

## **Executive Summary**

1. Produce energy fields via morphological operators.
2. Adapted and developed eye localisation from Choi et al. (2009) seen in section 3.2.
3. Implemented and developed deformable template based on work of Xie et al. (1994) for eye detection, see section 3.3.
4. Developed reflex localisation.
5. Developed deviation calculation.
6. Visually assessed all the images for result analysis.

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Background and Context</b>	<b>3</b>
2.1	Strabismus and Its Diagnosis . . . . .	3
2.1.1	Strabismus . . . . .	3
2.1.2	Hirschberg Test . . . . .	5
2.1.3	Quantification of Deviation . . . . .	8
2.1.4	Hirschberg Ratio . . . . .	8
2.2	Morphological Operation . . . . .	10
2.2.1	Erosion and Dilation for Binary Images . . . . .	10
2.2.2	Top-hat and Top-bottom Transformation . . . . .	10
2.2.3	Blur-minimum Operator on Greylevel Images . . . . .	12
2.3	Eye Extraction . . . . .	13
2.3.1	Coarse Position Estimation . . . . .	13
2.3.2	Template Based Extraction . . . . .	17
<b>3</b>	<b>Work Carried Out</b>	<b>24</b>
3.1	Preprocessing . . . . .	24
3.1.1	Energy Fields Generation . . . . .	24
3.1.2	Tilt Correction . . . . .	27
3.2	Coarse Position Estimation . . . . .	28
3.2.1	Vertical Estimation . . . . .	29
3.2.2	Horizontal Estimation . . . . .	31
3.3	Eye Extraction . . . . .	35
3.3.1	Modification of Deformable Template . . . . .	35
3.3.2	Implementation of Template-based Extraction . . . . .	36
3.4	Reflex Extraction . . . . .	40
3.5	Diagnosis . . . . .	42
<b>4</b>	<b>Result and Analysis</b>	<b>44</b>
<b>5</b>	<b>Conclusion and Further Work</b>	<b>49</b>
5.1	Conclusion and Critical Evaluation . . . . .	49
5.2	Recommendations for Further Work . . . . .	50

<b>A Full Experimental Data</b>	<b>55</b>
A.1 Esotropia Image Set . . . . .	55
A.2 Exotropia Image Set . . . . .	56
A.3 Pseudostrabismus Image Set . . . . .	56
A.4 Other Image Set . . . . .	56

# Chapter 1

## Introduction

Strabismus, often called “crossed-eyes” or “wall eyes”, is a medical condition where the eyes are not properly aligned with each other (Rutstein et al., 2011). Either the left or the right eye has a deviation when the eyes are fixed to an object straight ahead. Sometimes both eyes turn in or out. With an alternative name “squint”, strabismus is a common eye disease. Especially in both infants and the elderly, there are higher strabismic presence (Bremer et al., 1998; Martinez et al., 2011). Generally, this disease can be classified according to the direction of the iris deviation: esotropia is the condition of eyes turning in, while exotropia represents the eyes turning out. There are also hypertropia and hypotropia which correspond to the eyes abnormally turning up or down, however, they are less incident and are more likely to be complicated by a horizontal deviation.



Figure 1.1: Strabismus: left esotropia. Source: “Accommodative Esotropia”, [http://www.strabismus.org/esotropia\\_eye\\_turns\\_in.html](http://www.strabismus.org/esotropia_eye_turns_in.html), 12 September 2012.

Most strabismus can be treated by corrective glasses or surgery, in which the amount of correction is important to know before the treatment. Therefore, quantification of the deviation of strabismic eye is needed to give a further reference for treatment. There is a kind of professional equipment, called synoptophore, that can accurately measure the deviation. This examination can only be carried out in clinics. A patient who undergoes this test

should place their head onto the machine, and fully follow the instructions of practitioner. This complex process makes synoptophore test unsuitable for diagnosis of infants and young children. They are neither not able to sit properly in front of the machine, nor likely to calm down and follow the complex instructions. Therefore, visual assessments are frequently taken. Ophthalmologists often carry out a simple Hirschberg test to subjectively estimate the iris deviation. To hold a light in front of the patient can make a reflex on each eye, and the aim of this test is to estimate the iris deviation from where the reflex locates on the cornea. The result of this test can be referred to in further tests, such as ocular motility test, which assess how well the extraocular muscles can drive the eye fixating moving objects. In general, Hirschberg test produce rather qualitative than quantitative results. Depending on the cooperation of the patient and the experience of the practitioner, the results will necessarily vary between practitioners.

Ophthalmologists have the needed resources to quickly assess eye deviation accurately. The aim of this project is to develop a system that automatically accesses the condition of strabismus through the images of strabismic patient. Specifically, the system should be able to calculate the amount of iris deviation in a given image. Since Hirschberg test is the most common strabismus test in clinics, Hirschberg image (the facial image captured in the Hirschberg test) is relatively easy to get. In these images, blur rarely occurs because the light can always keep the shutter fast. It makes Hirschberg image ideal for the resource for strabismus analysis. The system should automatically detect the eyes in the Hirschberg image and discover the problem within the region of eyes.

In the field of computer vision, to accurately detect eyes in facial image has been developed for more than 20 years, and currently there are two approaches: feature-based and template-based. A disadvantage of the feature-based extraction is that information within the feature cannot be resolved. This makes it unsuitable for this project. However, template-based approach can do this job. Deformable eye template, a representative of template-based approach, was first proposed by Yuille et al. (1992). It can accurately locate the iris within the eye. It is not difficult to implement on Hirschberg image by calculating the deviation through the comparison of the iris and the reflex location. However, the challenge of this approach is to extract the eye with different head poses, and moreover, this system also has to be robust to low image resolution and different skin colours.

# Chapter 2

## Background and Context

### 2.1 Strabismus and Its Diagnosis

#### 2.1.1 Strabismus

Strabismus is a common eye disease. According to Mirzaie (2009), among Caucasians the incidence of strabismus is 3-4%, while its incidence among blacks is much lower, but it is still as much as 0.6%. Strabismus is more prevalent among infants; its incidence rises with the growth of age. A research on first year infants showed the presence of strabismus is 6.6% at 3 months, and 11.8% at 12 months (Bremer et al., 1998), and the presence of strabismus at a younger age was found to be a highly significant predictor of the presence at later age. Some incidence of strabismus are found among the aged and according to Martinez et al. (2011), there is an annual incidence of 0.05% of the residents in Olmsted County, Minnesota, U.S., and the median age at diagnosis was 65.2 years.

#### Extraocular Muscles

Each eyeball is directly controlled by 6 extraocular muscles (EOM). Each extraocular muscle mainly takes charge of movement in one particular direction. As shown in table 2.1, EOMs are controlled respectively by 3 nerves. Among them, 4 muscles are controlled by the oculomotor nerve. This proves that a nerve may control eye orientation in multiple axes, and the instruction to the muscles is cooperated by brain and the nerves. Strabismus may be caused by disability of either the nerve or the muscle. For this reason, it is hard to judge where the problem is by simply visual test. It is important to carry out multiple examinations, i.e. ocular motility test, to exclude the effect of well-working parts.

Extraocular muscle	Primary Action	Nerve
Lateral rectus (LR)	Abduction	Abducens nerve
Medial rectus (MR)	Adduction	
Inferior rectus (IR)	Elevation	
Superior rectus (SR)	Depression	Oculomotor nerve
Inferior oblique (IO)	Extorsion	
Superior oblique (SO)	Intorsion	Trochlear nerve

Table 2.1: Extraocular muscles function and the nerves controlling them:  
 Abduction - movement towards center; adduction - movement away from center; elevation - movement in a superior direction; depression - movement in a inferior direction; extorsion - rotation away from center; intorsion - rotation towards center.

### Classification by Direction

By the direction of deviation, strabismus can be categorised into esotropia (eye horizontally turns towards the centre of head), exotropia (horizontally away from the centre), hypertropia (vertically up), and hypotropia (vertically down). They are usually described with the left or right eye specified, e.g. left esotropia, or right exotropia. Generally, instances of esotropia are recorded to be more than that of exotropia. According to Mirzaie (2009), the presence of esotropia and exotropia in Europe is about 5:1, however, in Japan exotropia is more prevalent.

### Concomitant Strabismus

By the property of deviation, strabismus can be classified as concomitant or incomitant. Concomitant strabismus is to specify the eye, which always has the same deviation angle whatever or wherever it looks. For incomitant strabismus, the deviation angle varies in different orientations. Incomitant strabismus is also called paralytic strabismus, for the reason that it is more likely to be caused by nerve palsy, or sometimes extraocular muscle problems. In terms of concomitant strabismus, the cause of the disease is more of vision functions than extraocular muscle or nerve problems. Children with concomitant strabismus are always accompanied with problematic eyesight and poor binocular vision.

### Accommodation

Compared to other strabismus, concomitant esotropia has a higher incidence in infants and young children. Particularly, if esotropia onset between birth

and six month of age, and the misalignment of eye is stable and large, it is called congenital esotropia, or infantile esotropia. Moreover, by its relevance to accommodation problems, other concomitant esotropia can be classified into accommodative esotropia, partially accommodative esotropia, or non-accommodative esotropia. Accommodation is the ability to focus the eyes. Hyperopia (farsightedness), which is a defect of this ability, is a common cause of accommodative esotropia. According to Crick and Khaw (2003), if a concomitant esotropia can be fully corrected with spectacles, it is accommodative esotropia. Non-accommodative esotropia will get no improvements with spectacles attached, while partially accommodative esotropia is a mixed condition.

### Cover Test

In clinics, the initial deviation of the eye is measured in the first eye position (looking straight forward). If the displacement of eye is large enough, it is can be easily discovered by the practitioner. However, phoria, a recessive strabismus, is the condition where the eyes of the patients usually have no symptom of misalignment, if they are looking at an object with both eyes open. The ability of stereovision helps to automatically correct the eye direction. To diagnose phoria, cover-uncover test is carried out to disrupt the ability of stereo fusion. It is also carried out in diagnosis of dominant strabismus. In both the action of covering and uncovering an eye, to observe whether the other eye moves can help to discover the problem. There is also another cover test, called alternate cover test, to detect heterophoria.

Moreover, in clinical practice, very young children may not respond or not able to follow the instructions of the practitioner. This has made ophthalmologists to invent alternative options to assess first position deviation. According to Rutstein et al. (2011), optometrist can estimate the degree of strabismus by the corneal reflex test, when very young children cannot fixate on a target long enough for valid cover testing.

#### 2.1.2 Hirschberg Test

The complete name of this test is Hirschberg corneal reflex test. German ophthalmologist Julius Hirschberg proposed it in 1886. It estimates the degree of strabismus by the observing the displacement of the central point of gaze from the iris centre. As show in figure 2.1, the patient in this test is asked to keep looking straight forward. At the same time, the practitioner holds a point light in front of the patient. According to Parks et al. (2006), the light source described by Hirschberg was a candle flame, and currently a bright fixation light is used in clinics. The light will produce a glossy reflex

on the corneal. The reflex is marked as the central point of gaze. A correct eye should have the reflex located in the centre of the iris (but not perfect centre because of angle kappa, which will be explained later). Usually, this test is to produce qualitative results, which is by observing how far the reflex is from the edge of the eye. The closer they are, the larger deviation that exists.



Figure 2.1: Hirschberg test. Source: “Figure 3: Pam McCart (right) performs the Hirschberg or corneal light reflex test for ocular alignment on fellow student Martha Parker (left).”, <http://nas.sagepub.com/content/24/6/254.full>, 13 September 2012

Excluding accommodative strabismus (this can be diagnosed by means of seculars), the patient normally has a squint eye, and another one is fit. Even if the light source is not at the front centre of the child, the reflex of the other eye can still be a reference of a healthy condition. However, a well-working eye is not actually having the centre of gaze located right on the centre of the iris. The actual human eye has an angle between the line of sight and the pupillary axis, which is defined as angle Kappa (LeGrand and ElHage, 1980). It is also called angle Lambda in some literature. According to Parks et al. (2006), a nasal (direction towards nose) angle kappa is designated as positive, while a temporal (direction towards temple) angle as negative. It is illustrated in figure 2.2. Eyes of humans have this angle, although most of them are quite small, which is as a result of the fusion of stereovision. This is not a sign of strabismus. Since Hirschberg test estimates the angle between visual axis and the line of the point light, what it measures is just angle kappa for a healthy person. Angle Kappa always cannot be eliminated during the test. The acceptable deviation should not exceed 11PD (5mm) in nasal direction, when the angle Kappa is taken into account. Then this explains why most people do not have the Hirschberg reflex on the perfect centre of both eyes.

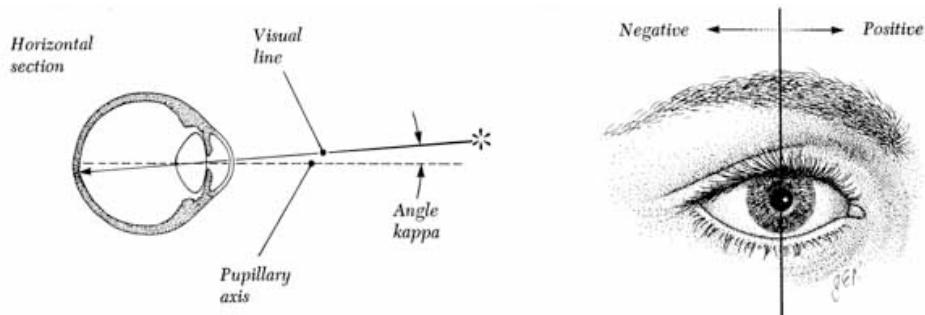


Figure 2.2: Angle Kappa and the designation of its value. (Parks et al., 2006).

Although it is a simple test, Hirschberg has a valuable advantage: the ability to exclude pseudostrabismus. Infants have a high incidence of strabismus in their first year, but some infants have both correct eyes although they look strabismic (figure 2.3). This condition is called pseudostrabismus. Due to their facial features, especially the bridge of nose, are not fully developed, one of the eye may look deviated. Simply observing the relative position of eyes in the eyelid will lead to a false judgment. Under Hirschberg test, pseudostrabismus can be easily distinguished from strabismus - both of the reflexes on the eyes have no obvious displacement away from the pupil.



Figure 2.3: Pseudostrabismus: although the right eye looks esotropia, the reflex of both eyes are on the centre. Source: “Fig. 2 Pseudostrabismus is the appearance of but not truly misaligned eyes.”, Esotropia, <http://www.aapos.org/terms/conditions/48>, 2010).

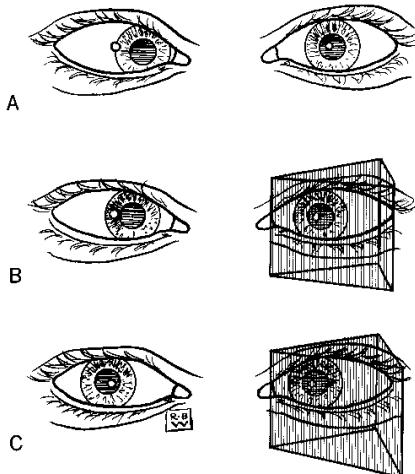


Figure 2.4: Krimsky test. Source: “Krimsky prism test”, <http://www.umsl.edu/garziar/krimsky.htm>, 15 September 2012.

### 2.1.3 Quantification of Deviation

To quantify the iris deviation, Krimsky test is invented as an improvement of Hirschberg test. It adds a prism before an eye (currently often the fixating eye) for turning the sight to a particular degree, which is shown in figure 2.4. Since the visual axis of the eye has been amended, the eye will turn, and the reflex on the iris will also move. This change is shown in figure 2.4. The angle the eye turns is proportional to the diopter of the prism. (Prism diopter was defined as being the deflection of one centimetre at a distance of one metre.) By trying different diopters, the best is one which makes the light to reflex symmetrical between two eyes. The diopter of prism (PD) is often used to describe the deviation of strabismus.

PD is widely used in all strabismus diagnosis, even when there is no prism in Hirschberg test. It is possible to roughly estimate the deviation by simple visual assessment. Schwartz (2006) summarised a way to estimate the prism diopter corresponding to a strabismic eye. (See figure 2.5.)

### 2.1.4 Hirschberg Ratio

The estimation by visual assessment is not accurate enough for further assessment. Fortunately, more detailed data can be extracted from the Hirschberg reflex. In order to make an accurate transformation from reflected location to prism diopter, Hirschberg ratio is defined as the corre-

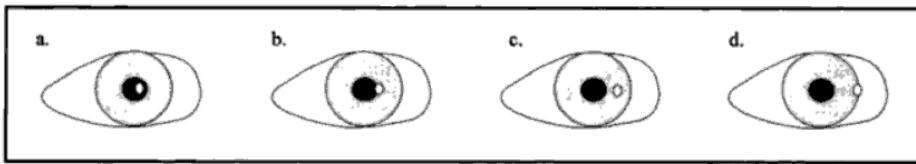


Figure 2.5: Estimating the deviation by the Hirschberg test (Schwartz, 2006): a, no deviation (ortho) because the reflex is at the centre of the pupil; b, a deviation of 30 PD, when the reflex is at the pupillary margin; c, a deviation of 60 PD, when the reflex falls halfway between the pupillary margin and the limbus; d, a deviation of 90 PD, the reflex is on the limbus.

sponding prism diopter added when the reflex moves by 1mm. The unit is PD/mm. In clinical assessment, this ratio is not very useful, because the practitioners rarely measure the deviation by ruler when calculating PD. They can use professional equipment to get far more accurate data. Nevertheless, Hirschberg ratio is important in research. With an accurate distribution of Hirschberg ratio, it is able to get PD value with wherever corneal reflex.

Eskridge et al. (1988) carried out several tests on subjects of all ages to find a reasonable Hirschberg ratio. He marked the reflex as an initial position when the eye is fixating, and then performed the Hirschberg test. To calculate the final deviation, he used the relative distance between the deviated position and the initial position. This method successfully eliminates the disturbance of angle Kappa, and gives a good result. The Hirschberg ratio he got is 22 PD/mm for all ages.

Riddell et al. (1994) proved a proper Hirschberg ratio should be 22 PD/mm for both adults and infants. But he discovered the angle Kappa (called angle lambda in his article) decreases from 8.4 at birth to 5 at 5 months, due to reflect axis growth of the eye.

However, the surface of the cornea is actually not a spherical surface. Brodie (1992) pointed out that if the Hirschberg ratio is constant, it will produce an obvious error when the angle is larger than 50 PD. He proposed an integral equation, as well as a differential approach to represent the surface of the cornea. Regarding the accuracy of this proposed system, the integral equation is not needed. When the eye is in a extreme position, the eye extraction result would also be unreliable. It is sure that when assessing the deviation, to use a constant Hirschberg ratio of 22 PD/mm is reasonable where the deviating angle is not large.

## 2.2 Morphological Operation

In the proposed method, the morphological operator will help to generate pre-processed image and a part of deformable template. All the needed morphological operations will be introduced in this section.

### 2.2.1 Erosion and Dilation for Binary Images

According to Maragos (2004), given a binary image, let the foreground(pixels with value 1) be represented by the set  $X$  and the background by the set compliment  $X^c$ . Dilation of  $X$  by a (window) set  $B$  is equivalent to the Minkowski set addition  $\oplus$ :

$$X \oplus B \triangleq \{z : (B^S)_{+z} \cap X \neq \emptyset\} \quad (2.1)$$

where  $(B^S) \triangleq \{x : -x \in B\}$  is the symmetric of  $B$  with respect to the origin. Likewise, erosion of  $X$  by  $B$  is equivalent to the Minkowski set subtraction  $\ominus$ :

$$X \ominus B \triangleq \{z : B_{+z} \subseteq X\} \quad (2.2)$$

There is a duality between dilation and erosion since  $X \oplus B = (X^c \ominus B^s)$ . In applications,  $B$  is usually called a structuring element. It normally has a simple geographical shape and a size smaller than image  $X$ . Dilation will shrink the input image foreground, and erosion will expand the input image foreground. For instance, if  $B$  is a flat disk with a radius of 2 pixels, dilation will add a boarder of 1-pixel width, and erosion will remove the boarder of that width.

### 2.2.2 Top-hat and Top-bottom Transformation

The peak and valley energy fields for deformable template is generated by top-hat and top bottom transformation. They are morphological operators who work on graylevel images. At first, some basic operations are introduced. Denote a grayscale image as  $f(x)$ . The dilation and erosion of image signal  $f(x)$  by a set  $B$  (Maragos, 2004):

$$(f \oplus B)(x) \triangleq \bigvee_{y \in B} f(x - y) \quad (2.3)$$

$$(f \ominus B)(x) \triangleq \bigwedge_{y \in B} f(x + y) \quad (2.4)$$

where  $\bigvee$  denotes maximum for finite  $B$  and  $\bigwedge$  denotes minimum for finite  $B$ . Flat erosion of a function  $f$  by a small convex set  $B$  reduces the peaks and enlarge the maxima of the function. In contrast, flat dilation increases the valley and enlarges the minima. The actual effect of both these operations

is shown in figure 2.6.

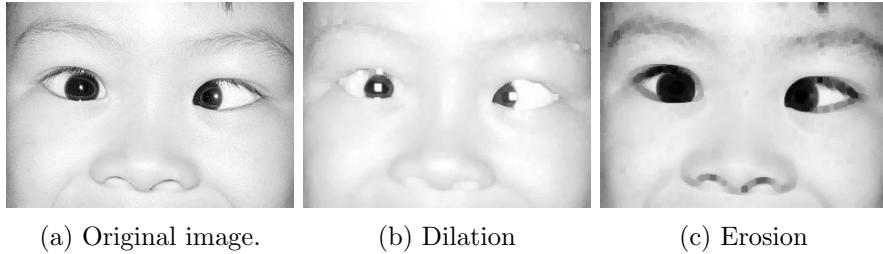


Figure 2.6: Dilation and erosion of greyscale image by a flat disk with  $r = 3$ . (Original image source: “Strabismus in Adults and Children (Part 2 - Symptoms)”, <http://info.thevisiontherapycenter.com/discovering-vision-therapy/bid/56472/Strabismus-in-Adults-and-Children-Part-2-Symptoms>, 17 September 2012.)

Based on erosion and dilation, opening and closing operation is defined as:

$$f \circ B = (f \ominus B) \oplus B \quad (2.5)$$

$$f \bullet B = (f \oplus B) \ominus B \quad (2.6)$$

Maragos (2004) summarised that “flat opening smooths the graph of  $f$  from below by cutting down its peaks, whereas the closing smooths it from above by filling up its valley”. The result of filtering the image of figure 2.6a is shown in figure 2.7.

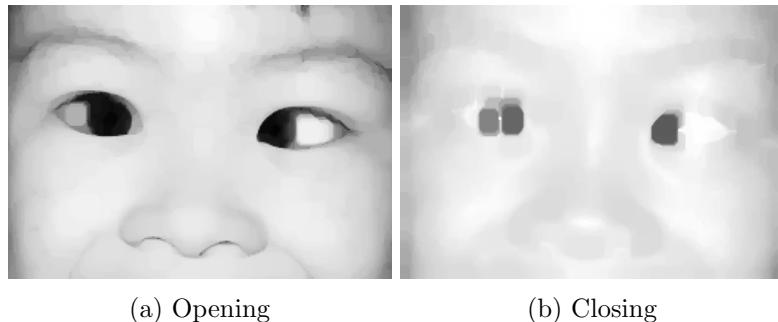


Figure 2.7: Opening and closing of greyscale image. We can see that in opening, sharp peaks have all been removed, while in closing, dark shadows have been levelled up.

Then top-hat transformation is given by:

$$\text{peak}(f) \triangleq f - (f \circ B) \quad (2.7)$$

It is also called peak operator. Obviously the set  $B$  is a very important parameter of peak operator, because it controls the shape and size of the leks' s support. The transformation of top-hat as completed is illustrated in figure 2.8.

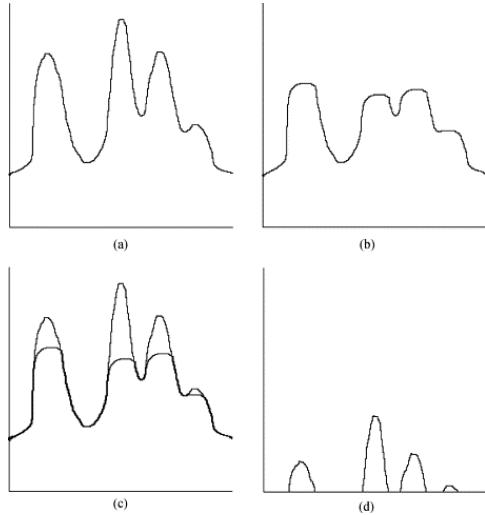


Figure 2.8: Tophat transformation through greylevel opening: (a) original function, (b) function opened with circular disk, (c) superposition of the previous two and (d) features after tophat transform (Mukhopadhyay and Chanda, 2000).

Similarly, here is the top-bottom transformation:

$$\text{valley}(f) \triangleq (f \bullet B) - f \quad (2.8)$$

These operators are very useful for detecting a region with significantly brighter (or darker) intensities relative to the surroundings (Maragos, 2004). Therefore, they are just right for highlighting local peaks and shadows. The result is shown in figure 2.9.

### 2.2.3 Blur-minimum Operator on Greylevel Images

Lee et al. (1987) proposed this edge detector, which is defined by

$$I_{\text{edge-strength}} = \min\{I_1 - \text{erosion}(I_1), \text{dilation}(I_1) - I_1\} \quad (2.9)$$

where  $I_1 = \text{blur}\{I_{\text{input}}\}$  and  $\text{blur}\{I_{\text{input}}\}$  is the input image with a blurring operation. The same size of kernels are used for the blur and the structuring element of the dilation and erosion. According to Lee et al. (1987), the advantage of this operator is that it will not detect a single noise point. Moreover, it can detect ideal step edges. Therefore in facial feature extraction, it can have a good response to the edge between the iris and the



(a) Top-hat

(b) Top-bottom

Figure 2.9: Peak and valley extraction of greyscale image. We can see that only the local peaks and valleys are extracted in detail.

eye whites, as well as robusting to spots on the face to some extent. Its performance is shown in figure 2.10.



(a)  $\min\{G_e, G_d\}$ .

(b)  $\max\{G_e, G_d\}$

(c) Blur-minimum.

Figure 2.10: Comparison of morphological edge detectors. The original image  $f$  is same as figure 2.6a.  $G_d$  is the dilation residue operator defined by  $\text{dilation}(f) - f$ , and erosion residue operator  $G_e$  is defined by  $f - \text{erosion}(f)$ . All the structure element  $B$  used here is a 3x3 flat disk. We can see the former ones are sensitive to noise and the edge extracted are highly determined by size of  $B$ . Only Blur-minimum operator have a better response to step edges in eyes and can result in continuous fine edges.

### 2.3 Eye Extraction

### 2.3.1 Coarse Position Estimation

The deformable template should be initialised within an estimated eye region (eye window). Therefore to find where the eyes are in the image as accurate as possible all the work should be done before fine extraction of the iris.

## Early Work

Kanade (1977) did a pioneering work on facial feature detection. He filtered facial image with a 9-by-9 Laplacian operator, and got a binary edge image. Then there are 7 steps for extracting the following features in order: head top; face sides; nose, mouth and chin; chin contour; cheer areas; eye. Since the analysis of each step is based on results of previous steps, if in any step a salient feature is not found, the whole program will stop and discard the current image.

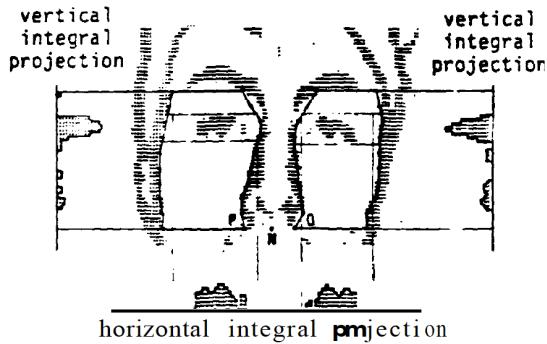


Figure 2.11: Cheek areas and rectangles which contain the eye (Kanade, 1977)

Bounding eyes with rectangles is the final step in this algorithm. Particularly, it works upon the estimated area of cheek, which is the result of the former steps. In this step, the projection of the edge binary image will be done both horizontally and vertically. After fusion (dilation) and shrink (erosion), the largest connected area in the estimated region can then be picked out. The bounding rectangle of the two selected areas are the result of eye detection. However, the drawback is that any error of the previous steps can result in the failure of the eye extraction.

## A Fast Method

Choi et al. (2009) proposed a set of method for fast coarse estimation of the eye location. Their publication illustrates the entire process for extracting eyes, but only the former steps are explained here. The fine eye extraction of them is rough and will not be used in this project. The estimating method is quite similar to the work done by Kanade (1977), but in a more straightforward way. It consists of 3 main epochs.

**Estimation of Face Tile Angle** In a facial image, the eyes, the eyebrows, and the mouth regions are darker than their neighbourhood and these shapes are often horizontally wide and vertically short. They can be easily extracted on the valley image (figure 2.12 (a)) produced by top-bottom transformation (see subsection 2.2.2). Once the valley region is extracted, a function  $V(y)$  can be defined:

$$V(y) = \text{the number of pixels}(x, y) \in \text{valley}, \text{for all } x. \quad (2.10)$$

The function  $V(y)$  is shown in figure 2.12 (c), in which the vertical y-axis represents the vertical image line and the horizontal x-axis represents the number of pixels belonging to the valley regions in each line. According to Choi et al. (2009), this function has the maximum variance when the face is kept upright, regardless of illumination change (shown in 2.13). The face tilt-angle can be reliably estimated by rotating the face image to the pose when  $V(y)$  has the maximum variance. The result is shown in figure 2.12 (c) and (d).

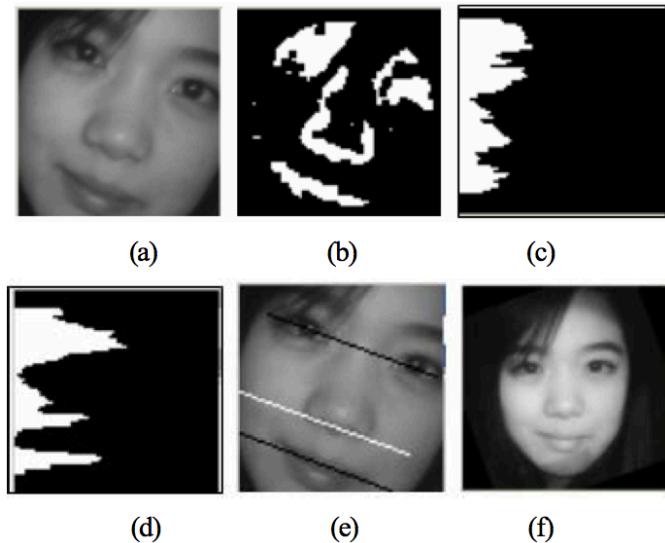


Figure 2.12: The estimation of face tilt angle: (a) The original image, (b) The valley detected using a morphological grey black top-head, (c) The function  $V(y)$  when the face tilted to the right, (d) The function  $V(y)$  when the face de-rotated to upright. (e) Three lines show the estimated tilt angle, (f) the normalised face image rotated to upright. (Choi et al., 2009)

**Coarse Estimation of Vertical Eye Location** According to Choi et al. (2009), the function  $V(y)$  in the equation has local maximum values at the vertical locations of the eyes, eyebrows, nose, or mouth. This is shown in figure 2.13. In general, the function  $V(y)$  has the maximum value at



Figure 2.13: Variation of the function  $V(y)$  due to the illumination and pose changes (Choi et al., 2009)

the vertical eye location or at the vertical eyebrow location. Since there is a property that eyes are closer to eyebrow than to the nose, the vertical location of eyes can be estimated. When detecting the local maximum location, they use a lowpass filter to smooth function  $V(y)$  for removing spurious local maximum. Then two local maximum located at the eye and the eyebrow can be merged, especially when the eyebrows are very thick and the eyes are very thin. Choi et al. (2009) created another function  $E(y)$  using horizontal edges:

$$E(y) = \text{the number of pixels}(x, y) \in \text{horizontal edges, for all } x. \quad (2.11)$$

Even though this function  $E(y)$  is more sensitive to illumination change, according to Choi et al. (2009) the eye location can be reliably separated from the eyebrow location.

**Coarse Estimation of Horizontal Eye Location** Once the vertical location of the eyes is obtained, they cut off a region which includes two eyes and two eyebrows tightly. Based on the valley image in this region (shown in figure 2.14 (b)), Choi et al. (2009) create a function  $F(x)$ :

$$F(x) = \text{the number of pixels}(x, y) \in \text{valley, for all } y. \quad (2.12)$$

According to Choi et al. (2009), the function  $F(x)$  is large in the eye region and small in the other region. Using this property, the horizontal zones of the eye can be located.

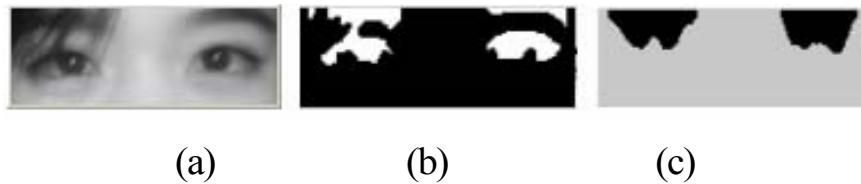


Figure 2.14: (a) The cut-off image including eyes and eyebrows, (b) The detected valley region, (c) The Function  $F(x)$  (Choi et al., 2009)

### 2.3.2 Template Based Extraction

Within the estimated eye region, the next process is to extract the iris. Yuille et al. (1992) proposed the deformable eye template, which can be attached to the shape of eye and accurately gets the iris location.

#### The Deformable Template

Yuille et al. (1992) first proposed a set of deformable templates for facial features in 1988, and then revised their publication in 1992. The later one is illustrated here. This method has been the classical way of doing facial feature extraction and recognition. Currently, many researches are still taking his template, or the variation of it as a fundamental, although this algorithm was proposed over 20 years ago.

Before extraction, it is essential to generate three potential fields from the facial image. The edge image  $\Psi_e(x, y)$ , valley image  $\Psi_v(x, y)$ , and peak image  $\Psi_p(x, y)$  can be produced by morphological operators (see section 2.2). But the edge field, for example, will be the largest near edges in the image. Therefore they smooth these three images by convolving them with an exponential function  $e^{-p\sqrt{x^2+y^2}}$  respectively. This operation results in energy fields  $\Phi_e(x, y)$ ,  $\Phi_v(x, y)$ , and  $\Phi_p(x, y)$ , which enable interactions to be effective over longer distances. The three potential fields, together with the image intensity  $\Phi_i(x, y)$ , are shown in figure 2.15. The potential fields (representations of image) are more effective for tracking objects than the image itself. According to Yuille et al. (1992), introducing potential fields will enable strong edges, valleys, or peaks to attract objects a large distance away.

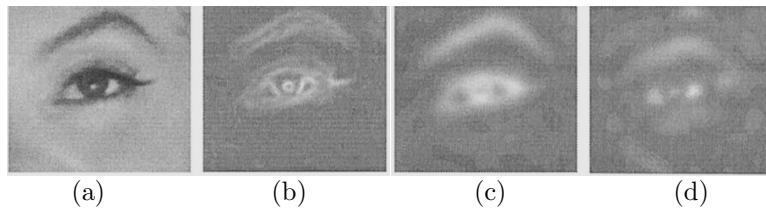


Figure 2.15: The potential fields for an “archetypal” eye. The eye (a), has edge, valley, and peak fields shown in order (b), (c), (d). The strengths of the fields are shown in greytone, white is strong and black is weak. (Yuille et al., 1992)

Then the original eye templates is illustrated in figure 2.16. According to Yuille et al. (1992), it consists of the following features:

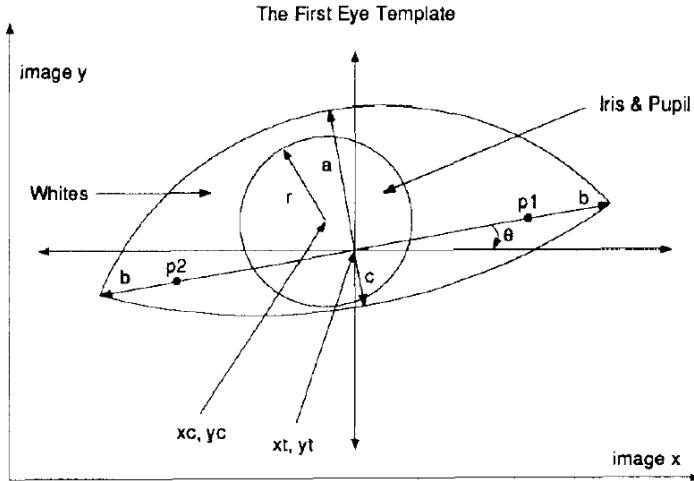


Figure 2.16: The deformable eye template (Yuille et al., 1992)

1. A circle of radius  $r$  and centre  $\vec{X}_c$ , representing the border of iris. Its interior is attracted to valleys in the image intensity.
2. A bounding contour of the eye, modelled by two parabolic sections. Both of the sections have centre  $\vec{X}_e$ , and width  $2b$ , and the height above and below is  $a$  correspondingly. The bounding contour has a rotation  $\theta$ . It is attracted to edges.
3. Two points  $\vec{X}_e + p_1(\cos\theta, \sin\theta)$  and  $\vec{X}_e + p_2(\cos\theta, \sin\theta)$  ( $p_1 \geq 0, p_2 \leq 0$ ), located at the centres of the white of eyes. They are attracted to peaks in the image intensity.
4. The regions between bounding contour and the circle also correspond to whites of the eyes. They are attracted to large values in the image intensity.

The template has eleven parameters represented by  $\vec{g} = (\vec{X}_c, \vec{X}_e, p_1, p_2, r, a, b, c, \theta)$ . Then the complete energy function  $E_c(\vec{g})$  is defined as

$$E_c = E_v + E_e + E_i + E_p + E_{prior}$$

where

- (i)  $E_v$  is the valley potentials, given by opposite value of average valley intensity inside the circle,

$$E_v = -\frac{c_1}{|R_c|} \int_{R_c} \Phi_v(X) dA;$$

- (ii)  $E_e$  is the edge potentials, given by the opposite sum of average edge intensity on circle and bounding contour,

$$E_e = -\frac{c_2}{\partial|R_c|} \int_{\partial R_c} \Phi_e(X) \, ds - \frac{c_3}{\partial|R_w|} \int_{\partial R_w} \Phi_e(X) \, ds;$$

- (iii)  $E_i$  is the image potentials, given by the average peak intensity inside the circle, minus the average valley intensity in the whites,

$$E_i = \frac{c_4}{|R_c|} \int_{R_c} \Phi_i(X) \, dA - \frac{c_5}{|R_w|} \int_{R_w} \Phi_i(X) \, dA;$$

- (iv)  $E_p$  is the peak potentials, given by the sum of peak intensity on p1 and p2,

$$E_p = c_6 \{ \Phi(\vec{X}_e + p_1(\cos\theta, \sin\theta)) + \Phi(\vec{X}_e + p_2(\cos\theta, \sin\theta)) \};$$

- (v)  $E_{prior}$  is the prior potentials, given by

$$\begin{aligned} E_{prior} = & \frac{k_1}{2} \|(\vec{X}_e) - (\vec{X}_c)\|^2 + \frac{k_2}{2} (p_1 - p_2 - \{r + b\})^2 \\ & + \frac{k_3}{2} (b - 2r)^2 + \frac{k_4}{2} ((b - 2a)^2 + (a - 2c)^2). \end{aligned}$$

Here  $R_c$  and  $R_w$  correspond to the iris and the eye, while  $\partial R_c, \partial R_w$  represents their boundaries (see figure 3.4). Their areas or lengths are given by  $|R_c|, |R_w|, |\partial R_c|, |\partial R_w|$ .

The search is divided into a number of epochs with different values of the parameters  $\{c_i\}$  and  $\{k_i\}$ . According to another publication of Yuille (1991), “changing the values of these coefficients enables us to use a matching strategy in which different parts of the template guide the matching at different stages”. To minimise  $E_c$ , the parameter values are updated by steepest descent.

In the simulation by Yuille et al. (1992), the process is divided into 8 carefully designed epochs:

1. The valley field helps the iris to drag the template toward the eye.
- 2 - 3. The iris finds its size and position.
4. The Peak field helps in finding parabolic locations.
5. Fine tune white boundaries.

## 6 - 8. Fine tune position.

This process is shown in figure 2.17. In each epoch, only some potentials are allowed to change. An epoch will terminate when the updated parameters converge, which is defined by another table of parameter tolerances in each epoch.

The simulation presents a good result. The image they test on is a simple eye photo with good lighting. The initial template will be placed in the centre of the image. As a first work of the deformable template based extraction, their result is acceptable on most test images. However, the epochs of this algorithm are pre-defined, which may be unsuitable for other images. The author also points out that this algorithm gives error during a test, when there is a strong intensity peak on the eyelid and shadow in the eye.

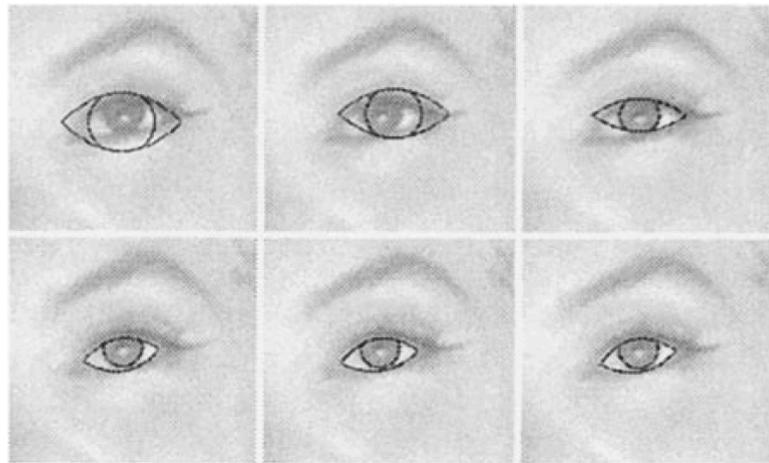


Figure 2.17: Eye template at different times during the minimisation. (Yuille et al., 1992)

## Improved Deformable Template by Xie et al. (1994)

Xie et al. (1994) point out some problem of the template by Yuille et al. (1992):

1. Long processing time.
2. Sequentially weights changing may result in trapping in local minima.
3. Hard to generalise - weights are determined by experiment.

Then they proposed an improved template. They removed the two points in whites (represented by  $p_1, p_2$ ) from the template, and used the middle

point between the centre and corner as the peak of whites. This is shown in figure 2.19.

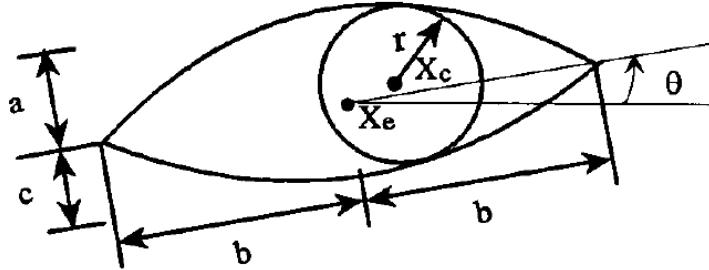


Figure 2.18: The modified eye template. (Xie et al., 1994)

In this template, the modified total energy is denoted as

$$E = \sum_{i=1}^{10} W_i E_i^2$$

Where  $E_i$ , called energy terms, are normalised between 0 and 1.  $\{C_i\}$  in the original template are replaced by weights  $\{W_i\}$ , and they are placed outside each term  $E_i$ . The definition of each energy terms is given as follows:

1. The valley potential here became a valley energy term.

$$E_1 = E_v = 1 - \frac{1}{Area_{circle}} \sum_{i \in circle} b_i^{(V)}$$

where  $b_i^{(V)}$  is the value of the point from valley field.

2. Similarly, the edge potential became edge energy term.

$$E_2 = E_e = 1 - \left( \frac{1}{Length_{circle}} \sum_i b_i^{(E)} + \frac{1}{Length_{parabolas}} \sum_i b_i^{(E)} \right)$$

where  $b_i^{(E)}$  is the value of the point from edge field.

3. With the parameters  $p_1, p_2$  removed, a new term is created to replace the peak potential. It sums the peak field over two small windows located at the middle points of the centre and parabola intersections.

$$E_3 = E_p = \sum_{i=1}^2 \frac{1}{Area_{window_i}} \sum_{j \in window_i} b_j^{(P)}$$

where  $b_i^{(P)}$  is the value of the point from peak field, which is applied to the next term as well.

4. Another new term is calculated in small windows at parabola intersections, to give the template a right orientation.

$$E_4 = E_c = 1 - \sum_{i=1}^2 \frac{1}{\text{Area}_{\text{window}_i}} \sum_{j \in \text{window}_i} b_j^{(V)}$$

The image potential had been spilt into three terms. They are the terms for image intensity inside the circle, over the whites, and on the edges:

$$5. \quad E_5 = E_{ic} = 1 - \frac{k_5}{\text{Area}_{\text{circle}}} \sum_{i \in \text{circle}} g_i$$

$$6. \quad E_6 = E_{iw} = 1 - \frac{k_6}{\text{Area}_{\text{whites}}} \sum_{i \in \text{whites}} g_i$$

$$7. \quad E_7 = E_{ie} = 1 - \frac{k_7}{\text{Length}_{\text{parabolas}}} \sum_{i \in \text{parabolas}} g_i$$

where  $g_i$  is the greylevel value of point in the image.

The prior potential had been simplified and spilt into two internal energy terms  $E_8$  and  $E_9$ . Another one  $E_{10}$  is newly added to constraint of the shape of lower eyelid.

$$8. \quad E_8 = E_{int\ 1} = k_8 |\vec{X}_e - \vec{X}_c|$$

$$9. \quad E_9 = E_{int\ 2} = k_9 |b - 2r|$$

$$10. \quad E_{10} = E_{int\ 3} = k_{10} |b - 4c|$$

In the above definition,  $\{k_i\}$  are normalisation constants. According to Xie et al. (1994),  $k_5, k_6, k_7$  are obtained from the grey level mean and standard deviation of the area under side ration and  $k_8, k_9, k_{10}$  are related to  $b$ .

Compared to Yuille et al.'s work, the greatest change of this method is on the way of processing. Firstly, in the part of preprocessing, Xie et al. (1994) used the binary image of potential fields. A bounding box is created for the deep valley of the image, and the template will be initialised within the box. They worked on the whole facial image and the eye window is obtained by searching both vertically and horizontally on the face. The algorithm was based on the prior knowledge of human faces, but was not explained in detail.

Then, most importantly, they remove the epochs of minimisation, but optimised all the parameters  $X = \{\vec{X}_c, \vec{X}_e, a, b, c, r, \theta\}$  simultaneously. According to Xie et al. (1994), this "prevents some parameters of the eye template from being overly changed, helps the algorithm to converge to



Figure 2.19: The valet, peak and edge fields of an eye image (from left to right). (Xie et al., 1994)

the global minimum, and reduces processing time". The normalised energy function is written as a sum of weighted nonlinear functions

$$F(X) = \sum_{i=1}^{10} W_i E_i^2(X). \quad (2.13)$$

They used the Levenberg-Marquardt algorithm rather than steepest descent method for energy minimisation. The iteration is given by

$$X_{k+1} = X_k + d_k \quad (2.14)$$

$$(J_k^T W J_k + \lambda I) d_k = -J_k^T W F_k \quad (2.15)$$

where  $J_k$  is a Jacobian matrix for the nonlinear functions  $E_i(\vec{X})$ , and  $\lambda$  is an adjustable coefficient. (Xie et al., 1994) In their implementation,  $\lambda$  is chosen as a small value, to balance the convergence speed and the ease of initialised parameter selection. A process of minimisation is shown in figure 2.20.

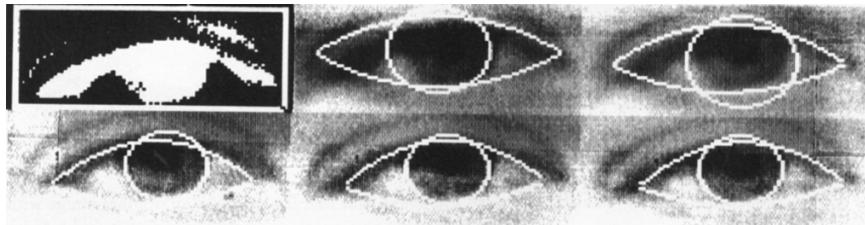


Figure 2.20: A sequence of eye templates with the initial parameters of the template based on the eye window. (Xie et al., 1994)

According to Xie et al. (1994), in their experiment, it takes typically 4 - 10 iterations for the algorithm to converge to the global minimum state. The best extraction comes when the weight for image energies equal 4 and for internal energies equal 0.5. However, they also conclude that the fitting of parabolas is more difficult than that of the circle, for the reason that shades exist around the eyelids and the area around lower eyelid has very low contrast.

# Chapter 3

# Work Carried Out

Based on the material illustrated above, this project will firstly use the fast method (Choi et al., 2009) to estimate the rough position of the eye and then implemented the improved deformable template (Xie et al., 1994) to locate the iris in Hirschberg image. Due to the special features of strabismic eyes and requirements of this project, both of these methods have been modified.

## 3.1 Preprocessing

Preprocessing include energy field generation and tilt correction. Tilt correction and the deformable template optimisation will rely on the generated energy fields.

### 3.1.1 Energy Fields Generation

The aim of this step is to extract three representation of the strabismic image for later analysis. They are: valley field (highlighting the dark iris and eyelid), peak field(highlighting the eye whites), and edge field (highlighting strong step edges, especially the boundary of iris and the eyelids).

#### Peak Field and Valley Field Generation

Creation of these two fields are similar, because they both highlight particular regions in the image. Regarding the property of the iris, the valley field should strengthen the shadows in the image. However, the iris is probably not the darkest region in the whole image (background or black hair may have lower greyscale value), but it should be the darkest among its neighbourhood due to the bright eye whites around it. Therefore the creation of valley field is actually to find local minimum in the image. Similarly, since the highlight on the face may be higher than the eye whites, generation of peak field should be extraction of local maximum. When given a test image as figure 3.1a, simple threshold is not capable to compete this job. It is easy

to see that in figure 3.1b, the extracted shadow area contains the unwanted background, while in figure 3.1c, the bright areas contain the cheek, also the eye whites are not fully extracted.

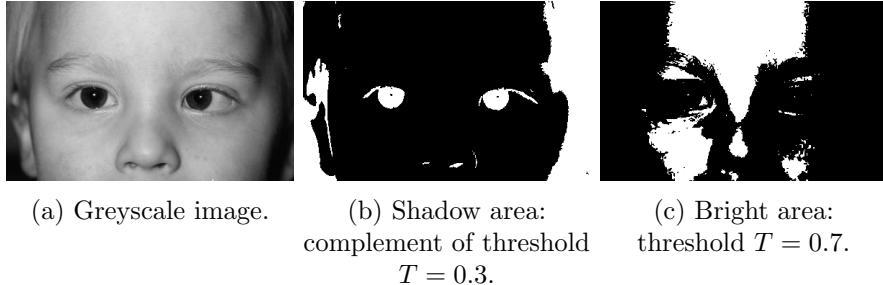


Figure 3.1: Binary thresholding is not suitable for energy field generation. Original image source: “Strabismus Picture - Strabismus is commonly caused by an eye muscle abnormality”, Strabismus, <http://lasereyesurgeons.net/strabismus>, 22 September 2012.

The author also tried the adaptive threshold, which can binarise an image according to local brightness. According to Fisher et al. (2003), this algorithm will first calculate the mean  $M$  of a  $N * N$  square neighbourhood around each pixel, and then compare the value of central pixel with  $M - C$  (where  $C$  is a thresholding constant). If the value is greater than  $M - C$ , it will be set to foreground, otherwise, it is set to background. Compared to direct thresholding, the shadows and highlights found are local rather than global. Applied on the image of figure 3.1a, adaptive threshold can result in images illustrated figure 3.2. In figure 3.2a, it is obvious that more local shadow has been extracted and the dark background can be effectively removed. However, the peak image does not pick out all white areas. The noise on it could also mislead the deformable template during optimisation.

Most importantly, binary image are not suitable for energy fields however they are generated. Although Xie et al. (1994) used binary energy fields in their experiment, the author here prefers the representations created by Yuille et al. (1992). Energy field for template of Yuille et al. (1992) are grayscale images, which guarantee them to convey more information on each pixel than binary images. Deformable template optimisation can be unstable and fluctuating on binary images, since the difference of nearby pixels on the edge are much greater than that in grayscale images.

To create grayscale energy field, morphological top-hat and top-bottom operations (see subsection 2.2.2) are chosen. It is not hard to see their advantage when figure 3.3 is compared with the binary images above. Effects of the background is reduced because both fields have relatively low value



(a) Valley image: complement of adaptive threshold where  $C = 15$  (b) Peak image: adaptive threshold image where  $C = -10$

Figure 3.2: Adaptive threshold images of figure 3.1a.

there. In the peak field, eye whites have highest values while in valley field two irises have great response. This makes the deformable template much more easy to be attracted.



(a) Top-hat peak field (b) Top-bottom valley field

Figure 3.3: Peak and valley potential fields generated with morphological operators.

The size of the operator should be proportional to the size of the feature which is to be extracted. In this step the size of eye is not known, but all image resources are facial image. Assuming all image are covering the eyes with a little bit extra space horizontally, the author allows the kernel size to be proportional to width of the image. The experiments show that when the size is equal to the  $width/30$ , the extracted fields have the best performance.

### Edge Field

The edge field is also extracted through morphology operations. Similar to the method which Choi et al. use, blur-minimum operator (see subsection 2.2.3) is implemented to extract edges. The comparison of sobel edge detector, morphological gradient operator, and blur-minimum operator is shown in figure 3.4: the classical sobel edge is strong but choppy, which could mislead the eye template; the gradient edge performs much better on fine edges but it is sensitive to noise; only the blur-minimum edge has strong and clear response to step edge, and robustness to noise as well.

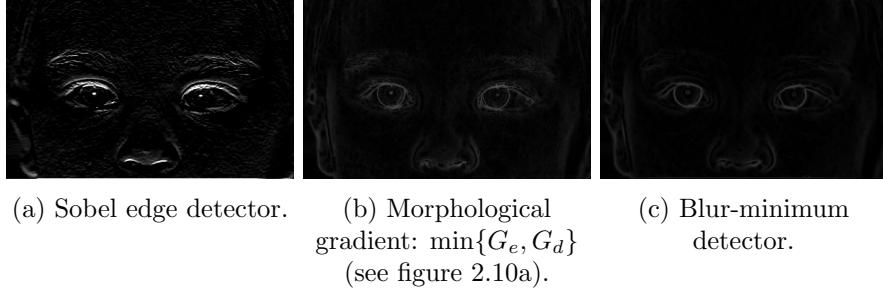


Figure 3.4: Comparision of edge detectors. The source image is the same as figure 3.1a. Kernel size for all of these operations are  $width/100$ .

Similar to the peak and valley fields, the size of the blur-minimum kernel should also be proportional to size of the eyes. Here the author still uses the width of image to estimate the kernel size, but the best value is  $width/100$ , which resulted from the experiments.

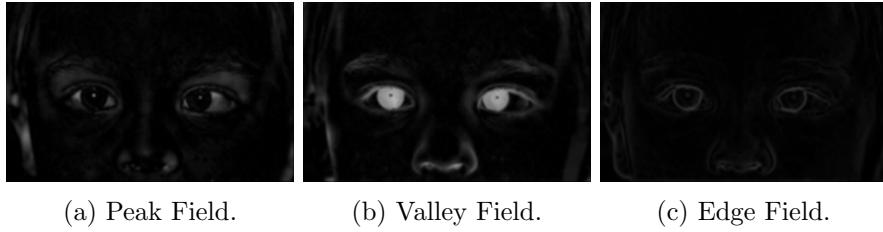


Figure 3.5: Smoothed potential fields.

According to Yuille et al. (1992), smoothing potential fields can help attract the template from long distance. However, if too large kernel is used, the smoothed image can lose the detailed information of step edges. Therefore,  $width/100$  is chosen again for the size of kernel, to produce a smooth but informative energy field. Images in 3.5 are the energy fields after smoothing by mean filter.

### 3.1.2 Tilt Correction

The algorithm of tilt correction is adapted from the work of Choi et al. (2009). When the function  $V(y)$  (definition see equation 2.10 ) is calculated, it is not hard to calculate its variance

$$Var(y) = \frac{1}{H} \sum_{y=1}^H (V(y) - \mu)^2$$

where

$$\mu = \frac{1}{H} \sum_{y=1}^H V(y)$$

and  $H$  is the height of image. Choi et al. (2009) point out that the standard variation of  $V(y)$  will reach its maximum value when the face is upright. Since the square root operation is monotonically increasing on  $[0, +\infty)$ , variance  $Var(y)$  can replace its square root of the standard variation in Choi et al.'s method is used to simplify the calculation. For instance, when given the valley field of a facial image, the algorithm will try to tilt the image clockwise and anticlockwise within a range. (To balance speed and accuracy, during this process the range is chosen as  $\pm 15^\circ$ , and the step for between each trial is  $1^\circ$ .) The function of  $Var(y)$  on tilt degree is shown in figure 3.6c). The degree where the variance reaches its maximum is chosen to be the degree for correction, which will then correct all the three energy fields and the grayscale image. The correction result of the valley image in figure 3.6a is shown in figure 3.6b.

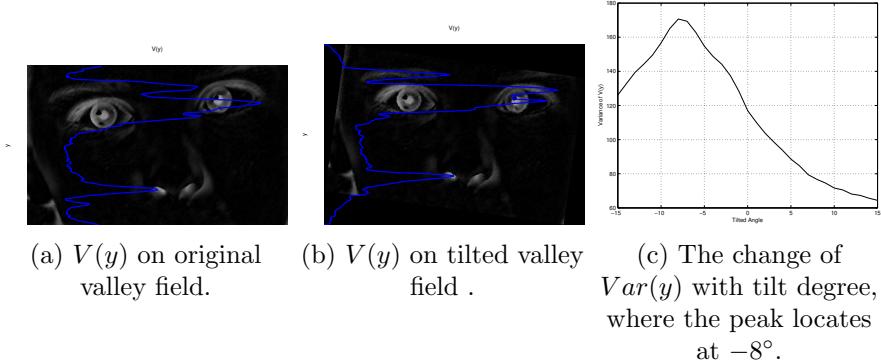


Figure 3.6: Tilt correction. In both a and b,  $V(y)$  is represented by blue line, and the vertical coordinates correspond to each horizontal line on the valley image. Original image source: <http://strabismus.com/>, 23 September 2012

## 3.2 Coarse Position Estimation

The entire part is developed from the work of Choi et al. (2009). The detail for region extraction is not described in their publication, so a different method is tried below to find the best approach. Different from their method, only the valley field is used here, since the edge field contains more irrelevant information on the eyebrow and the nose, and it varies much on different image.

### 3.2.1 Vertical Estimation

According to Choi et al. (2009), this step is to extract the peak located on the eyes in the valley field. In figure 3.7, all  $V(y)$  (which have been smoothed) has great responses in the region of eyes, then it should be quite straightforward to extract these peaks. The paragraphs below describes all the trials the author has made.

#### Gaussian Fitting

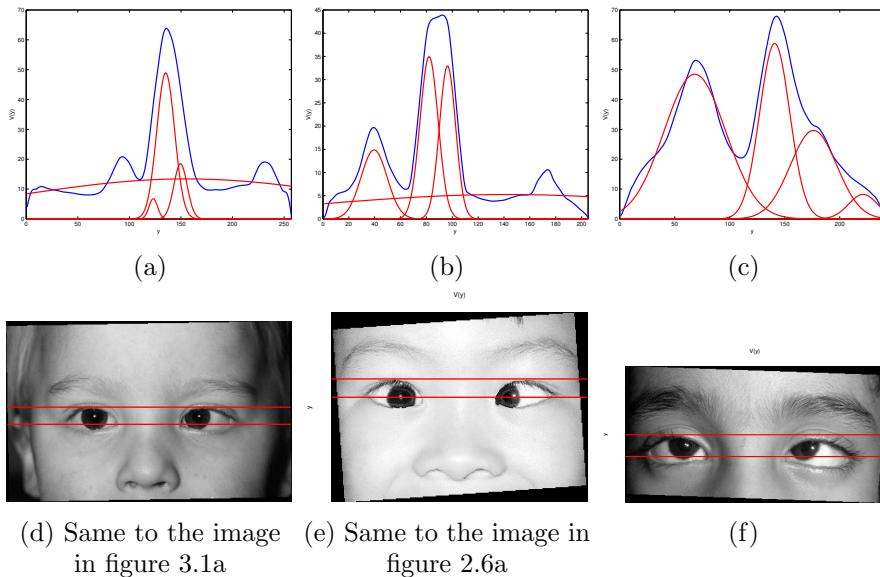


Figure 3.7: Result of fit  $V(y)$  with 4 gaussian distribution. The estimated eye region, bounded by red horizontal lines in each image, is given by  $[\mu - \sqrt{2}\sigma, \mu + \sqrt{2}\sigma]$  of the highest distribution. (Image source for (f): Karna Eye Clinic Netra Niwas, <http://squintneuroophthal.com/squint>, 23 September 2012.)

The peak representing the eye is always the highest. The most common way to extract this peak is to use multiple gaussian distribution to fit the curve. A distribution with highest amplitude can represent all information of vertical eye range. When the method is applied onto  $V(y)$  in MATLAB, it is needed to specify the number of peaks fitting the curve. The author tried 4 of every image type since from the observation there will be 4 peaks at most in each  $V(y)$ . However, the result is not ideal. In the illustration in figure 3.7 (a), (b), and (c) (corresponding to each image below), some peak are more likely to concentrate on same highest peaks, which makes every distribution narrower than the width of highest peak. From the result

in figure 3.7 (c), (d), and (e), it can easily be seen that the estimated eye range given by the highest distribution is smaller than the real eyes. They are more likely to be attracted by the strong valley of upper eyelid. It is also not applicable to reduce the amount of peaks in the fitting, because it is still hard to ensure them corresponding to all the peaks in  $V(y)$ .

### Maximum Derivative Searching

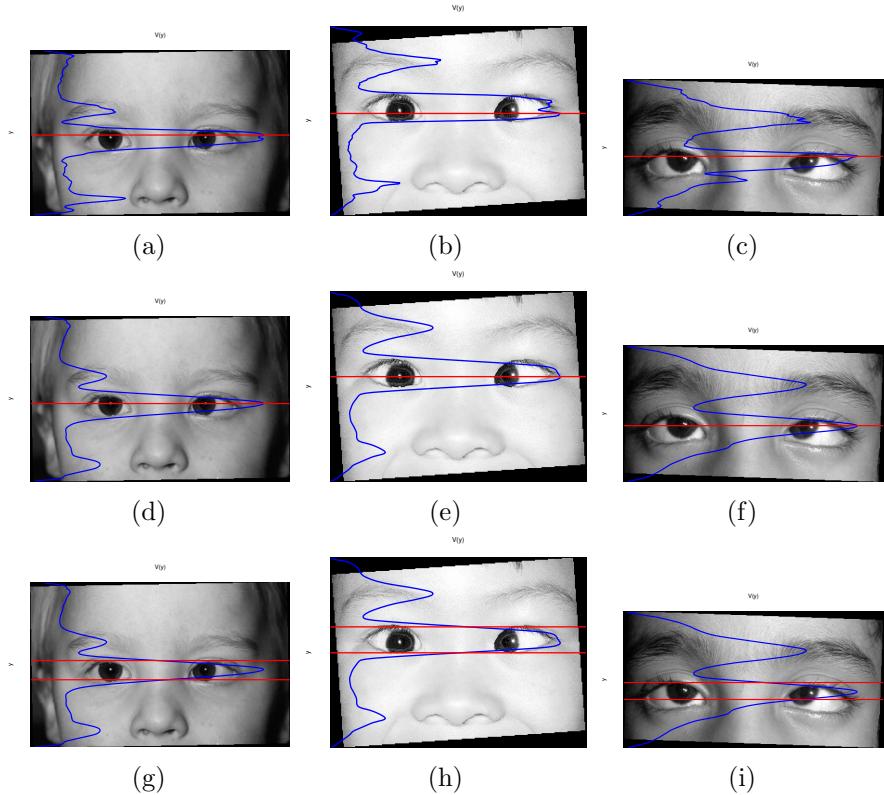


Figure 3.8: Peak extraction and maximum derivative searching

Observing each  $V(y)$ , it is easy to get the location of the maximum value, which can be used as the central line for iris. This is shown in first row of figure 3.8. However the result is not accurate. So smoothing is done on the function to spread the energy in the iris. The smoothed  $V(y)$  is denoted by  $V_s(y)$ . It is achieved through a mean filter with a span of  $H/20$ , where  $H$  is the height of the tilted image. After that, the central lines become more accurate (see second row in figure 3.8 ).

Then the author try to use some strategy focusing on local features. The eye region is actually bounded by the width of highest peak. When the top and bottom wall of the peak is found, the eye then should be bounded by

the walls. This is achieved by searching on the derivative of  $V_s(y)$ , which is denoted by  $V'_s(y)$ . The algorithm is described as below.

```

Initialise:  $y_{top} \leftarrow y_{peak}$ ;  $y_{bottom} \leftarrow y_{peak}$ ;
while  $V'_s(y_{top}) < V'_s(y_{top} + 1)$  do
     $y_{top} \leftarrow y_{top} - 1$ ;
end while
while  $V'_s(y_{bottom}) > V'_s(y_{bottom} - 1)$  do
     $y_{bottom} \leftarrow y_{bottom} + 1$ ;
end while
```

What it actually found are steepest edges of the peak, denoted as  $y_{top}$  and  $y_{bottom}$ . The third row in figure 3.8 shows their position on the image. If these boundary is used as eye region, it is still a little bit short to cover the eye. Especially, in figure 3.8i, due to the condition of hypertropia, the region covers only the revealed iris but lost some parts of eye whites. This method is still unsuccessful.

### Mixture of Interval Searching and Gaussian Fitting

To make the boundary cover eye completely, the author searches monotone intervals on  $V_s(y)$  instead of the monotone intervals on  $V'_s(y)$ . Then it can find 2 nearest local minimums around the peak. The results on the three images above is shown in the first row of figure 3.9. It is obvious that these ranges are too wide for the eye region.

Therefore, the author combines interval searching method with the gaussian fitting method together. A gaussian distribution will be fitted on the data (the blue area in images of second row in figure 3.9), which comes from  $V(y)$  but have been cut into the intervals corresponding to the monotone intervals in  $V_s(y)$  (blue area in the first row figures). Finally, the mean  $\mu_y$  and standard variation  $\sigma$  of the gaussian can perfectly represent the eye region, which is given as  $[\mu_y - \sqrt{2}\sigma_y, \mu_y + \sqrt{2}\sigma_y]$ . The final estimated vertical eye region is shown in the third row of 3.9 (bounded by red lines). Although in the image of i, it seems taller than it should be, this can be corrected during the optimisation of the deformable template. From the experiments in later steps, it is found that the template is more likely to shrinking than extending.

#### 3.2.2 Horizontal Estimation

There are different approaches to estimate horizontal eye position. Due to the property of strabismic eyes, the iris is not always in the centre when it is fixating forward. To speedup the optimisation, the initialised deformable template should hot set the iris in the centre - it should place the iris at an

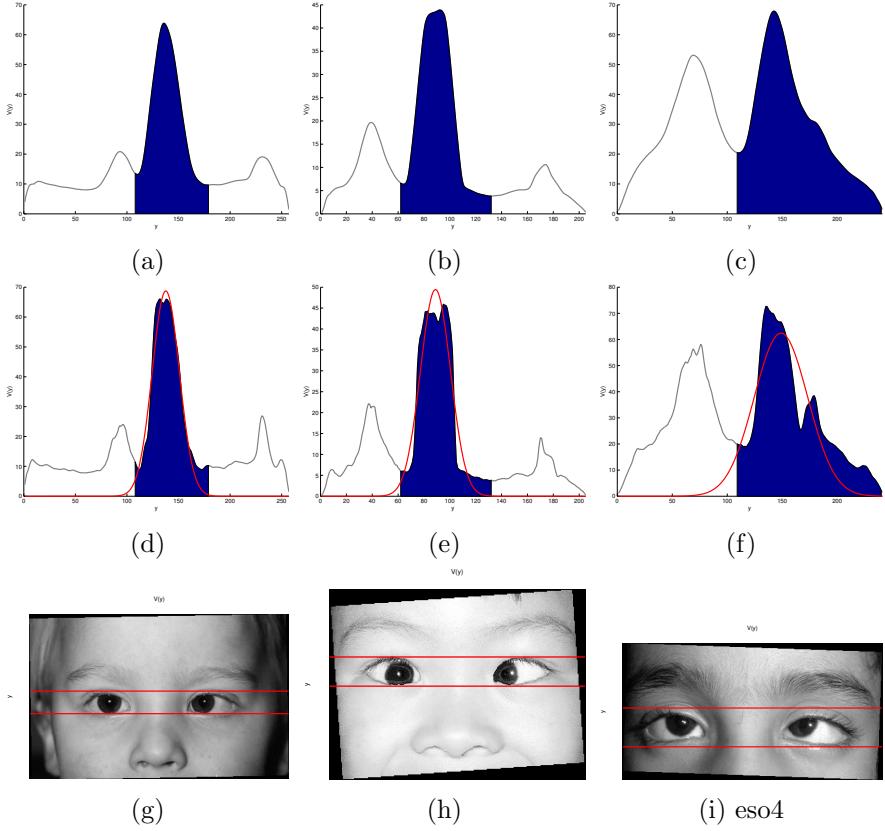


Figure 3.9: Mixture method of interval searching and gaussian fitting

estimated position within the eyelids. Then, the rough estimation can be divided into two steps - eye range estimation and iris position estimation.

### Eye Range Estimation

Blue dots in figure 3.10 represents the function  $F(x)$  (for definition, see equation eq:fx) of image in figure 3.1a and image in figure 2.6a respectively. To let it correspond to the similar function  $V(y)$ , its name is changed to  $V(x)$ . There are two symmetry bunch of peaks in each  $V(x)$ . The aim of eye region extraction is to extract the width of each bunch of peak.

Similarly to the condition on vertical direction, to fit this with 2 gaussian distribution is still not applicable. The reason for this is the iris concentrates too much energy on valley field, and what the gaussian distributions fit are energy of the irises. The fitted distribution cannot represent the whole eye region. As shown in figure 3.10a, nearby the great peaks, there are small peaks resulted by eyelids. They should contribute to the total eye width together with the great peaks. Since background may disturb  $V(x)$  and

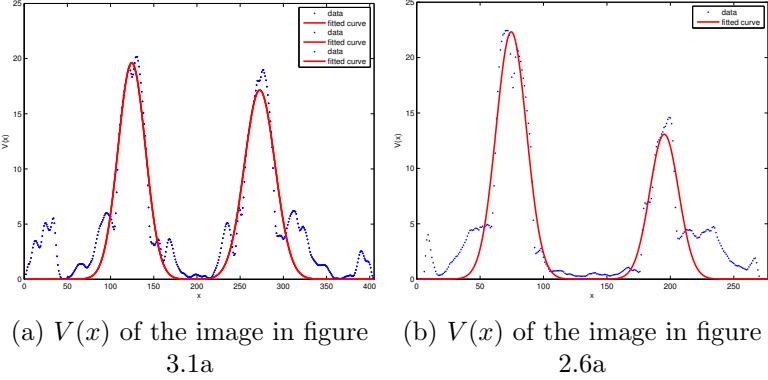


Figure 3.10:  $V(x)$  and the result of 2 gaussian fitting of it.  
 (a)  $V(x)$  of the image in figure 3.1a      (b)  $V(x)$  of the image in figure 2.6a

(a)  $V(x)$  of the image in figure 3.1a      (b)  $V(x)$  of the image in figure 2.6a

(a)  $V(x)$  of the image in figure 3.1a      (b)  $V(x)$  of the image in figure 2.6a

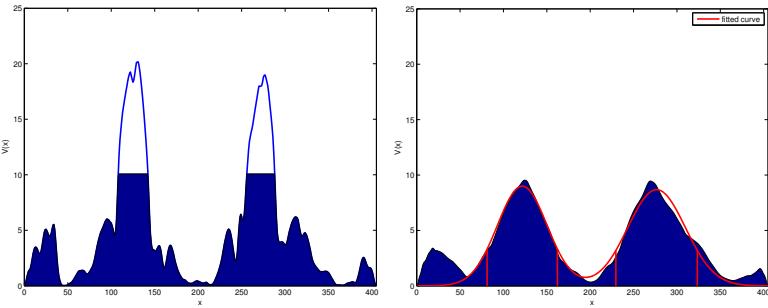
(a)  $V(x)$  of the image in figure 3.1a      (b)  $V(x)$  of the image in figure 2.6a

(a)  $V(x)$  of the image in figure 3.1a      (b)  $V(x)$  of the image in figure 2.6a

(a)  $V(x)$  of the image in figure 3.1a      (b)  $V(x)$  of the image in figure 2.6a

produce peaks at the corner (see figure 3.10a), to add up the quantity of distributions may let the fitting be misled and produce fault result. Moreover, to simply expand the extracted width of main peaks will not work, since in strabismic images the iris may locate at the side of eye. As it shown in figure 3.10b, extend width of right peak will result in an error towards left.

The solution of this problem is to reduce the effect of high energy concentration in the iris. A threshold of  $\frac{1}{2} \max\{V(x)\}$  is used to cut off the high peak in  $V(x)$ . Any value greater than this value will be set equal to it, then  $V(x)$  has only the lower half left. The function  $V(x)$  is shown in figure 3.11a, and the filled area represents the lower half of  $V(x)$ . Then, to spread the energy, it is smoothed, and the result is denoted as  $V_s(x)$ . The smoothing is completed by the mean operator with a span of  $W/10$  (where  $W$  is the width of the tilted image).



(a) The cut operation on figure 3.10a, where all energy above the threshold is removed.

(b) Gaussian Fitting of  $V_s(x)$ .

Figure 3.11: Extraction of horizontal eye ranges.

Now two gaussian can be fitted to  $V_s(x)$  (shown in figure 3.11b). Although the energy from background is still existing, it is much lighter when compared to the solid strong energy from the two eyes. It will not strongly affect the fitting, Then the range  $[\mu_x - \sqrt{2}\sigma_x, \mu_x + \sqrt{2}\sigma_x]$  of each gaussian (represented by red vertical lines in figure 3.11b) can describe the estimated location of corresponding eye. The estimated region is also marked in figure 3.12 by red line. Although in figure 3.12b, the left eye region still has an inward error, it has already been minimised when compared with the method above. Please note since the resource is facial photograph, the eye in the left of the image is actually the right eye, whereas the eye on the right is the left eye.

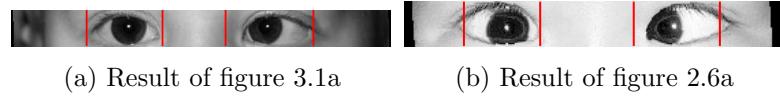


Figure 3.12: Estimated horizontal eye regions shown within the vertical range of eye.

### Iris Position Estimation

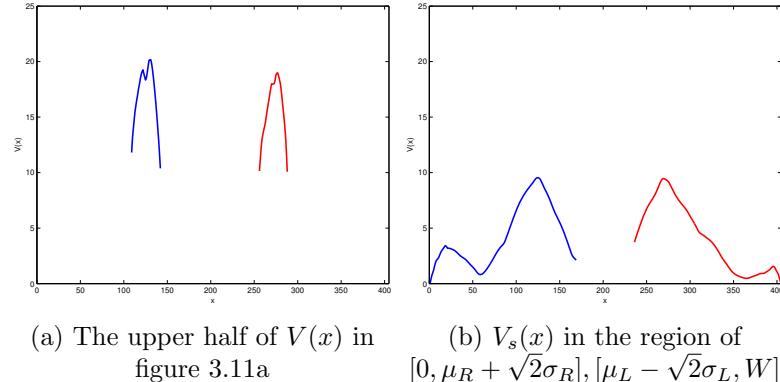


Figure 3.13: Iris horizontal position estimation.

It would be ideal if the upper halves can be used to extract iris position. However, it fails. Figure 3.13a illustrates the upper half of  $V(x)$  in figure 3.11a. The peaks in blue and red resulted from the right and left iris correspondently. The author has tried to use the locations of maxima (result on images in figure 3.1a and figure 2.6a is shown in the left column of figure 3.14), and the mean of the peak (shown in the right column in figure 3.14) as horizontal coordinate for the iris. It is clear that the later method performs better than the former one.

However, when they are compared with the location of peaks  $p_R, p_L$  (shown in the central column in figure 3.14) in the  $F_s(x)$ , it seems that the peaks are closer to the centre of irises. Consequently, the iris central position is estimated from the coordinates of  $p_R, p_L$ . Probable reason for this is the smoothing operation has polish the top plane produced by threshold, and the peaks in  $F_s(x)$  are just centres of the iris energy peak .

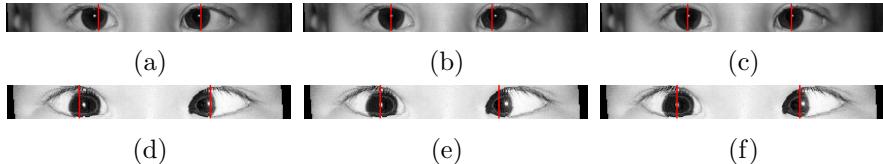


Figure 3.14: Different approach of iris centre estimation.

### 3.3 Eye Extraction

#### 3.3.1 Modification of Deformable Template

Due to probable mistake in publication, some of the energy terms in the deformable template proposed by Xie et al. (1994) is unreasonable. Based on the personal understanding of both Yuille et al.'s and Xie et al.'s template, the author here does some correction on the improved template. Regarding the property of strabismus image, some terms are also being replaced.

Firstly, the energy term  $E_3$  is attracted by peak field, so it should have a smaller value when the eye white in template covers more peak energy. Therefore a “1 -” should be added in front of the right hand of the definition. Due to the property of strabismic eye, the eye whites may not appear if the eye always poses extremely. Then to calculate the peak within small window on mid-point may be meaningless. To attach to the eye whites more precisely, the author use the area of the entire exposed eye whites to calculate summation of peak field. This area is got by the area within parabolas subtracting the area of iris. Then  $E_3$  becomes:

$$E_3 = E_p = 1 - \frac{1}{Area_{eye\ whites}} \sum_{j \in eye\ whites} b_j^{(P)}$$

Secondly, the image potential of iris should have lower value when it covers more dark pixels:

$$E_5 = E_{ic} = \frac{k_5}{Area_{circle}} \sum_{i \in circle} g_i$$

and the image potential of edge should also have lower value when it covers more dark edge:

$$E_7 = E_{ie} = \frac{k_7}{Length_{parabolas}} \sum_{i \in parabolas} g_i$$

Thirdly, due to the condition of strabismus again, the internal energy term  $E_8$  will not work. The task of it is trying to pull the iris and the centre of parabolas together, but iris of a strabismic eye is more likely to go to extreme positions. If this energy term is applied to strabismus image, it can probably slow down the optimisation or lead to large error.

On the other hand, the radius of the iris in the image is a crucial parameter for final diagnosis. Usually, radius of both irises should be the same. If the radius of irises  $r_L$  and  $r_R$  is compared and the difference between them can be minimised, they can help each other to find a suitable value. Therefore, the internal term  $E_8$  now has been replaced by:

$$E_8 = k_8 |r_L - r_R|$$

This change will make an improvement on the the accuracy of the final diagnosis.

Finally, it is obvious the iris cannot appear beyond the parabolas. So the author redefines the area of the iris as the intersection of the disk and area within parabolas. This change is especially useful when the eye is in a extreme position. It can help the template to be less affected by the skin around the eye.

### 3.3.2 Implementation of Template-based Extraction

#### Calculation Energy Terms

That optimisation of deformable templates uses Levenberg-Marquardt algorithm, where the partial derivative of each energy terms will be calculated and supervised the step of parameter change. Since all the image terms ( from  $E_1$  to  $E_7$  ) are not linear function of the parameters. Calculation of these terms will be quite time-consuming. Moreover, getting the partial derivative means to calculate them several times with different parameter values in a trial.

To accelerate the optimisation, the author created a mask for each energy. Each mask is a binary image, which has the size to all the fields and the grayscale image. A mask will set all “active” pixel for the current term to 1, while the “inactive” pixel will be set to zero in the mask. The word

“active” here means this pixel will take part in the calculation of current term. To calculate each energy term, the mask corresponds to it will be covered on the energy field, then the summation of pixels leaked by the mask is the energy value for this term. Any change in parameter set  $X$  will change the related mask, but the way to calculate energy summation keeps the same, which will certainly accelerate the process of partial derivatives.

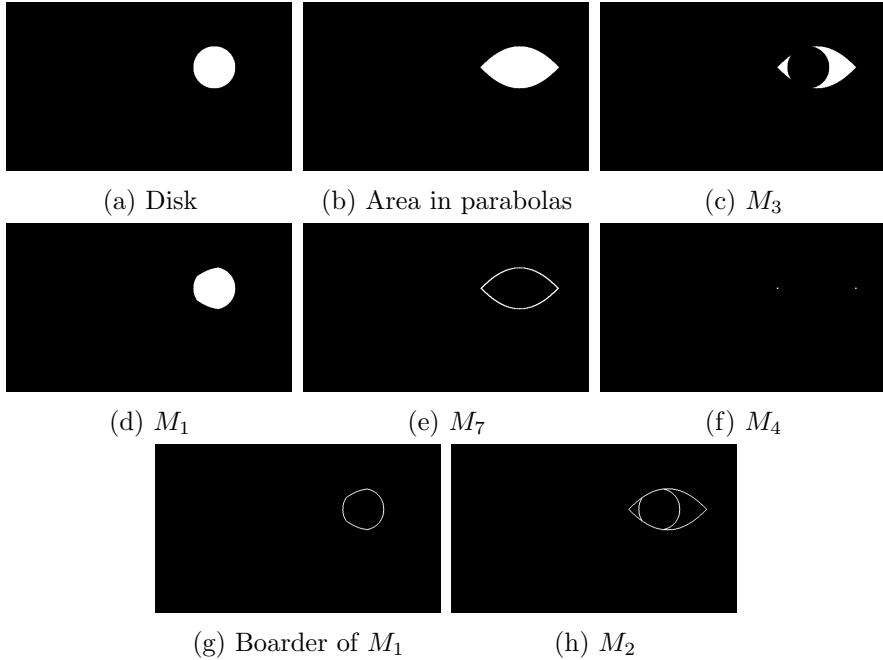


Figure 3.15: Masks for energy calculation. (These masks are created for the left eye in the image of figure 3.17.)

Since some masks are related to others, the creation of masks is not only drawing shapes, but also do logical and morphological operations between them. All the created mask is illustrated in figure 3.15, but some shared masks is only illustrated once (i.e. masks shared by  $E_1$  and  $E_5$ , and by  $E_3$  and  $E_6$ ).

The masks are created consequently. Firstly, a disk is created for iris. The condition of it is given as :

$$(x - x_c)^2 + (y - y_c)^2 \leq r^2$$

where  $x_c, y_c$  are coordinates of iris centre  $\vec{X}_c$  on x and y direction. The pixels satisfying this condition will be set to 1 in the mask. It is shown in figure 3.15a. Another basic shape is the area bounded by parabolas, whose

condition is given as

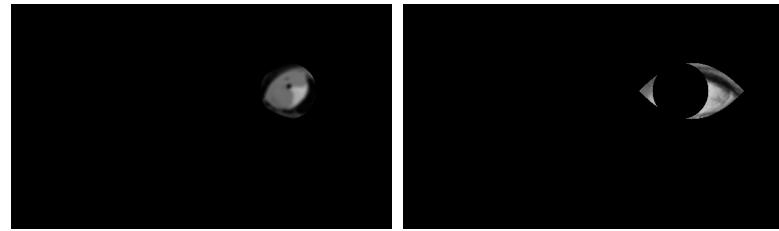
$$\left(\frac{\sin(\theta)(y + y_e) + \cos(\theta)(x + xe)}{b}\right)^2 - \frac{\cos(\theta)(y + y_e) - \sin(\theta)(x + xe)}{a} \leq 1$$

$$\left(\frac{\sin(\theta)(y + y_e) + \cos(\theta)(x + xe)}{b}\right)^2 + \frac{\cos(\theta)(y + y_e) - \sin(\theta)(x + xe)}{c} \leq 1$$

where  $x_e, y_e$  are coordinates of eye centre  $\vec{X}_e$  on x and y direction. The pixels satisfying both conditions will be set to 1 in the mask.

Then  $M_1$  (shown in figure 3.15d) is created by the intersection of the disk and area in parabolas.  $M_7$  (figure 3.15e) is created by the area in parabolas when its erosion is subtracted (see subsection 2.2.1).  $M_3$  (figure 3.15c), representing the eye whites, is created by subtraction of area inside parabolas and  $M_1$ . The same way to extract  $M_7$  can be used to get the boarder of iris (figure 3.15g), which then union with  $M_7$  to create the mask  $M_2$  (figure 3.15h).

Finally, creating the mask  $M_4$  is done by finding two corner points of the eye, and then drawing a 3-by-3 around each point.



(a)  $M_1$  Covering on the Peak Field for  $E_1$  Calculation. (b)  $M_3$  Covering on the Image Field for  $E_6$  Calculation.

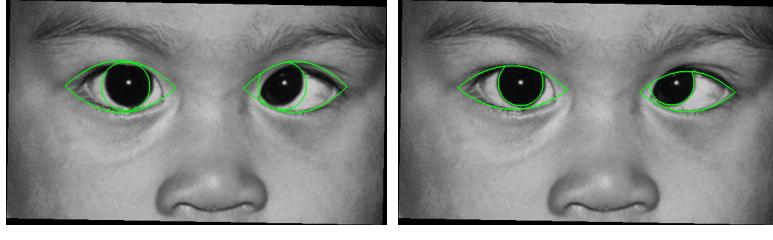
Figure 3.16: Masks applied on the potentials for energy calculation.

Then, the masks can be pixel-by-pixel multiplied with the corresponding potential fields, and the summation of all nonzero pixel values is the energy for current term. The pixel-by-pixel multiplication can be completed by the “.\*” operator in MATLAB. The effect of multiplication is shown in figure 3.16.

### Levenberg-Marquardt Optimisation

In the Euclidean plane, x and y axis are orthogonal, which requires that the coordinate on each axis should be optimised separately. So in the implementation, parameters  $\vec{X}_c$  is split into two coordinates  $x_c, y_c$  for computation. Similarly,  $\vec{X}_e$  is split into  $x_e, y_e$ . Altogether, there are 9 scalars controlling a template. The parameters  $X = \{x_c, y_c, x_e, y_e, a, b, c, r, \theta\}$  are first initialised

with estimated values in the former step. Take the left eye for example, the parameters will be initialised by:  $y_c, y_e \leftarrow \mu_y$ ;  $a, c \leftarrow \sqrt{2}\sigma_y$ ;  $x_e \leftarrow \mu_L$ ;  $b \leftarrow \sqrt{2}\sigma_L$ ;  $x_c \leftarrow p_L$ ;  $\theta \leftarrow 0$ . The initialised template is shown in 3.17a. The green curves for illustrating the template is complete by an addition of  $M_2$  to the image.



(a) Initialised template      (b) Optimised template

Figure 3.17: Deformable template optimisation. The template is represented by green curves. Original image source: “Child with amblyopia”, when should my child have an eye exam, URL <http://www.oregoneyes.net/blog/when-should-my-child-have-an-eye-exam/>, 24 September 2012.

According to Xie et al. (1994), the template can find its optimum state through the Levenberg-Marquardt algorithm. The algorithm is to change parameters to a state where the total energy function is minimised. Each eye in the image has an individual template. Only in the calculation of  $E_8$ , they share the radius of the iris to each other. The framework of the implementation is illustrated below. Here parameters  $X = \{x_c, y_c, x_e, y_e, a, b, c, r, \theta\}$  in step  $k$  are denoted by  $X_k = \{x_{1,k}, x_{2,k}, \dots, x_{9,k}\}$ .

```

for  $k = 1 \rightarrow k_{limit}$  do
     $X_k \leftarrow X_{k-1} + d_k - 1$ ; (equation 2.14)
     $F_k(X) \leftarrow \sum_{i=1}^{10} W_i E_i^2(X_k)$ ; (equation 2.13)
    if  $F_k > F_{k-1}$  then
         $\lambda \leftarrow \lambda * \alpha$ ;
    end if
    if  $|d_k| \leq d_{limit}$  then
        terminate;
    end if
end for

```

where weights  $W$  are constants set before the optimisation process.  $W$  are the only coefficients controlling the length of the increment  $d_k$ . Ideally, it should be proportional to the size of the facial features. Here the author use the estimated distance between the eye as a measurement of the facial size. To make reasonable values,  $W$  also has to be normalised before the multiplication with eye distance. To sum up, given a set of preset weights,

it will be processed via the following expression:

$$Wi = \beta|x_{eL} - x_{eR}| \frac{W_i}{\sum_{i=1}^{10} W_i}$$

where  $\beta$  is a constant coefficient controlling the length of  $W$ . Based on experiments, it performs best with a value of 1/15.

The increment  $d_k$  in each step is given by the solution of equation 2.15:

$$(J_k^T W J_k + \lambda I) d_k = -J_k^T W F_k$$

where  $J$  is the 10-by-9 Jacobian matrix and  $\lambda$  is a coefficient to balance the speed and accuracy. Based on the observation of the derivative of energy functions, the Jacobian matrix can supervise the parameter change towards optimum. Each element in the Jacobian matrix is defined as a partial derivative  $J_{i,j,k} = \partial E_i / \partial x_{j,k}$ . Since each parameter in the experiment is discrete, and not all image energy terms are differentiable functions, the partial derivative is not achievable. Each element is then replaced by the central differences of a discrete function:

$$J_{i,j,k} = \frac{Ei(x_{1,k}, \dots, x_{j,k} + 1, \dots, x_{9,k}) - Ei(x_{1,k}, \dots, x_{j,k} - 1, \dots, x_{9,k})}{2}$$

In the experiment, all the elements in  $J$  are consequently calculated by the above expression in each optimisation step, and the optimisation will terminate the optimising cycle when it reaches the upper limit  $k_{limit}$ . The limit is set to 10 in the experiment. Moreover, the author implemented some external strategy to control the optimisation. The coefficient  $\lambda$  is first set to 0.2, and it will be multiplied by a ratio  $\alpha$  once  $F_k(X)$  becomes greater. The effect of it is to make the algorithm attempt to find a greedy route when its total energy does not decrease. Another process of control is to terminate the loop when the length of  $d_k$  is less than the lower limit  $d_{limit}$ , which is 1 in the experiment. This can stop optimisation when the template is in its optimum, therefore total time can be saved during processing. An example of optimised template is shown in figure 3.17b. Due to a good prior estimation, the optimisation terminates after 6 steps and reached a very good state.

### 3.4 Reflex Extraction

The last step carried on the the image is to get the location of the reflex. Within the extracted iris, the reflex appears as a small bright dot. Whatever the colour of the iris is, the reflex on cornea should be much brighter than the iris for most people. This should be compared on the greyscale image but not peak image, since the peak image can highlight local bright features

and mislead the process. In a grayscale facial image, it is reasonable to pick up the brightest pixel as the reflex. However, on a medium size image, the reflex may appear like a small bright disk (seen in figure 3.18a). All pixels within this disk may have the same high value. Determined by the internal algorithm of the experimental platform, the chosen maximum pixel may not be located at the centre of the disk. If the coordinate of this pixel is used for the location of the reflex, the difference between this and the real reflex centre may lead to an obvious error in the diagnosis. A error of 6 PD may result from the difference of 1 pixel when the iris has a size of 40-pixel width.

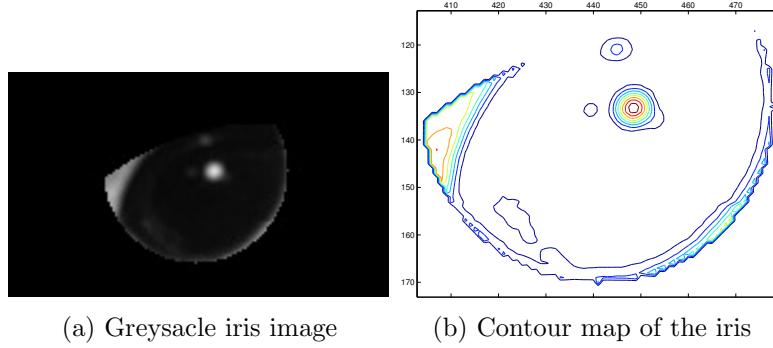


Figure 3.18: Corneal reflex extraction on the image of figure 3.17

Therefore, to extract the location as accurate as possible, the average of the coordinates of these bright pixels is needed. Due to the performance of the template in the previous step, the area of the extract iris may still cover a part of eyelid, which is also bright on the image (the small triangle on the left in figure 3.18a). To use a gaussian distribution will be disturbed by the eyelid, while to calculate the weighted average of all pixels will not work for the same reason. The author finally used the average of the brightest pixels as the result. If the iris image is denoted by  $I$ , and location of reflex is  $(x_r, y_r)$ , the method can be given by:

$$x_r = \bar{x}_i, y_r = \bar{y}_i, \text{ where } x_i, y_i \in \{I(x_i, y_i) = \max I\}$$

The contour map of values is shown in figure 3.18b, from where it is seen that although the triangle of the eyelid has a large area, its amplitude is lower than that of the main reflex. So if only the pixels with highest value are selected, the eyelid can have no effect. A result of the reflex extraction is shown in figure 3.19. It is easy to prove that this method works on whatever size of reflex: if only one pixel has the peak value, it should be the centre of the reflex; otherwise, the pixels that share the peak value should all come from the reflection and their centre can be used as the reflex centre. This method also works on non-circle reflexes.

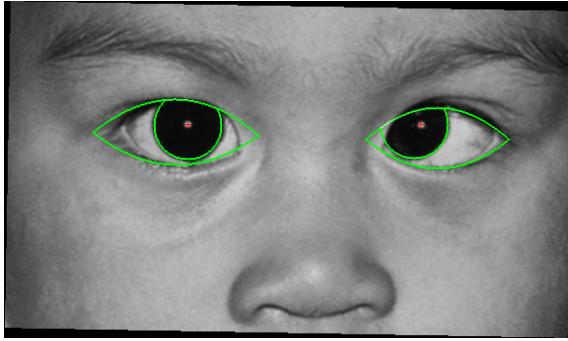


Figure 3.19: Extracted reflexes on the image of figure 3.17. They are represented by the small red crosses in the iris.

### 3.5 Diagnosis

The definition of Hirschberg ratio can directly give the expression to calculate the iris deviation in the prism diopter: multiplying the ratio of distance between the centre of the iris and the reflex has its unit as mm. The reflex for the eye had been extracted in former step, and the centre of the iris is just  $X_e$ . However, they are all in the image coordinates. A transition is needed to convert the distance between these two points to the world coordinates. Fortunately, there is an important property for the human eye: the size of the iris will be unchanged from the age of 1 years old. According to Davis and Becherer (2005) the average iris diameter (formally called Horizontal Visible Iris Diameter, HVID) of adults and children are the same 11.8mm. Based on this constant, the ratio between the world coordinates and image coordinates can be easily calculated, given as  $HVID/2r$ . Then it makes the strabismic deviation become accessible from the image. The horizontal deviation is given by

$$D = |x_r - x_c| * \frac{HVID}{2r} * HR;$$

where HR is the Hirschberg Ratio, a constant of 22 PD/mm according to Eskridge et al. (1988). In order to distinguish the deviation of esotropia and exotropia, the author defines the positive direction of the deviation: a positive value corresponds to the eye turning outward, and a negative value corresponds to “squint” eyes.

In the experiment, only the horizontal prism diopter is calculated. The reasons for this are listed below.

1. In the experiment, the test images there are conditions of either esotropia or exotropia. These two condition also have an overall higher presence than vertical deviated strabismus.

2. The experiment sometimes shows that the deformable template can have a larger error on vertical direction. Focusing on the horizontal direction can minimise the vertical error.
3. The estimation method by (Schwartz, 2006) is only applicable for horizontal misalignment. It is hard to visually estimate the reference value on the vertical direction. Therefore the correction of vertical strabismus cannot be quantified.

The final diagnosis is given by a very conservative strategy. Considering the probable error, values absolutely greater than 20 PD can trigger a statement of strabismic eyes. If the eye has a deviation smaller than -20PD, the system will output the text “Esotropia” and show the prism diopter. A similar output of exotropia will be made when the deviation is greater than 20PD. Figure 3.20 shows a screen shot of a diagnostic output.

**Diagnosis:**  
**Left eye: Esotropia -25PD**

Figure 3.20: Screen output of the diagnosis on image 3.17.

## Chapter 4

# Result and Analysis

Quantification of the experimental result is done by comparing the result value in PD with visually estimated reference value. The diagnosis does not take part in the result because it includes only conditional statements, and will not show an error. The reference value is produced according to the estimating method by (Schwartz, 2006), illustrated in figure 2.5. Here more conditions (15 PD, 45 PD, and 75 PD) have been added to describe the deviation between the key values of 0 PD, 30 PD, 60 PD, and 90 PD. Regarding the error of visual assessment, the experiment values within the interval of (-15 PD, 15 PD) of the reference value can all be marked as reasonable values. For instance, the experimental result of either 36 PD or 18 PD can both be the correct value when a reference of 30 PD is given.

The experiment is carried out in the MATLAB environment on a Macbook Pro with an Intel i7 quad core processor. Normally the initialisation will complete instantly and each step of the template optimisation takes less than 0.5 seconds. The speed highly depends on the image size. It can be obviously slow when the image resolution is higher than 800x600. Therefore, most image have been scaled to a width of 400 pixels to accelerate the processing.

There are a total of 42 images for the experiment. The whole image set covers the conditions of esotropia, exotropia, and pseudostrabismus. In the experiment, the system can measure the deviation of both eyes correctly on 24 out of 42 images. The correction rate of the images is 57.14%. There are two images on which the algorithm has extreme fault in the eye location estimation. The detailed experimental result and reference values of each image set are presented in a table which can be found in the appendix A.

If the statistics is carried on individual eyes, 64 out of 84 eyes can get a correct measurement by this algorithm, which makes the correction rate

of eyes to be 76.19%. It is obvious that the eye correction rate is higher than the image correction rate. The reason for that is some partially correct image are classified to the problematic group, but the correctly measured eye in them contributes to the correct eye group. Figure 4.1 illustrates the correspondence of experimental result and their reference values. Then figure ?? shows the distribution of the error.

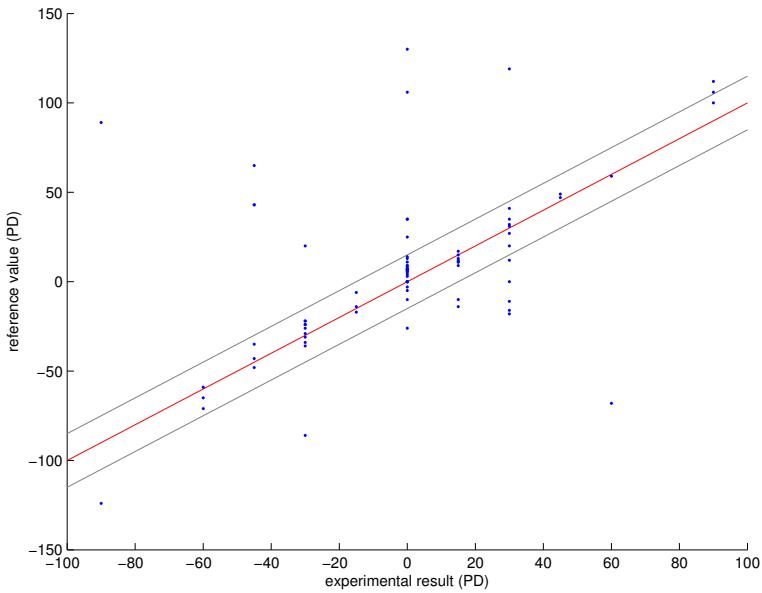


Figure 4.1: Correction of the experimental result. The  $y$  axis represents the experimental value and the corresponding reference value is represented by  $x$  coordinate. The red line is the place where  $y = x$ . Ideally all dots should be located on this line. The grey lines  $y = x - 15$  and  $y = x + 15$  are the boundaries for correct data points. 64 out of 84 data points are locate within the boundary. On the points with reference values of 15 PD and 30 PD, there is a trend for their experimental value becoming smaller, while the points with reference value 0 have a concentration of greater values.

By observing Figure 4.1 and 4.2 together, it is obvious that an experiment value is more likely to have a small positive error especially when it is around 0 PD. This phenomenon is more clear in figure 4.3, where the experimental values of only pseudostrabismus image set are illustrated. A possible reason for this phenomenon is the error of reference value . The author's subjective assessment may result in inaccurate reference values. Taking pseudostrabismus for example. When the image is marked as pseudostrabismus, the deviation of both eye will more likely be set as 0 by the author. However, the reflex may not be located at the perfect centre of the iris, and the true deviation can be calculated by the proposed method. So

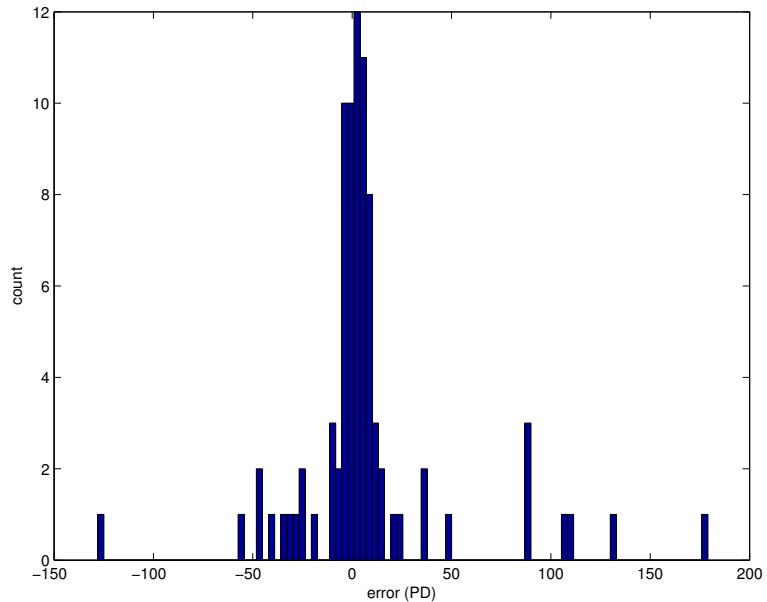


Figure 4.2: Error distribution. Most of the errors are in the interval of [-10 PD, 20 PD], and there is a data concentration around the error of -20 PD.

the error of this type is actually not due to the mistake of algorithm, but the inaccuracy of reference values. Moreover, the positive experimental values reveal that the angle kappa for these pesudostrabismic infants are mostly temporal.

In terms of individual images, the experimental results are mostly visually acceptable. As shown in figure 4.4, the parabolas describing the eyelid may not fit perfectly to the real eyelid, but this is not signification to the extraction of the iris. The job of shapes other than the circle in the template is to help the circle find its right location. Both images shown have the iris accurately extracted.

Some images failed in the experiment. Figure 4.5 and figure 4.6 are examples of the failed images. The former example reveals its performance when the eye is in extreme position, while the latter example proves this method is sensitive to blurring and environmental lighting direction.

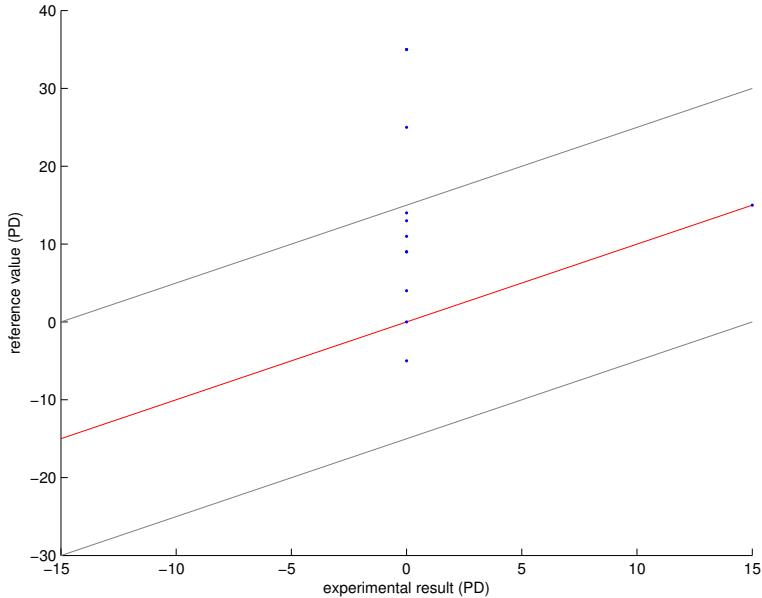


Figure 4.3: Correspondence of the experimental value and reference value of pseudostrabismus image set. Since the eyes are all pseudostrabismic, most reference values are set to 0. But most of the experimental values turned out to be acceptable positive value.

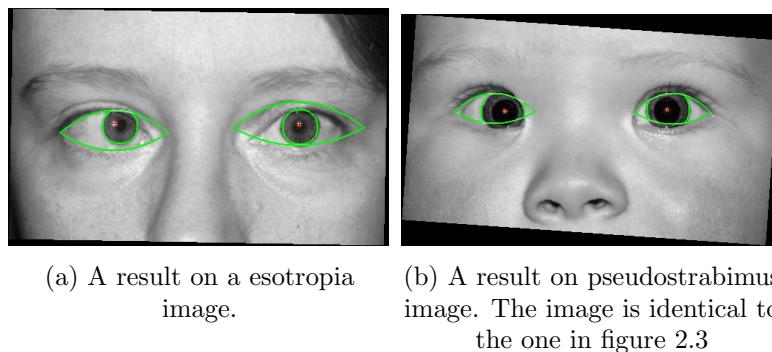
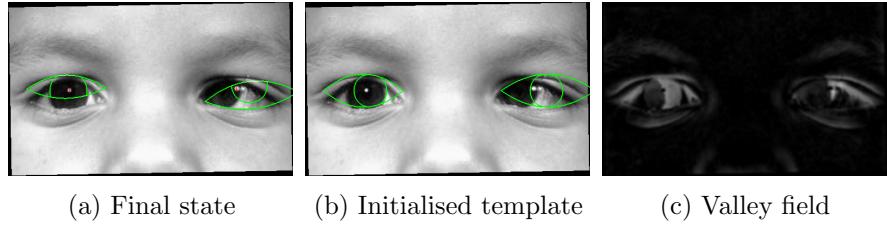
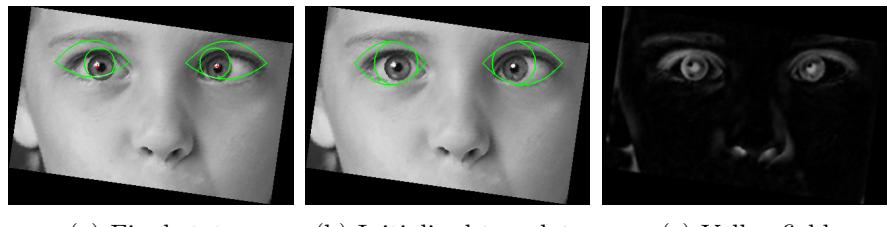


Figure 4.4: Some good results. (Image source for a: Fig.1 “Adult strabismus, or misaligned eyes, can be surgically corrected.”, Adult Strabismus, URL <http://www.aapos.org/terms/conditions/11>)



(a) Final state      (b) Initialised template      (c) Valley field

Figure 4.5: A failed example where the reflex is close to eye whites. (Image Source: “Strabismus”, Eye clinic of Racine, Ltd., <http://www.eyeclinicofracine.com/index.cfm/pediatric/strabismus>.) The left eye has an obvious esotropia, but the iris in the template is attached to the eye whites. The error begins during the initialisation, where the estimated left iris location is already in the whites. These faults are caused by the valley field, in which the left iris has a very low response, while the texture created by the reflex on the white has a concentration of more energy. Similar conditions are present more on the exotropia images, where the reflex can be located on the eye whites. Some representation other than valley field is needed to resolve this problem.



(a) Final state      (b) Initialised template      (c) Valley field

Figure 4.6: Another failed example caused by blurring. (The image is identical to the one in figure 3.6.) The result is also revelent to the light, which is not a typical Hirschberg light but located on top of the patient. They both make the upper eyelid dark and strong, which responds greatly on the valley field. Then it attracts the template in both initialisation and optimisation.

## Chapter 5

# Conclusion and Further Work

### 5.1 Conclusion and Critical Evaluation

In conclusion, this method achieved the objective described in chapter 1. It can automatically access the condition of strabismus (currently only esotropia and exotropia) from a properly captured facial image, although the final correction rate is 57% overall. The experimental system could probably make greater error when the eye is in an extreme position.

The result of the diagnosis highly depends on the quality of the Hirschberg image, where any blur or problematic lighting can cause a failure in the proposed method. To get the best image as possible, the photographer is advised to use the built-in flashlight as the light source and stand with a distance away from the patient. This is to avoid blurring as well as protecting the patient from the bright light source.

The deformable template optimisation process is the key algorithm of this project and the template should be initialised within an estimated eye region. A well-optimised template highly depends on a suitable eye region. The estimating algorithm is aimed at giving a quick response by analysing the valley field. Although multiple different methods have been studied, the best algorithm for vertical and horizontal estimation are still not totally reliable. Some robust representation of the images may be able to correct wrong estimation in minor occasions.

The iris extraction is done by deformable template estimation, which is effective and accurate with a reasonable initialisation. The drawback of it is that the processing time is relatively long due to the partial difference calculation in the Levenberg-Marquardt minimisation. To process a

medium-sized image usually takes a few seconds, which makes the proposed method not suitable for real-time diagnostic system.

The author then proposed the method to extract the reflex and adapted the equation to quantify the eye deviation based on the Hirschberg ratio. Both of these steps make few errors.

When the implementation has been completed, the author found a research carried out by Dallyson Sousa De Almeida et al. (2012), which has exactly the same goal as this project. They take a similar framework but implement some more complex approaches. They first build an aid eye window by similar geographical functions. Then, to build the narrower eye bounding box, an SVM based detector is implemented to give a reliable estimation. This approach is much more robust and smart than the one in this project. The drawback of it is that prior training is needed for the SVM, and the computational complexity is higher than the fast method in this project.

Another main algorithm of the eye extraction by Dallyson Sousa De Almeida et al. (2012) is importing the Hough transformation to detect the circle in the image, and then pick out the border of the iris by its strength in Hough space. An obvious advantage of this algorithm is the short processing time. However, in their publication, there is no description of the response of Hough transformation when the eye is in the extreme position. If the Hough transformation has a light response to an extremely posed iris, it may not be able to replace the approach of deformable template.

The research conducted by Dallyson Sousa De Almeida et al. (2012) has a very high matching rate, which is as a result of the high quality image sets and professional ophthalmological supervision. In terms of this project, the accuracy is enough for a qualitative analysis but the result is not currently reliable because a data set for accurate diagnosis is needed for detailed error analysis. Simple visual assessment of the image is not enough. Lack of the professional reference data is one of the limitations of this project. All the prior knowledge in ophthalmology is learnt by the author, and all the data set is collected by using Google image search. For each image, there is no professional reference diagnosis. Furthermore, all the reference data is personally estimated by the author. All these reasons make this project hard to get a high reputation in terms of accuracy.

## 5.2 Recommendations for Further Work

Other than implementing SVM for eye bounding box estimation and replacing the deformable template with the approach of hough transformation,

there are some probable improvements on this project.

1. Capture a high quality image set for the experiment and let ophthalmologists give corresponding diagnosis for result reference. This needs the cooperation of clinics and patients, which currently cannot be achieved by the author.
2. Add more representation of the image for eye window estimation. If a smile threshold image is used together with the valley image, the iris may be extracted more accurately.
3. Model the cornea following the expression given by Brodie (1992). Then the calculation of deviation is no longer a linear function, and the result for large deviation can be more accurate.
4. Adapt to a real time video system to produce further diagnosis. If the extraction has been speed by adapting the method of Dallyson Sousa De Almeida et al., the system can try to track the iris and access the ocular motley of the patient.

# Bibliography

- D. L. Bremer, E. A. Palmer, R. R. Fellows, J. D. Baker, R. J. Hardy, B. Tung, and G. L. Rogers. Strabismus in premature infants in the first year of life. *Archives of Ophthalmology*, 116(3):329–333, March, 1998. ISSN 0003-9950.
- S. E. Brodie. Corneal topography and the hirschberg test. *Applied Optics*, 31(19):3627–3631, Jul 1992. ISSN 0003-6935.
- J. K. Choi, S. T. Chung, J. Kim, S. Cho, Y. Kim, S. Ki, S. Choi, J. Yang, and T. Kim. Reliable and fast eye detection. *Proceedings of the 9th Wseas International Conference on Signal Processing, Computational Geometry And Artificial Vision (ISCGAV'09)*, pages 78–82, 2009. ISSN 1790-5109.
- R. P. Crick and P. T. Khaw. Squinting eyes (strabismus) – disorders of ocular motility. In *A Textbook of Clinical Ophthalmology: A Practical Guide to Disorders of the Eyes and Their Management (3rd edition)*, pages 221–228, 239. World Scientific Pub. Co. Inc., February 2003.
- J. a. Dallyson Sousa De Almeida, A. CorrêA Silva, A. Cardoso De Paiva, and J. Antonio Meireles Teixeira. Computational methodology for automatic detection of strabismus in digital images through hirschberg test. *Comput. Biol. Med.*, 42(1):135–146, Jan. 2012. ISSN 0010-4825. doi: 10.1016/j.combiomed.2011.11.001. URL <http://dx.doi.org/10.1016/j.combiomed.2011.11.001>.
- R. L. Davis and P. D. Becherer. Techniques for improved soft lens fitting, January 2005. URL <http://www.clspectrum.com/articleviewer.aspx?articleid=12852>.
- J. B. Eskridge, B. Wick, and D. Perrigin. The hirschberg test - a double-masked clinical-evaluation. *American Journal Of Optometry And Physiological Optics*, 65(9):745–750, September 1988. ISSN 0093-7002.
- R. Fisher, S. Perkins, A. Walker, and E. Wolfart. Adaptive thresholding, 2003. URL <http://homepages.inf.ed.ac.uk/rbf/HIPR2/adpthrsh.htm>.

- T. Kanade. Computer recognition of human faces. *Interdisciplinary Systems Research*, 47, 1977.
- J. Lee, R. Haralick, and L. Shapiro. Morphologic edge detection. *IEEE Journal of Robotics and Automation*, 3(2):142 –156, april 1987. ISSN 0882-4967. doi: 10.1109/JRA.1987.1087088.
- Y. LeGrand and S. ElHage. *Physiological Optics*. Springer Series in Optical Sciences, 1980.
- P. Maragos. Morphological filtering for image enhancement and feature detection. In A. C. Bovik, editor, *Handbook Of Image And Video Processing*, Communications, Networking and Multimedia Series, chapter 3.3. Elsevier Academic Press, November 2004. ISBN 9780121197926. URL <http://books.google.co.uk/books?id=UM\GCfJe88sC>.
- J. M. Martinez, J. M. Holmes, N. N. Diehl, and B. G. Mohney. Incidence of strabismus in an adult population. *ARVO Annual Meeting Abstract Search and Program Planner*, 2011, May 2011.
- H. Mirzaie. Review of strabismus, March 2009. URL <http://www.slideshare.net/hmirzaeee/review-of-strabismus>.
- S. Mukhopadhyay and B. Chanda. A multiscale morphological approach to local contrast enhancement. *Signal Processing*, 80(4):685 – 696, 2000. ISSN 0165-1684. doi: 10.1016/S0165-1684(99)00161-9. URL <http://www.sciencedirect.com/science/article/pii/S0165168499001619>.
- M. M. Parks, C. J. Kelly, and P. R. Mitchell. Alignment. In *Duane's ophthalmology on CD-ROM, 2006 editon*, volume 1, chapter 6. Philadelphia: Lippincott Williams & Wilkins, 2006.
- P. M. Riddell, L. Hainline, and I. Abramov. Calibration of the hirschberg test in human infants. *Investigative Ophthalmology & Visual Science*, 35 (2):538–543, February 1994. ISSN 0146-0404.
- R. P. Rutstein, M. S. Cogen, S. A. Cotter, K. M. Daum, R. L. Mozlin, and J. M. Ryan. *Optometric Clinical Practice Guideline, Care of the Patient with Strabismus: Esotropia and Exotropia*. American Optometric Association, 243 N. Lindbergh Blvd., St. Louis, MO 63141-7881, 2011.
- G. S. Schwartz. *The Eye Exam: A Complete Guide*, page 102. Slack Incorporated, 1st edition, January 2006.
- X. Xie, R. Sudhakar, and H. Zhuang. In improving eye feature-extraction using deformable templates. *Pattern Recognition*, 27(6):791–799, Jun 1994. ISSN 0031-3203. doi: 10.1016/0031-3203(94)90164-3.

- A. L. Yuille. Deformable templates for face recognition. *Journal of Cognitive Neuroscience*, 3(1):59–70, 1991. ISSN 0898-929X. doi: 10.1162/jocn.1991.3.1.59.
- A. L. Yuille, P. W. Hallinan, and D. S. Cohen. Feature-extraction from faces using deformable templates. *International Journal of Computer Vision*, 8(2):99–111, Aug 1992. ISSN 0920-5691. doi: 10.1007/BF00127169.

# Appendix A

## Full Experimental Data

### A.1 Esotropia Image Set

Left Eye			Right Eye		
Experiment	Reference	Correction	Experiment	Reference	Correction
89	-90		7	0	✓
-16	30		-22	-30	✓
20	-30		0	0	✓
-14	-15	✓	7	0	✓
6	0	✓	-34	-30	✓
-11	30		43	-45	
-65	-60	✓	0	0	✓
-26	0		-26	-30	✓
7	0	✓	-29	-30	✓
-24	-30	✓	-3	0	✓
-31	-30	✓	6	0	✓
-48	-45	✓	5	0	✓
Initialisation Fail			-	-	
-124	<-90	✓	11	15	✓
65	-45		-36	-30	✓
-24	-30	✓	0	30	
-18	30		0	0	✓
-71	-60	✓	31	30	✓
-59	-60	✓	11	15	✓
-22	-30	✓	11	15	✓
-35	-45	✓	8	0	✓
130	0		43	-45	
20	30	✓	-43	-45	✓
-86	-30		-10	15	

## A.2 Exotropia Image Set

Left Eye			Right Eye		
Experiment	Reference	Correction	Experiment	Reference	Correction
12	15	✓	-68	60	
106	0		59	60	✓
119	30		106	>90	✓
-10	0	✓	27	30	✓
12	30		6	0	✓
49	45	✓	3	0	✓
112	>90	✓	0	0	✓
-14	15		7	0	✓
100	90	✓	13	15	✓
35	30	✓	47	45	✓

## A.3 Pseudostrabismus Image Set

Left Eye			Right Eye		
Experiment	Reference	Correction	Experiment	Reference	Correction
-5	0	✓	15	15	✓
Initialisation Fail			-	-	
0	0	✓	14	0	✓
13	0	✓	9	0	✓
11	0	✓	25	0	
35	0		35	0	
4	0	✓	9	0	✓

## A.4 Other Image Set

Left Eye			Right Eye		
Experiment	Reference	Correction	Experiment	Reference	Correction
32	30	✓	41	30	✓
9	15	✓	-17	-15	✓
-6	-15	✓	17	15	✓