table 4.1 summarizes the characteristics of the three databases used in our experiment.

<center>(a)   (b)</center>
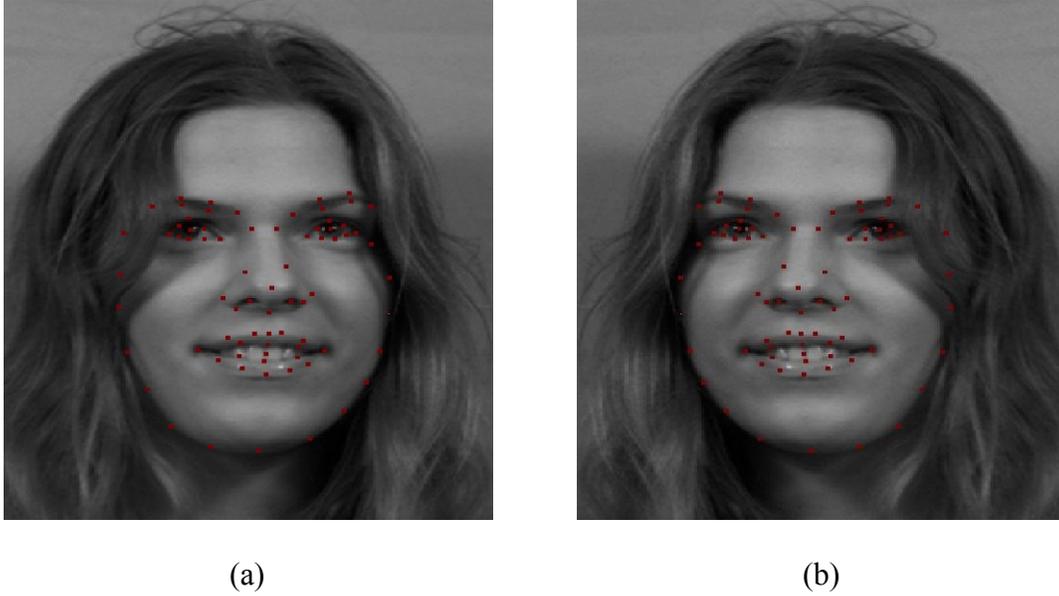
Figure 4.7: An example of the mirrored pairs of the samples from the MUCT database [54]. (a) The image of a sample from the MUCT database. (b) The mirrored image of the same sample.

Table 4.1: Summary of the three databases used for building the PDM and LGGM parts of the proposed ASM and performing the facial shape search using the ASM.

	Total Number of Samples	Number of Landmarks	Number of Individuals	Image Size (Pixels)
MUCT Database	7510	76	276	480 x 640
BioID Database	3042	20	23	284 x 286
IMM Database	480	58	40	640 x 480

4.3 Evaluation Method

In order to measure the effectiveness of the proposed ASM in fitting the landmarks of the

model shape onto a target facial shape in an image, the fitting accuracy is computed by

determining the distance between each landmark in the target shape and its corresponding

<center>68</center>

landmark in the model shape. The distance between the two points in the spatial domain is computed by using the Euclidean distance given by

$$d_i = \sqrt{(\mathbf{a_i} - \mathbf{b_i})^{\mathrm{T}}(\mathbf{a_i} - \mathbf{b_i})} \qquad (4.7)$$

where $\mathbf{a_i}$ is a vector containing the x- and y-coordinates of the ith landmark of the model shape, and $\mathbf{b_i}$ is a vector containing the x- and y-coordinates of the ith landmark of the target shape. Since facial shapes come from diverse conditions and individuals, an evaluation method, namely, the average normalized fitting error, which takes the difference in facial shape size into account, is utilized for computing the fitting accuracy.

The normalized fitting error is obtained by taking the average of the Euclidean distance between the manually annotated landmarks from a target facial shape and the corresponding landmarks from the model shape, and normalizing the result by a factor that ensures that s, the distance between two landmarks representing the extreme points on the left and right eyes, as shown in Figure 4.8, is equal to 50 pixels [36]. These two landmarks are selected, since they are sufficiently far apart to represent the width of a facial shape, and all three databases used in our experiment consist of these landmarks, as seen from Figures 4.1, 4.5, and 4.6. As a result of normalization of the distance, the fitting accuracy obtained by using shapes with different scales can be directly compared. In order to compute the overall fitting accuracy of the proposed ASM, the average normalized fitting error is computed by taking the mean of the normalized fitting error over all the samples from the test set as

$$E_{average} = \frac{1}{N_{test}} \sum_{i=1}^{N_{test}} \left\{ \frac{50}{ns} \sum_{j=1}^{n} d_j \right\} \qquad (4.8)$$

where N_{test} is the number of samples from the test set, and n is the number of landmarks in a sample.
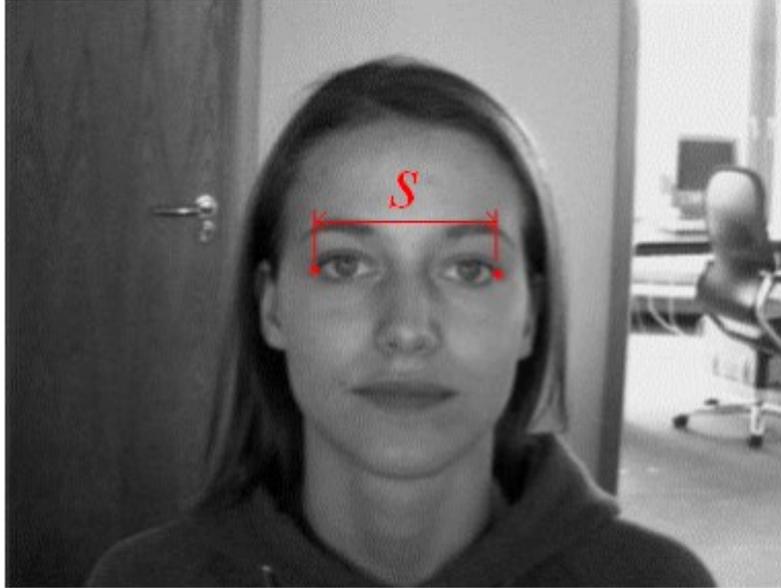
Figure 4.8: Two landmarks representing the extreme points on the left and right eyes, and the distance *s* between the two landmarks.

4.4 Experimental Results

In order to study the effectiveness of the proposed ASM in the application of an automatic facial landmark annotation of frontal faces, the facial annotation using the scheme described in Section 4.1 is performed on various facial images. In this section, we conduct an experiment using two groups of training and test sets of images and landmarks of the facial shapes. The effectiveness of the ASM is evaluated by computing the average normalized fitting error, and by measuring the execution time of the facial shape search. An empirical study is conducted in order to select the optimal number of DCT coefficients and to choose a suitable filter mask, which are used for building the

compressed DCT profile of the ASM. We also present and compare the experimental results of an automatic facial landmark annotation of frontal faces using the proposed and various other active shape models, namely, stacked ASM [35] and ASM with a PCA-based LGGM [36].

4.4.1. Experimental Settings

In order to examine the performance of the proposed ASM experimentally, two groups of training and test sets of images and landmarks of the facial shapes are selected from the three databases presented in Section 4.2. The first group of samples consists of a training set of 3000 samples from the MUCT database and a test set of 1471 samples from the BioID database. As seen from Table 4.1, the number of landmarks in each sample of the MUCT database and that in the BioID database are, respectively, 76 landmarks and 20 landmarks. Although the number of landmarks in the former is larger than that in the latter, 17 of these landmarks, as shown in Figure 4.9, are common in each of samples in the two databases. Thus, these 17 landmarks ($n = 17$) are utilized for computing the average normalized fitting error in the experiments using this group of samples. The second group of samples consists of a training set of 240 samples of the first 20 individuals of the IMM database and a test set of 240 samples of the other 20 individual of the same database. Since both the training and test sets of samples are chosen from the same database, all the landmarks of a facial shape ($n = 58$) are used for computing the average normalized fitting error in the experiments using this group of samples. Table 4.2 summarizes the two groups of training and test sets of images and landmarks of the facial shapes belonging to the three databases.
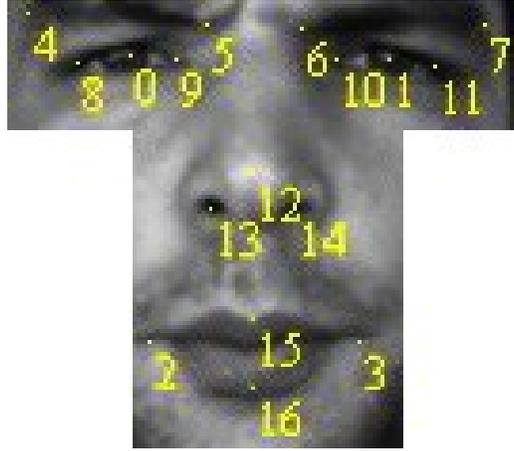
Figure 4.9: The 17 landmarks internal to the face utilized for computing the average normalized fitting error in the experiment with the first group of samples [59].

Table 4.2: Summary of the number of training and test sets in the two groups of the images chosen from the three databases and the number of landmarks in each facial shape.

	Group 1		Group 2	
	Training Set	**Test Set**	**Training Set**	**Test Set**
Database	MUCT	BioID	IMM	
Number of Samples	3000	1471	240	240
Number of Landmarks	76	20	58	

Publicly available STASM software [59] is modified and utilized to produce the experimental results for the proposed and various other active shape models under the same experimental condition. The active shape models used in the experiments are all implemented by using Microsoft Visual Studio 2012 and the C++ programming language. The experiments are performed on a 2.6 GHz Intel Core i7 CPU with 6-GB RAM and Windows 7 operating system.

The execution time of a facial landmark annotation of the frontal face in an image is given by the sum of the execution time for locating the face in the image and the execution time for conducting the facial shape search in the image. In order to conduct

the experiments under the same condition for all the ASMs, the location of a rectangular region containing the face in each test image is pre-computed using the Viola-Jones face detector [43]. Thus, the execution time for locating the face in the image is a constant time for loading the pre-computed location of the region containing the face. The average execution time is computed by taking the average of the execution time over all the test samples after the first 50, since the initialization of the ASM may have an effect on the computation time for some samples at the beginning of the execution.

4.4.2. Results and Analysis of the Proposed Active Shape Model

As described in Chapter 3, the proposed ASM employs a compressed DCT profile of a landmark of a facial shape. In order to build the compressed DCT profile of the ith landmark, the profile matrix $\mathbf{G_i}$ is obtained by calculating the response of a spatial filter mask at each pixel position of an m x m region around the landmark. A 2-D array $\mathbf{C_{G_i}}$ of DCT coefficients is obtained by computing 2-D DCT of $\mathbf{G_i}$, and a subset of n_c DCT coefficients is selected through a zig-zag scanning of $\mathbf{C_{G_i}}$. The selected DCT coefficients are normalized by dividing each element by the mean of the absolute values of the selected DCT coefficients, which are subsequently equalized by applying a sigmoid transform to each element. The resulting DCT coefficients are finally used as the compressed DCT profile of the landmark. The number of DCT coefficients n_c and the filter mask used in building the profile are important choices for the proposed ASM, since they have an effect on the fitting accuracy and the computational complexity.

The proposed ASM is utilized in an automatic facial annotation system that

automatically locates facial features of the face in an input image. The facial annotation system maps the initial model shape onto the input image using a rectangular region containing a face, reforms the model shape iteratively, and finally yields the model shape that fits closely to the face in the image. An example of the initial and final model shapes of the automatic facial annotation system is shown in Figure 4.10. In our experiments, the proposed ASM is first built by using a training set of samples, and the facial annotation is performed on each image in a test set of samples using different combinations of the number of DCT coefficients and a filter mask. Three different filter masks, given in Figure 3.4 (d), (e) and (f), are used for generating the profile matrix. The number of DCT coefficients utilized for building the compressed DCT profile is varied from 10 to 35 with increments of 5. The fitting accuracy of the model shape to the face in the image is also computed by using the normalized fitting error presented in Section 4.3. In order to study the performance of the facial annotation using the proposed ASM, we run an experiment using the two groups specified in Table 4.2 of training and test sets of images and landmarks of the facial shapes specified in Section 4.4.1, and the average normalized fitting error is computed in each experiment.

Table 4.3 lists the average normalized fitting error obtained from the experiment using the training and test sets of Group 1. It is seen from this table that the lowest average normalized fitting error results from using 25 DCT coefficients for each of the three filter masks. It is also observed that when the number of the DCT coefficients is made smaller or larger, the error in facial annotation becomes progressively larger. The overall lowest average normalized fitting error of 1.5071 is obtained by using the 4-Laplacian filter mask with 25 DCT coefficients.

74

| (a) | (b) |

Figure 4.10: An example of the initial and final model shapes of the automatic facial annotation system. (a) The initial model shape (shown in red color) mapped onto an input image using a rectangle region containing a face (shown in yellow color). (b) The final model shape (shown in red color) that fits closely to the face.

Table 4.3: Average normalized fitting error obtained from the experiment with Group 1 of samples by using different combinations of the number of DCT coefficients and a filter mask.

Number of DCT Coefficients	Average Normalized Fitting Error (pixels)		
	Type of Filter Masks		
	Gradient (Figure 3.4 (d))	4-Laplacian (Figure 3.4(e))	8-Laplacian (Figure 3.4 (f))
10	2.0000	1.8738	1.8838
15	1.6868	1.6408	1.6284
20	1.5738	1.5492	1.5595
25	**1.5221**	**1.5071**	**1.5273**
30	1.5611	1.5222	1.5341
35	1.5685	1.5323	1.5452

Table 4.4 gives the average normalized fitting error obtained from the experiment using the training and test sets of Group 2. It is seen from this table that the lowest average normalized fitting error is obtained by using 30 DCT coefficients for the gradient

and 4-Laplacian filter masks, and that by using 20 DCT coefficients for the 8-Laplacian filter mask. The overall lowest average normalized fitting error of 3.0520 is obtained by using the gradient filter mask with 30 DCT coefficients.

Table 4.4: Average normalized fitting error obtained from the experiment with Group 2 of samples by using different combinations of the number of DCT coefficients and a filter mask.

| Number of DCT Coefficients | Average Normalized Fitting Error (pixels) | | |
| | Type of Filter Masks | | |
	Gradient (Figure 3.4(d))	4-Laplacian (Figure 3.4 (e))	8-Laplacian (Figure 3.4 (f))
10	3.2982	3.4735	3.3901
15	3.1746	3.2606	3.2117
20	3.1287	3.1902	**3.1401**
25	3.0652	3.2298	3.2110
30	**3.0520**	**3.1895**	3.1767
35	3.0706	3.2071	3.1454

It should be noted that choice of the filter mask used for building the compressed DCT profile has an effect not only on the average normalized fitting error but also on the computational complexity. As seen from Figures 3.4 (d), (e) and (f), each computation of the filter response using the gradient filter mask requires three multiplications and two additions, while that using the 4-Laplacian filter mask requires 5 multiplications and 4 additions and that using the 8-Laplacian filter mask requires 9 multiplications and 8 additions. Thus, from the view point of the number of arithmetic operations involved with the filter masks, one would tend to select the gradient filter mask provided that the fitting error using this mask is acceptable. Hence, we need to examine as to how this choice of the gradient filter mask will impact the achievement of the lowest average normalized fitting error. In the experiment using Group 1 samples, it is seen from Table

4.3 that although the lowest average normalized fitting error is obtained by using the 4-Laplacian filter mask with 25 DCT coefficients, the error changes only insignificantly when the mask is changed to a gradient filter mask with the same number of DCT coefficients, i.e., $n_c = 25$. On the other hand, in the experiment using Group 2 samples, it is seen from Table 4.4 that the lowest average normalized fitting error happens to be achieved by using the gradient filter mask but with 30 DCT coefficients. If we reduce the number of coefficients from 30 to 25 using this mask, the error does not change significantly. Thus, a change of the number of DCT coefficients to $n_c = 25$ will make the choice of the mask and the number of DCT coefficients for the Group 2 samples in line with that for the Group 1 samples with only negligible loss in the value of the lowest average normalized fitting error. Therefore, from the results of the experiments, one can conclude that the use of the gradient filter mask with 25 DCT coefficients used for building the compressed DCT profile provides good performance as well as computational efficiency for either group of the training and test samples.

4.4.3. Comparison of the Facial Annotation Performance using the Proposed and Other Active Shape Models

In order to examine the comparative effectiveness of the proposed ASM, we apply this and other ASMs to the problem of facial annotation and compare the performance in terms of the average normalized fitting errors and the average execution times. We conduct experiments using the proposed ASM, stacked ASM [35], and ASM with a PCA-based LGGM [36]. Table 4.5 summarizes the parameters utilized for a facial landmark annotation using the three ASMs.

Table 4.5: Summary of the parameters utilized for a facial landmark annotation using the stacked ASM, ASM with a PCA-based LGGM and the proposed ASM.

Parameter	Description of Parameter	Stacked ASM	ASM with a PCA-based LGGM	Proposed ASM
m	1D profile size	17	-	-
m x m	2D profile size	169	169	25
n_c	Number of DCT coefficients	-	-	25
n_s	1D search length	7	-	-
n_s x n_s	2D search region	5x5	5x5	5x5
c	Shape constant for sigmoid transform	100	100	100
$3\sqrt{\lambda_k}$	Shape coefficients bound	1.8 (1D), 1.4 (2D)	1.4	1.4
N_{max}	Maximum number of iterations per pyramid level	4	4	4
N_{level}	Number of levels for multi-resolution search	4	4	4

As in the previous section, the two groups of training and test sets of images and landmarks of the facial shapes described in Section 4.4.1 are utilized for conducting experiments for the various ASMs. The facial landmark annotation is conducted 10 times for each ASM using the Group 1 samples and then using the Group 2 samples, and the mean values of the average normalized fitting errors and the average execution times are obtained and compared. The execution time of a facial landmark annotation corresponds to the time required for fitting the landmarks of the model shape onto a target facial shape in an image. The average execution time is computed by taking the mean of the complete execution time over all samples from the test set.

Table 4.6 gives the mean values of the average normalized fitting error and the average execution time obtained from the experiments using the training and test sets of Group 1. It is observed from this table that the lowest average normalized fitting error is

obtained by using the proposed ASM. It is also to be noted that the average execution time obtained by using the proposed ASM is much smaller than that obtained by using the stacked ASM and ASM with a PCA-based LGGM.

Table 4.6: Mean values of average normalized fitting error and average execution time obtained by using the proposed and two other active shape models from the experiment using the training and test sets of Group 1 (the MUCT and BioID databases).

Method	Average Normalized Fitting Error (pixels)	Average Execution Time per Facial Image (seconds)
Stacked ASM [35]	1.5807	0.1251
ASM with a PCA-based LGGM [36]	1.8033	2.8076
Proposed ASM	**1.5220**	**0.0675**

Table 4.7 presents the mean values of the average normalized fitting error and the average execution time obtained from the experiments using the training and test sets of Group 2. It is seen from this table that the lowest average normalized fitting error is obtained by using the proposed ASM. It is also to be noted that the average execution time obtained again by using the proposed ASM is much smaller than that obtained by using the other two ASMs used for comparison.

Table 4.7: Mean values of average normalized fitting error and average execution time obtained by using the proposed and two other active shape models from the experiment using the training and test sets of Group 2 (the IMM database).

Method	Average Normalized Fitting Error (pixels)	Average Execution Time per Facial Image (seconds)
Stacked ASM [35]	3.4666	0.1406
ASM with a PCA-based LGGM [36]	3.2164	2.1565
Proposed ASM	**3.0652**	**0.0784**

The block size of 8 x 8 pixels around a landmark used in the proposed ASM is sufficient to capture all the necessary features of the landmark, and the use of larger-size block only increases the computational complexity without improving the fitting accuracy. This is supported by the accuracy of the results using the 8 x 8 block of pixels around the landmark.

It is also seen from Tables 4.6 and 4.7 that the use of a smaller number of DCT coefficients significantly reduces the computational complexity, and, in turn, the average execution time. Specifically, the average execution time obtained by using the proposed ASM for the two groups of samples is 45.07% lower than that using the stacked ASM and 97.06% lower than that using ASM with a PCA-based LGGM. Since the reduced set of DCT coefficients chosen by performing a zig-zag scanning of the 2-D array of the DCT coefficients contains most of the pertinent information about the 8 x 8 region around a landmark, the reduction in the computational complexity resulting from the use of this reduced set is achieved without compromising on the fitting accuracy.

The noise is generally associated with the high-frequency components of the DCT coefficients of an image. Thus, in the proposed method, the process of choosing the low-frequency DCT coefficients filters out the noise contained in the image. In conclusion, the use of the proposed ASM in a facial landmark annotation significantly reduces the execution time without compromising on the shape fitting accuracy.

4.5 Summary

In this chapter, we have examined the effectiveness of the proposed active shape model (ASM) presented in Chapter 3 in an application of automatic facial landmark annotation

of frontal faces. The proposed ASM is utilized in a facial shape search method described in Chapter 2 for localizing a facial shape in an image. The steps of such a method consist of generating an initial model shape, refining the landmark positions using the local grey-level gradient model (LGGM), finding a model shape by calculating the model and transformation parameters, and finally, updating the model and transformation parameters. The initial model shape is generated by roughly fitting a shape, which is given by the model and transformation parameters, to a rectangular region containing the face in an input image. The shape search method iteratively improves the fitting of the model shape to the face in the image by moving each landmark of the model shape to a new position using LGGM for generating a new shape, and by updating the model and transformation parameters for closely approximating the new shape.

In order to perform the facial annotation using the proposed ASM, two groups of training and test sets of images and landmarks of the facial shapes are created from the samples of three databases, namely, the MUCT, BioID and IMM databases. The accuracy in fitting the landmarks of the model shape onto a target facial shape in an image is measured by using the normalized fitting error. The average normalized fitting error over all the samples from the test set is computed for measuring the overall fitting accuracy of the proposed ASM.

For measuring the effectiveness of the proposed ASM quantitatively, we have run experiments with the two groups of training and test sets of images and landmarks of the facial shapes using different combinations of the number of DCT coefficients and a filter mask. From the results of the experiments, we have concluded that the gradient filter mask with 25 DCT coefficients for building the compressed 2-D DCT profile for the

proposed ASM provides good performance in terms of the fitting accuracy as well as in terms of the computational efficiency. We have also compared the average normalized fitting error and the average execution time obtained by using the proposed and two other ASMs, namely, stacked ASM and ASM with a PCA-based LGGM. The experimental results have shown that the lowest average normalized fitting error is obtained by using the proposed ASM, and the average execution time obtained by using the proposed ASM is much smaller than that obtained by using the other two ASMs.

The block size of 8 x 8 pixels around a landmark is sufficient to capture all the necessary features of the landmark. Hence, the use of larger-size block (for example, 13 x 13 block of pixels used in the other two methods) would only increase the computational complexity without improving the fitting accuracy. The proposed compressed 2-D DCT profile has provided two distinct advantages: (i) The use of a smaller number of DCT coefficients significantly reduces the computational complexity. This is achieved without sacrificing the fitting accuracy, since the reduced set of the DCT coefficients is chosen in such a way that they contain most of the pertinent information about the 8 x 8 region around a landmark. (ii) The process of choosing the low-frequency DCT coefficients filters out the noise contained in the image, since the noise is generally associated with the high-frequency components of the DCT coefficients, which are discarded in the proposed scheme. Hence, the use of compressed DCT profile has, in fact, reduced the fitting error as well. Hence, it can be concluded that the proposed active shape model is an attractive and viable alternative to the existing ASMs.

CHAPTER 5

Conclusion and Future Work

5.1 Concluding Remarks

Facial recognition systems, which automatically identify a human in a digital image or a video frame using faces stored in a database, have been successfully applied in security, law-enforcement and human identification applications. In a feature-based facial recognition system, which employs a set of features extracted from each of the prominent facial components, an automatic and accurate localization of facial features is an essential pre-processing step. A flexible shape model, referred to as active shape model (ASM), was originally proposed to automatically locate a set of landmarks representing the facial features. Various improved versions of ASM for facial landmark annotation have been developed in order to increase the shape fitting accuracy. However, these improvements have been achieved at the expense of a substantial increase in the computational complexity.

In this thesis, a low-complexity ASM has been developed by incorporating the energy compaction property of the discrete cosine transform (DCT). The proposed ASM utilizes a novel 2-D profile of a landmark, which is based on the DCT, in order to reduce the computational complexity without affecting the facial shape fitting accuracy. The development of the proposed ASM has been aimed at reducing the execution time of a

facial shape search while keeping the shape fitting accuracy of the ASM comparable to that of other improved versions of ASM.

The proposed ASM has been developed by introducing a compressed DCT profile of a landmark, which is a 2-D profile based on the DCT of the local grey-level gradient pattern around the landmark. In order to build the profile of the landmark, a profile matrix of the landmark is first obtained by computing the response of a spatial filter mask at each pixel position of an m x m region around the landmark. A 2-D array of the DCT coefficients is obtained by computing 2-D DCT of the profile matrix, and then a subset of these coefficients is chosen by performing a zig-zag scanning of the 2-D array of the DCT coefficients. The selected DCT coefficients are normalized by dividing each element by the average of the absolute values of the selected DCT coefficients, which are, in turn, equalized by applying a sigmoid transform to each element. The resulting DCT coefficients are finally utilized as a compressed DCT profile representing the landmark.

The proposed ASM has been employed in a scheme of facial landmark annotation for locating facial features of the face in an input image. This scheme first maps the initial model shape onto the input image using a rectangular region containing a face, refines the model shape iteratively, and, finally yields a model shape that fits closely to the face in the image. In order to perform the facial annotation using this scheme, two groups of training and test sets of images and landmarks of the facial shapes have been created from the samples of three databases, namely, the MUCT, BioID, and IMM databases. Experiments with the two groups of training and test sets of samples have been conducted using different combinations of the number of DCT coefficients and a filter mask. It has been concluded from the experimental results that the gradient filter mask

with 25 DCT coefficients provides good performance in terms of the fitting accuracy as well as in regard to the computational efficiency. The average normalized fitting error and the average execution time obtained by using the proposed and two other ASMs, namely, stacked ASM and ASM with a PCA-based LGGM, have also been compared in order to examine the comparative effectiveness of the proposed ASM. The experimental results have shown that the lowest average normalized fitting error is obtained by using the proposed ASM, and the average execution time obtained by using the proposed ASM is significantly lower than that obtained by using the other two ASMs. It has been noted from the experimental results that the use of a smaller number of DCT coefficients in building the compressed DCT profile significantly reduces the computational complexity, and the process of choosing the low-frequency DCT coefficients filters out the noise contained in the image, which is associated with the high-frequency coefficients.

In conclusion, this study has shown that the use of the proposed ASM in an automatic facial landmark annotation of frontal faces has significantly reduced the computational complexity without losing the shape fitting accuracy.

5.2 Future Work

In this study, a fixed-size window has been used for generating the profile matrix. One could also undertake a study in which the size of the window is varied depending on the location of a landmark within the face from the view point of improving the fitting accuracy further. For the development of the proposed ASM, an exhaustive search in a fixed region has been used for finding a new position of each landmark of the model shape. One could use non-exhaustive search techniques in order to reduce the search time.

Effectiveness of the proposed ASM based on a compressed DCT profile has been examined through its application in facial annotation problems. Other applications, such as object segmentation in medical images, can also be considered to examine whether their performance could possibly be improved from using the proposed ASM.

References

[1]   R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: a survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705-740, May, 1995.

[2]   M. Kirby and L. Sirvoich, "Application of the Karhunen-Loeve procedure for the characterization of human faces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp.103–108, Jan., 1990.

[3]   M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proc. Int. Conf. on Comp. Vision and Pattern Recognition.*, Maui, HI, 1991, pp. 586–591.

[4]   S. Akamatsu, T. Sasaki, H. Fukamachi, and Y. Suenaga, "A robust face identification scheme—KL expansion of an invariant feature space," in *Proc. SPIE, Intell. Robots and Comput. Vision X: Algorithms and Techniques*, Boston, MA, 1991, vol. 1607, pp. 71–84.

[5]   A. Pentland, B. Moghaddam, T. Starner, and M. Turk, "View-based and modular eigenspaces for face recognition," in *Proc. IEEE Computer Soc. Conf. On Comput. Vision and Pattern Recognition*, Seattle, WA, 1994, pp. 84–91.

[6]   M. Lades, J. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, and R. Wurtz, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300–311, Mar., 1993.

[7]   D. Swets and J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug., 1996.

[8]   P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul., 1997.

[9]   R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*. New York, NY: Wiley, 1973.

[10]   T. Sakai, M. Nagao, and S. Fujibayashi, "Line extraction and pattern recognition in a photograph," *Pattern Recognition.* vol. 1, no. 3, pp. 233–23, Mar., 1969.

[11]   M. Kelly, "Visual identification of people by computer," Ph.D. dissertation, Dept. Comp. Sci., Stanford Univ., Stanford, CA, 1970.

[12] V. Govindaraju, S. Srihari, and D. Sher, "A Computational model for face location," in *Proc. 3rd Int. Conf. on Comput. Vision*, Osaka, Japan, 1990, pp. 718–721.

[13] A. Yuille, D. Cohen, and P. Hallinan, "Feature extraction from faces using deformable templates," in *Proc. IEEE Comput. Soc. Conf. on Comput. Vision and Pattern Recognition*, San Diego, CA, 1989, pp. 104–109.

[14] P. Burt, "Multiresolution techniques for image representation, analysis, and 'smart' transmission," in *Proc. SPIE, Visual Comm. and Image Proc. IV.*, Philadelphia, PA, 1989, vol. 1199, pp. 2–15.

[15] I. Craw, D. Tock, and A. Bennett, "Finding face features," in *Proc. 2nd Europe. Conf. on Comput. Vision*, Santa Margherita Ligure, Italy, 1992, vol. 588, pp. 92–96.

[16] J. Shepherd, "An interactive computer system for retrieving faces," in *Proc. NATO Advanced Research Workshop*, Aberdeen, Scotland, U.K., 1986, vol. 28, pp. 398–409.

[17] T. Kanade, "Picture processing system by computer complex and recognition of human faces," Ph.D. dissertation, Dept. Info. Sci., Kyoto Univ., Japan, 1973.

[18] L.D. Harmon and W.F. Hunt, "Automatic recognition of human face profiles," *Comput. Graph. and Image Process.* vol. 6, no. 2, pp. 135–156, Apr., 1977.

[19] R. Brunelli, and T. Poggio, "Face recognition: Features versus templates," *IEEE Trans. on Pattern Anal. and Mach. Intell.*, vol. 15, no. 10, pp. 1042–1052, Oct., 1993.

[20] Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform," *Int. J. Comput. Vis.* vol. 43, no. 3, pp. 167-188, Aug., 2001.

[21] A. L. Yuille, P. Hallinan and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. J. Comput. Vis.*, vol. 8, no. 2, pp. 99-112, Aug., 1992.

[22] M. Kass, A. Witkin and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 259-268, Jan., 1988.

[23] G. E. Hinton, C. K. I. Williams and M. D. Revow, "Adaptive Elastic Models for Hand-Printed Character Recognition," in *Advances in Neural Inform. Process. Syst.*, San Mateo, CA, 1992, vol. 4, pp. 512-519.

[24]   L. H. Staib and J. S. Duncan, "Parametrically Deformable Contour Models," in *Proc. IEEE Comput. Soc. Conf. on Comput. Vision and Pattern Recognition*, San Diego, CA, 1989, pp. 427-430.

[25]   A. Pentland and S. Sclaroff, "Closed-Form Solutions for Physically Based Modelling and Recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 7, pp. 715-729, Jul., 1991.

[26]   P. Karaolani, G. D. Sullivan, K. D. Baker and MJ. Baines, "A Finite Element Method for Deformable Models," in *Proc. 5th Alvey Vis. Conf.,* Reading, 1989, pp. 73-78.

[27]   C. Nastar and N. Ayache, "Non-Rigid Motion Analysis in Medical Images: a Physically Based Approach," in *Proc. 13th int. conf. IPMI '93*, Flagstaff, AZ. 1993, pp. 17-32.

[28]   U. Grenander, Y. Chow and D. M. Keenan, *Hands. A Pattern Theoretic Study of Biological Shapes.* New York, NY: Springer-Verlag, 1991.

[29]   K. V. Mardia, J. T. Kent and A. N. Walder, "Statistical Shape Models in Image Analysis," in *Proc. 23rd Symp. on the Interface,* Seattle, WA, 1991, pp. 550-557.

[30]   U. Grenander, M. I. Miller, "Representations of Knowledge in Complex Systems," *J. Roy. Statistical Soc. Series B*, vol. 56, no. 4, pp. 549-603, 1994.

[31]   T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam, "The use of active shape models for locating structures in medical images," *Image and Vision Computing* vol. 12, no. 6, pp. 355–366, Jul., 1994.

[32]   T. F. Cootes and C. J. Taylor, "Statistical Models of Appearance for Computer Vision," Imaging Sci. and Biomedical Eng., Univ. Manchester, Manchester, Tech. Rep., Mar., 2004.

[33]   T. F. Cootes, C. J. Taylor, D. H. Cooper and J. Graham, "Active Shape Models - Their Training and Application," *Comput. Vision and Image Understanding*, vol. 61, no. 1, pp. 38-59, Jan., 1995.

[34]   T. F. Cootes, C. J. Taylor, A. Lanitis, "Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search," in *Proc. 5th British Mach. Vision Conf.*, York, 1994, pp. 327-336.

[35]   S. Milborrow and F. Nicolls. "Locating Facial Features with an Extended Active Shape Model," in *Proc. 10th European Conf. On Comput. Vision*, Marseille, France, 2008, pp. 504-513.

[36] K. Seshadri and M. Savvides, "Robust modified active shape model for automatic facial landmark annotation of frontal faces," in *Proc. 3rd IEEE Int. Conf. Biometrics: Theory, Applications Systems*, Washington, DC, 2009, pp. 1–8.

[37] W. Wang, S. Shan, W. Gao, B. Cao, B. Yin, "An Improved Active Shape Model for Face Alignment," in *Proc. 4th IEEE Int. Conf. on Multimodal Interfaces*, Pittsburgh, PA, 2002, pp. 523-528.

[38] M. H. Mahoor and M. Abdel-Mottaleb, "Facial Features Extraction in Color Images Using Enhanced Active Shape Model," in *Proc. 7th IEEE Int. Conf. on Automatic Face and Gesture Recognition*, Southampton, 2006, pp. 144-148.

[39] K. Seshadri, M. Savvides, "An analysis of the sensitivity of active shape models to initialization when applied to automatic facial landmarking," *IEEE Trans. Inf. Forens. Security,* vol. 7, no. 4, pp. 1255–1269, Aug., 2012.

[40] S. Milborrow, "Locating Facial Features with Active Shape Models." M.S. thesis. Dept. Elec. Eng., Univ. Cape Town, Cape Town, South Africa, 2007.

[41] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, N.J.: Prentice Hall, 2008.

[42] J. C. Gower, "Generalized Procrustes Analysis," *Psychometrika*, vol. 40, no. 1, pp. 33-51, Mar., 1975.

[43] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. 2001 IEEE Comput. Soc. Conf.*, Kauai, HI, 2001, vol. 1. pp. I-511 - I-518.

[44] S. A. Khayam. (2003, Mar.). *The Discrete Cosine Transform (DCT): Theory and Application* [Online]. Available: http://www.egr.msu.edu/waves/people/ Ali_files/DCT_TR802.pdf

[45] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, N.J.: Prentice Hall, 1999.

[46] S. A. Broughton and K. M. Bryan, *Discrete Fourier analysis and wavelets: applications to signal and image processing*, Hoboken, N.J.: John Wiley & Sons, 2009.

[47] C. Loeffler, A. Ligtenberg, and G. S. Moschytz, "Practical fast 1-D DCT algorithms with 11 multiplications," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, Glasgow, 1989, vol. 29, pp. 988 – 991.

[48]   W. B. Pennebaker and J. L. Mitchell, *JPEG: Still image data compression standard*. Norwell, MA: Kluwer Academic Pub., 1992.

[49]   Z. M. Hafed and M. D. Levine, "Face recognition using the discrete cosine transform", *Int. J. Comput. Vis.*, vol. 43, no. 3, pp. 167-188, Jul., 2001.

[50]   S. Eickeler, S. Muller, and G. Rigoll, "Recognition of JPEG compressed face images based on statistical methods," *Image and Vision Computing*, vol. 18, no. 4, pp. 279–287, Mar. 2000.

[51]   A. Amine, S. Ghouzali, M. Rziza, and D. Aboutajdine, "Investigation of feature dimension reduction based DCT/SVM for face recognition", in *Proc. IEEE Symp. on Comput. and Commun.*, Marrakech, 2008, pp. 188-193.

[52]   K. Yasuda and M. O. Ahmad, "Development of an Active Shape Model using the Discrete Cosine Transform," to be submitted for publication to *IET Computer Vision.*

[53]   G. K. Wallace, "The JPEG still picture compression standard," *Commun. ACM*, pp. 30-44, Apr., 1991.

[54]   S. Milborrow, J. Morkel, and F. Nicolls, "The MUCT Landmarked Face Database," in *Proc. Pattern Recognition Assoc. of South Africa*, Stellenbosch, South Africa, 2010, pp. 179-184.

[55]   M. Vetterli, H. Nussbaumer, 'Simple FFT and DCT Algorithms with Reduced Number of Operations," *Signal Processing*, vol. 6, no. 4, pp. 267-278, Aug., 1984.

[56]   O. Jesorsky, K. Kirchberg, R. Frischholz, "Robust Face Detection Using the Hausdorff Distance," in *Proc. 3rd Int. Conf. on Audio- and Video-based Biometric Person Authentication*, Halmstad, 2001, pp. 90-95.

[57]   M. B. Stegmann, B. K. Ersbøll, and R. Larsen, "FAME – a flexible appearance modelling environment," *IEEE Trans. Med. Imag.*, vol. 22, no. 10, pp. 1319-1331, Oct., 2003.

[58]   D. Cristinacce and T. Cootes, "Feature Detection and Tracking with Constrained Local Models," in *Proc. British Mach. Vision Conf.*, Edinburgh, 2006, pp. 95.1-95.10.

[59]   S. Milborrow. (2010, Sept.). Stasm User Manual. [Online]. Available: http://www.milbo.org/stasm-files/stasm-manual-13.pdf