Automatic Localization of Trackable Features in Retinal Fundus Images

by

Paul Inder

A Thesis submitted to
the Faculty of Graduate and Postdoctoral Affairs
in partial fulfillment of the requirements for the degree of
Master of Applied Science
in
Biomedical Engineering

Ottawa-Carleton Institute for Biomedical Engineering (OCIBME)
Department of Systems and Computer Engineering
Carleton University
Ottawa, Ontario, Canada
January 2014

Copyright ©
2014 - Paul Inder
Abstract

A surgical technique known as laser photocoagulation is used to reduce vision loss due to degenerative eye conditions. This treatment is performed by guiding a treatment laser over diseased tissue on the retina. However, due to natural motions of the eye healthy retinal tissue can be damaged by the laser. This makes it necessary to either 1) reduce these motions using anesthesia or physical restraints, or 2) compensate for these motions using sensors and a computer.

The retinal surface can be imaged directly and noninvasively. A scanning laser ophthalmoscope (SLO) is an imaging instrument that can produce high resolution video images of the retina. The Automated Laser Eye Treatment project aim is to integrate motion analysis from SLO video with the treatment laser. Motion analysis from the SLO video provides information suitable for automatic laser aiming.

Motion estimation from SLO video can be achieved using landmark-based motion tracking. In the ALET implementation, landmarks are selected from a reference image by an expert operator. This thesis work automates the landmark selection process. The proposed automatic landmark localization approach has two steps: first, a set of candidate landmarks is found using a corner detection procedure on a vessel-enhanced retinal image; and second, the set of candidate landmarks are ranked in decreasing order of their estimated tracking integrity.
Acknowledgments

Firstly, I am forever grateful for the kind help and support from my supervisor Professor Victor Aitken.

Next, I would like to thank Professor James Green for his helpful discussions.

Last, but not least, I would like to thank the two most important people in my life: my mother for providing her unconditional love and support; and my loving partner, Stephanie, for standing by me and pushing me through all the difficult moments.
Table of Contents

Abstract i

Acknowledgments ii

Table of Contents iii

List of Figures v

1 Introduction 1
   1.1 The Retinal Fundus ........................................ 1
   1.2 Automated Laser Eye Treatment ............................. 3
   1.3 Landmark Tracking .......................................... 4
   1.4 Objective .................................................. 5
   1.5 Contributions ............................................... 5
   1.6 Organization of Thesis ...................................... 6

2 Background 7
   2.1 The ALET Project ............................................ 7
      2.1.1 The Simulated Image Sequence Generator ............... 8
      2.1.2 Offline Saccade Detection ................................ 9
      2.1.3 Landmark Tracking ...................................... 10
   2.2 Finding Vessel Bifurcations/Crossovers ....................... 12
# List of Figures

1. Anatomy of the eye ................................................. 1
2. Sample DRIVE fundus image ................................. 2
3. Example retinal vessel crossover and bifurcation ......... 5
4. Saccade detection using edge motion between images .... 10
5. Illustration of landmark tracking in ALET ................. 11
6. Example distortion matrix ........................................ 15
7. Bianco and Thompson's calculation of landmark reliability.. 16
8. Sample preprocessing results ..................................... 24
9. Chen, *et. al.*, sample Harris corner results ............... 25
10. Bifurcations and crossovers as corners ...................... 26
11. Harris corner detection results ................................. 29
12. Distortion matrix for landmark on linear edge ............. 31
13. Distortion matrix for landmark over uniform values ...... 32
14. Distortion matrix on highly textured region ............... 33
15. Distortion matrix for reasonably salient landmark ........ 34
16. Sample benchmark normalized distortion matrix ($\lambda = 3$) .... 35
17. Top scoring landmarks - Example 1 .......................... 38
18. Top scoring landmarks - Example 2 .......................... 39
19. Top scoring normalized distortion matrices - Example 1 .... 40
20. Top scoring normalized distortion matrices - Example 2 .... 41
<table>
<thead>
<tr>
<th>Section</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>Plot of p metric results a set of candidate landmarks</td>
<td>42</td>
</tr>
<tr>
<td>22</td>
<td>Bottom scoring normalized distortion matrices</td>
<td>43</td>
</tr>
<tr>
<td>23</td>
<td>Landmarks selected from first simulation frame</td>
<td>45</td>
</tr>
<tr>
<td>24</td>
<td>Motion profile for SLO simulation</td>
<td>45</td>
</tr>
<tr>
<td>25</td>
<td>Landmark tracking results</td>
<td>46</td>
</tr>
<tr>
<td>26</td>
<td>Comparison of motion estimation to true motion</td>
<td>48</td>
</tr>
<tr>
<td>27</td>
<td>Motion estimation errors</td>
<td>48</td>
</tr>
<tr>
<td>28</td>
<td>Rotational distortion during a saccade</td>
<td>49</td>
</tr>
<tr>
<td>29</td>
<td>Umeyama measurement of image rotation</td>
<td>50</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

1.1 The Retinal Fundus

The interior surface of the eye is known as the retinal fundus. This surface contains the retina, macula, and optic nerve head (a.k.a. optic disc). These structures are shown in the diagram of gross eye anatomy provided in Figure 1.

Figure 1: Anatomy of the eye. \(^a\)

\(^a\)http://www.blundelloptometry.com/images/eye_diagram.gif
The retinal fundus can be imaged directly and noninvasively. Imaging of the retinal fundus has clinical, commercial, and research applications. In the clinic, fundus images can be used to monitor the progress of pathologies such as diabetic retinopathy [1], and hypertensive changes in the retinal vessels can be used to indicate the onset of a stroke [2]. Commercial applications include biometric security [3]. Fundus imaging applications in research may be focused in the study of retinal manifestations of systemic diseases (e.g. systemic sclerosis in [4]), or the study of visual nerve pathways through targeted light delivery to rods and cones [5]. A sample wide-field, color retinal fundus image from the DRIVE retinal image dataset [6, 7] is provided in Figure 2.

![Sample DRIVE fundus image](image)

**Figure 2:** Sample DRIVE fundus image.

The circular bright region on the left side of the field of view is the optic disc. The macula is the dark region located in the middle, and the retinal vasculature is the branched structure extending throughout the image.
1.2 Automated Laser Eye Treatment

Amtech Aeronautical Ltd., in collaboration with the Canadian Space Agency, have worked on an Automated Laser Eye Treatment (ALET) system [8, 9]. The goal of the ALET project is to develop a system for safe, automated photocoagulation treatment of retinal lesions and microaneurysms. This type of treatment is used to reduce vision loss due to degenerative eye conditions, such as diabetic retinopathy, and retinal detachment.

During a laser treatment, there exists a significant risk of damaging healthy eye tissue due to the presence of natural eye motions. These eye motions are fast in comparison to human reaction times, so it is necessary to either 1) minimize the motions through the use of anesthetics or physical restraints, or 2) compensate for the motions through the use of a computer equipped motion-detecting sensors and software. Elimination of motion through anesthetic or physical restraint is not ideal - there exists a risk of adverse reaction to anesthetic, and physical restraints can be cumbersome and painful. The ALET project aims to compensate for eye motion by tracking the motion of several landmarks in a retinal video feed provided by a scanning laser ophthalmoscope. Scanning laser ophthalmoscopy (SLO) [10, 11] is a method of obtaining retinal fundus images with high spatial resolution. In this method, a focused laser beam enters the eye through the pupil and is moved over the retina by horizontal and vertical scanning mirrors. Light reflected off of the retinal surface exits the eye, and is directed into a CCD camera to be translated into a brightness value for the corresponding pixel in a video frame.
1.3 Landmark Tracking

Landmark tracking has been used to estimate motion in SLO image sequences [8, 9, 12, 13]. Landmarks are small, reliably identifiable regions within an image. A typical landmark tracking procedure is summarized as follows:

1. Select landmarks from an image captured during a moment of minimal motion.

2. Predict the locations of landmarks in the next image using a motion estimation model.

3. Find original landmarks in the new image.

4. Use the new landmark positions to compute an estimate of retinal motion. Update the motion estimation model using the current landmark position measurements.

5. Repeat steps 2 through 4 until the end of the image sequence.

In the ALET project implementations [8,9,12,13], the tracking algorithm proceeds after the operator manually selects the pixel location of 4-6 landmarks. This is the only human intervention necessary for the algorithm to proceed. Landmark positions in the next frame are predicted by a motion estimation model; estimated locations of each landmark are found using a correlation-based search over a small search window around the predicted position. The positions of landmarks in the new frame are compared to the positions of landmarks in the reference frame in order to compute the transformation parameters (rotation, translation, scaling) between the two point sets.

It has been observed that points of retinal vessel bifurcation or intersection (crossover) work well as landmarks for retinal image registration and tracking procedures [14]. See Figure 3 for an example of these vessel structures.
Figure 3: An example of retinal vessel crossover (left) and bifurcation (right).

Automatic selection of landmarks from a reference image would eliminate the need for expert human intervention in the laser eye repair process.

1.4 Objective

The goal of this work is to implement an automatic landmark detection procedure to work with the previously implemented ALET landmark tracking system. This landmark detection procedure will operate by using a reference retinal fundus image as input to generate a list of landmark locations to be used in the tracking algorithm.

In the current landmark tracking implementation, only 4 to 6 landmarks are required. Therefore, the aim will be to rank the list of candidate landmarks in order of descending tracking integrity. From this ordered list, the $n$ best landmarks can be selected for use in the tracking system.

1.5 Contributions

The contributions of this work are as follows:

1. A method for automatic landmark localization from a static retinal image is
1. Formulated.

2. Verifying corner detection as a suitable method for finding vessel bifurcation and crossover points in retinal images.

3. Methodology for qualitative and quantitative assessment of a candidate landmark's tracking integrity is developed.

4. Legacy ALET project code has been modernized for use in future work.

1.6 Organization of Thesis

This remainder of this thesis is organized as follows:

- Chapter 2 presents a brief summary of the ALET project. Focus is given to the simulation technique developed in the ALET project, which is used in testing the landmark localization procedure developed in this work. Following this, a survey of previous literature and theory related to this thesis is presented.

- In Chapter 3, the procedure for detecting landmarks suitable for tracking is presented. The procedure is first described in brief, and an in-depth explanation follows.

- Chapter 4 contains an overview of the results of the landmark detection procedure. These results are presented in two parts:
  1. Results of the candidate landmark ranking procedure are examined.
  2. Using the ALET software simulation, the tracking accuracy of landmarks found using the method developed in this work are examined.

- Chapter 5 provides a summary of the work and results presented in this thesis, along with proposals for future work.
Chapter 2

Background

The aim of this chapter is twofold. First, a brief overview of the ALET project is presented in order to help the reader understand where this thesis work fits within the existing ALET system. Second, a discussion on landmark localization and assessment is provided. Given that retinal vessel bifurcations and crossovers have been empirically recognized as good landmarks for tracking, existing techniques for localizing these points are considered. This is followed by an examination of existing methods for tracking integrity assessment.

2.1 The ALET Project

The ALET project aims to provide an automated system for safe photocoagulation treatment of retinal lesions and microaneurysms [8,9]. The safety component comes in the form of compensation for retinal motion so that healthy eye tissue is not damaged by the treatment laser. This compensation has two elements. The first is a routine for detecting the onset of fast, large amplitude eye motions called saccades. Upon detecting a saccade event, the ALET system will signal the laser to shut off until the eye has returned to low-motion state. The second element of safely compensating for retinal motion is to follow the low-amplitude drift motion using landmark tracking.
Motion estimates computed using the landmark tracking procedure can be used to accurately aim the treatment laser at the desired treatment location. Both of these elements (saccade detection and landmark tracking) are performed using the video feed from a scanning laser ophthalmoscope.

A software simulation procedure was developed during the ALET project to provide a means for evaluating the performance of the motion detection and tracking system. Under simulation conditions the true positions of each landmark in an image can be stored, and this data allows for the calculation of a tracking error. Although testing on real SLO image sequences is important, information about the true eye motion is not available in this case.

What follows is a description of the ALET simulation procedure, which has three separate phases. First, a single retinal image is used in an image sequence generator to create a simulated SLO video feed. Second, the simulated SLO video feed is examined frame-by-frame in the saccade detection routine. Third, a landmark tracking routine is used to estimate motion in the simulated SLO image sequence.

2.1.1 The Simulated Image Sequence Generator

A simulated image sequence generator was developed during the ALET project in order to provide a means for testing the motion detection and estimation algorithms. The approach to image sequence generation is based on information about retinal motion available in the literature, and observations made while viewing real SLO image sequences. The image sequence is generated by sampling a high resolution, wide-angle retinal image - referred to as the base retinal map. In this implementation, simulated images are of size 256×256. Pseudo-random motion parameters are defined, which mimic both the natural saccadic and drift motions observed in real sequences.

An image sequence is generated by simulating the laser scanning pattern used in SLO instruments. A simulated real-world clock defines the speed in which the
'laser' is scanned over the base retinal map. Pixel sampling occurs at regular time intervals, and the four pixels nearest to the laser position in the base retinal map are averaged to form the resulting pixel in the simulated image. Pseudo-random motion simultaneously occurs throughout the simulated scanning procedure, which distorts the output images as observed in real SLO image sequences. This type of distortion is discussed in Chapter 4.

2.1.2 Offline Saccade Detection

A saccade is a high speed, large amplitude motion of the eye. In the ALET project, a saccade reflex detection procedure was developed to detect the earliest sign of the onset of this type of motion. This procedure can also provide the horizontal landmark motion predictions to be used in the landmark tracking algorithm.

The saccade detection procedure is designed to operate in harmony with the laser scanning of an SLO instrument. As the laser scans over the retinal surface, individual pixels are detected for image construction. When a horizontal line of a SLO image is completed (all pixels in that row have been detected), an edge mask is used to compute the positions of edges within the row. A strong positive response to the edge mask is a left edge, and a strong negative response to the edge mask is a right edge. The positions of the strongest left and right edges in row \( r \) of image \( n \) are compared with the edge positions in row \( r \) of image \( n + 1 \) (see Figure 4). If a match is found, an estimate of horizontal motion is computed for this row. If the horizontal motion exceeds a threshold, the system will signal the onset of a saccade. During a laser treatment session, this would shut down the treatment laser until the eye returns to a low-motion state. Note that there is no estimation of motion in the vertical direction during the saccade detection procedure. Saccade motion is strongest in the horizontal direction.

The simulated images used in this implementation have 256 rows. In the ALET
Figure 4: Saccade detection using edge motion between images. Reprinted, with permission, from [8,9].

project, each SLO image is divided into 8 image bands - each band containing 32 image rows. Horizontal motion estimates for each row inside a band are averaged to form a single band motion estimate. These horizontal motion estimates are used in the landmark tracking system to predict landmark motion. Note that this method does not provide an estimate of vertical motion.

2.1.3 Landmark Tracking

The ALET landmark tracking implementation requires an image sequence and corresponding band motion estimates as input. Landmarks are selected from the first image. In the original ALET implementation, this landmark selection step was done manually - an expert user visually inspected the first image and selected the pixel
Figure 5: The ALET landmark tracking algorithm uses horizontal motion estimates to locate the landmark in a subsequent image. (Reprinted, with permission, from [8,9])

locations of 4-6 landmarks. This thesis work has automated the landmark selection process.

Following the selection of landmarks, the landmark position is located in each subsequent frame. Previous landmark positions are used to decide whether a landmark is located within the current band. The horizontal band motion estimate provided by the saccade detector is combined with the previous landmark position to derive an estimate of the position of the landmark in image $i+1$ (see Figure 5). The landmark window position is defined as the coordinate of the middle pixel of the landmark window. A search window is defined surrounding the predicted landmark position. The size of the search window is large enough to ensure that the true position of the landmark window is within the boundary of the search window. The new landmark
position is found using a direct, correlation-based search. The optimal estimate of the landmark position is found by calculating a similarity measure between the pixel values of the landmark window, and same-sized regions within the search window. The estimated position will be at the pixel location of the minimum (or maximum) of all similarity measure results. In the existing ALET system, a sum of absolute differences is used as a similarity measure. This similarity measure is discussed in sections to follow.

Once the new landmark positions have been found, an estimate of retinal motion is computed by finding the transformation parameters between the new landmark locations, and the original landmark locations. Umeyama provides a least-squares optimization method for computing the rotation, translation, and scaling parameters between the two sets of landmark locations [15]. In this simulation, a simplified Umeyama method is used, where the scaling factor remains locked at 1. Therefore, for each frame in the image sequence, comparison of new and original landmark positions provide translation and rotation parameters with respect to the reference image.

### 2.2 Finding Vessel Bifurcations/Crossovers

Vessel bifurcation and crossover points have been used in retinal image registration procedures. Published methods for finding retinal vessel bifurcation/crossover points can be grouped into one of two general strategies: 1) finding points from a direct analysis of the retinal image (e.g. [16–18]); or 2) by first extracting the vascular tree and searching the resulting binary image (e.g. [19–21]).

For example, in [18], Hu, et al., describe a method for finding vessel bifurcation- 

crossovers directly from a retinal image. In their method, a set of retinal images are examined and all the bifurcations/crossovers are labeled by a human expert. Using
the annotated images, a large set of image filters are generated. These image filters are convolved with each training image - and the results of convolution at each labeled bifurcation/crossover pixel (in addition to some non-bifurcation/crossover pixels) are collected to be used in a supervised classifier. In order to find the locations of vessel bifurcations/crossovers in a new image, each of the image filters are convolved with the new image. Using the set of convolution results at every pixel of the new image, each pixel is compared to the training set in a k-Nearest Neighbours fashion. In this way, each pixel is classified as either bifurcation/crossover/neither.

Bhuiyan, et al. [21] find vessel bifurcations after first segmenting the vascular tree from the retinal image. The segmentation is performed by classifying each image pixel as vessel or non-vessel. The binary segmented vessel image can be reduced to the vessel center lines. Finally, for each pixel in the vessel center line image, an examination of the 8 pixel neighbourhood around each pixel can be used to indicate whether a pixel is a bifurcation or crossover point. A bifurcation will be represented by a Y-shaped pixel structure, and a crossover will be represented by a perpendicular cross shape.

A common theme among bifurcation/crossover-finding procedures is the use of supervised classification. These strategies require a significant amount of manual image labeling, and are not guaranteed to generalize well beyond the images used to develop the method. Additionally, many of these methods focus on accuracy. They aim to find all vessel bifurcation/crossover points - sometimes distinguishing between the two types of points, with minimal error.
2.3 Image Landmarks and Assessment of Tracking Integrity

An image landmark is a subimage (a small portion of an image) which is useful for image registration, recognition, and tracking. Sebastian Thrun [22] describes a good trackable feature as being 1) stationary, 2) reliably recognizable, and 3) sufficiently unique. In line with this view, this thesis defines a good trackable landmark as a window having a reasonable degree of salience with respect to the neighbourhood surrounding the landmark window. A landmark which meets this criteria is said to have good tracking integrity. High tracking integrity is a requirement because we must have confidence that the precise location of the landmark can be reliably found in subsequent views of the scene.

The concept of tracking integrity, in an abstract sense, describes how we expect a landmark will perform when used for tracking. If a landmark has poor tracking integrity, we expect that it may not be easily localized in future images, and the tracking algorithm will not succeed. Tracking accuracy would be of the utmost importance in, for example, a targeted delivery of light to a single retinal photoreceptor. We need to consider how to quantify tracking integrity so that we can choose the most reliable landmarks for tracking.

Existing literature on measuring tracking integrity is sparse. A small group of papers from the field of robot localization using computer vision provide some insight. Thompson et al. [23, 24], building on the concepts put forth by Mori et al. [25], and Bianco and Zelinsky [26], describe the concepts of dynamic and static tracking integrity 1. Static tracking integrity refers to a measurement of tracking integrity calculated using a single image. Dynamic tracking integrity refers to a measurement of the change in static tracking integrity over a sequence of images.

1the authors use the term reliability in place of tracking integrity, but they refer to the same idea.
Consider a $M \times N$ pixel landmark $T$ located at the center of a search window $S$. Mori et al. [25], and Bianco et al. [26] demonstrated a technique for landmark localization known as the \textit{valley method}. In this method, the landmark $T$ is correlated with all possible regions of the \textit{same size} in search window $S$. At each $M \times N$ pixel region $R$ in $S$, where $T$ overlaps $R$ and does not extend beyond the boundary of $S$, the sum of absolute differences of the corresponding pixel values is computed:

$$SAD = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} |R_{i,j} - T_{i,j}|.$$ \hspace{1cm} (1)

Each $SAD$ value represents a measure of how closely the pixel values of landmark $T$ and region $R$ correspond to one another. Computing a SAD value for all valid regions in $S$ results in a \textit{distortion matrix} (or correlation matrix; see Figure 6). Take note that the distortion matrices presented in this work are individually normalized relative to the maximum $SAD$ value.
Ideally, the distortion matrix will have a global minimum at the best estimated position of a landmark. This minimum corresponds to the pixel in the search window which corresponds to the location of the landmark. If the landmark and search window are from the same image, this minimum will be zero (a perfect match). If a landmark is correlated with a search window in a different image, this minimum may not be zero.

The tracking integrity metric used by Bianco and Thompson is as follows:

\[ r = 1 - \frac{g}{g'} \]  

(2)

where \( r \) is the reliability, \( g \) is the value at the global minima of the distortion matrix, and \( g' \) is the minimum distortion value from the sixteen \( SAD \) values in the extended neighbourhood, as shown in Figure 7. The \( r \) value ranges between 0 and 1, and characterizes the depth of the global minimum of the distortion matrix.

There are two issues with the use of this metric for our work. First, this metric is not useful for determining the tracking integrity of a landmark using a single image because the distortion value at the global minimum \( g \) will always be zero - which,
in turn, means that all landmarks will have an $r$-value of 1. Second, the $g/g'$ term provides a very limited local description of the shape of the distortion matrix. For example, there may exist more than one minima within a search window around the landmark location. In this case, the tracking algorithm may incorrectly localize a landmark at another minima - despite the landmark having a high reliability value. A good tracking integrity metric should account for the distortion values outside of the small neighbourhood around the global minimum.

In the chapter to follow, a landmark localization procedure is developed which requires a single image for distortion matrix analysis, and uses the entire distortion matrix in assessing tracking integrity.

### 2.4 Data and Software

To conclude this background chapter, a brief note on the data and software used in this work is given.

**Data**

Several sets of retinal images have been made publically available in order to promote the development of retinal image analysis methods. Two popular image sets are the DRIVE [6, 7], and STARE [27, 28] databases.

In this work, the DRIVE dataset was used for developing the landmark-finding algorithm. This set of images was originally established to enable comparative studies on segmentation of blood vessels in retinal images.
Software

Existing ALET code was written in MATLAB. Most of this code was developed nearly 15 years ago. This legacy code was updated to work in a modern version of MATLAB (R2012a). This updated and modified code was used to perform the validation simulations presented in Chapter 4.

The landmark detection procedure was developed using the Python programming language. The Python standard library does not have built-in scientific computing capabilities. The ability to perform scientific computing in Python is made possible through the use of the base NumPy (numerical python) package. NumPy provides support for transformations of $N$-dimensional arrays, along with a set of standard mathematical functions. The real power of using Python for scientific computing comes in the form of the many packages built on top of the foundation provided by NumPy. Among these packages are:

SciPy A collection of modules for scientific computation. Some example modules:

- integration, interpolation, optimization, signal processing, multidimensional image processing.

matplotlib A powerful plotting library. The syntax is similar to that of Matlab.

scikit-image This image processing and analysis package fills the gaps in the small collection of image processing routines provided by SciPy.

scikit-learn This package provides an extensive library of machine learning routines.

All of these packages are free (as in freedom and as in beer), and open source.
Chapter 3

Methodology

This chapter reveals the automatic landmark detection strategy developed in this thesis. This strategy employs an image preprocessing step designed to 'normalize' retinal images before being further analyzed. The preprocessing step effectively enhances the appearance of the retinal vessels. Following this image enhancement step, candidate landmarks are located through the use of a Harris corner detector. Finally, candidate landmarks are ranked in decreasing order of tracking integrity by comparison to an ideal tracking integrity template.

3.1 Preprocessing

Every retinal image will possess different degrees of vessel contrast, image noise, and light distribution. If our first objective is to find vessel bifurcations and crossings, the preprocessing procedure should enhance the appearance (contrast) of vessels, in addition to reducing noise and improving the light distribution. Such a procedure effectively 'normalizes' the images being put into subsequent steps in the landmark detection procedure, which will allow for consistent operation on images of different quality, using the same parameters in the corner detector.

Retinal image analysis procedures typically examine only the green channel from
color (RGB) fundus images. The green channel contains the most vessel contrast, while the red channel is saturated and the blue channel contains very little information. Walter et al. [29] suggests the reason for this is, in part, due to hemoglobin having a high absorption coefficient in the green region of the visible spectrum. Blood vessels contain hemoglobin, and absorb much more green light than the surrounding retinal tissue. In the procedure which follows, this convention is observed and only the green channel from each retinal image is used.

Vessel-enhancement procedures are commonly used in retinal vessel segmentation procedures. In this work, we make use of the enhancement techniques described in [30–32]. The procedure has four steps: 1) noise reduction 2) background subtraction, 3) contrast scaling, and 4) vessel enhancement.

### 3.1.1 Noise Reduction

A simple noise reduction procedure is used. First, high frequency noise is removed by using a $3 \times 3$ mean kernel as a low-pass filter. Following this, a smoothed image $I_\gamma(x, y)$ is produced through convolving the previously filtered image with a $9 \times 9$ Gaussian kernel ($\mu = 0, \sigma^2 = 1.8$).

$$I_\gamma(x, y) = (I(x, y) \ast \frac{1}{9} I_3) \ast N_{9 \times 9}(0, 1.8)$$  \hspace{1cm} (3)

where $I_3$ is the $3 \times 3$ identity matrix, and $N_{m \times m}(\mu, \sigma^2)$ is an $m \times m$ Gaussian kernel. The $\ast$ symbol denotes a convolution operation.

### 3.1.2 Background Subtraction

Images of the retina often possess gradients and irregularities in the distribution of light on the surface of the retina. Background subtraction is intended to homogenize the light distribution throughout the image. This is performed by computing an
image containing only the estimated background of the retinal image \((I_B(x, y))\), and subtracting this estimate from the noise-reduced image:

\[
I_{BS}(x, y) = I_\gamma(x, y) - I_B(x, y)
\]

(4)

An estimate of the background \((I_B(x, y))\) can be obtained by filtering \(I_\gamma(x, y)\) with a large mean kernel. The size of this mean kernel depends on the size and/or scale of the image, and may require some experimentation. In the case of images from the DRIVE dataset, experimentation in this thesis found that a \(31 \times 31\) pixel mean kernel provided good background estimates of the images having size \(565 \times 584\). The resulting array \(I_{BS}\) contains the background-subtracted image.

### 3.1.3 Contrast Scaling

Contrast between vessels and the background can be enhanced by scaling the pixel values of \(I_{BS}\) to cover the entire range of possible grayscale values \([0 \rightarrow 255]\). This is achieved using the following pixel transformation:

\[
g_{\text{out}} = \begin{cases} 
0, & \text{if } g_{\text{in}} < 0 \\
255, & \text{if } g_{\text{in}} > 255 \\
g_{\text{in}} + 128 - g_{\text{in}\rightarrow\text{max}}, & \text{otherwise}
\end{cases}
\]

(5)

where \(g_{\text{in}}\) corresponds to the pixel values of \(I_{BS}\), and \(g_{\text{in}\rightarrow\text{max}}\) is the greyscale value corresponding to the peak in the histogram of greyscale values occurring in \(I_{BS}\). The resulting image is \(I_{CS}\).
3.1.4 Vessel Enhancement

The final preprocessing step is to generate a vessel enhanced image $I_{VE}$. A top-hat transform is an image processing operation that enhances small elements and details in an image. Two types of top-hat transforms exist; the white top-hat transform enhances the appearance of objects that are brighter than their surroundings, and the black top-hat transform enhances elements that are darker than their surroundings.

The white top-hat transform on image $X$ is given by:

$$T_w(X) = X - X \circ b \quad (6)$$

where $\circ$ denotes an image opening operation, and $b$ is a structuring element. The white top-hat transform returns an image containing objects from the input image that are both smaller than the structuring element and brighter than their surroundings in the original image. Similarly, the black top-hat operation on image $X$ is given by:

$$T_b(X) = X \bullet b - X \quad (7)$$

where $\bullet$ denotes an image closing operation, and $b$ is a structuring element. The black top-hat transform returns an image containing objects from the input image that are both smaller than the structuring element and darker than their surroundings in the original image. More on these morphological image transformations (top-hat, opening, closing, etc.) can be found in [33, 34].

The vessel enhanced image $I_{VE}$ is produced by performing a white top-hat operation on $I_{CS}^c$, the image complement of $I_{CS}$. An image complement is computed by subtracting an image from the maximum pixel value. For an 8-bit image, $I_{CS}^c = 255 - I_{CS}$.

$$I_{VE} = I_{CS}^c - I_{CS}^c \circ disk \quad (8)$$
disk(8) is a disk structuring element with a radius of 8 pixels. The pixel diameter of this structuring element must be larger than the diameter of the largest vessels in the image. The purpose of using the image complement is to change the vessels from being dark with respect to their surroundings, to being bright (whiter) than their surroundings.

3.1.5 Sample Preprocessing Images

Figure 8 contains a set of images depicting the effects of the above preprocessing operations. These images were chosen to illustrate the performance of the preprocessing procedure on images with different levels of illumination and pathology.

The top row of Figure 8 are the unmanipulated green channel retinal images. The left-top image contains no pathology, and is well illuminated. The centre-top image contains no pathology but has darker illumination. The retina in the right-top image is showing signs of diabetic retinopathy and non-ideal illumination. Pathologies such as diabetic retinopathy can occlude the view of the vascular structure.

The processed images in the last row illustrate that the preprocessing algorithm can handle different illumination conditions (the quality of the results of image left and centre columns are similar). However, the presence of pathologies in the eye can create problems - there is a noticeable degree of 'noise' present in the rightmost image of the bottom row in Figure 8.

3.2 Finding Candidate Trackable Features

The first step in locating features for tracking is to obtain a set of candidate features. We obtain a set of candidate features through a procedure designed to locate vessel bifurcation and crossover points. Typically, these points appear in retinal images as a strong edge possessing a high degree of curvature. This description is very similar
**Figure 8:** Preprocessing steps. Top row: Unmanipulated green-channel images. Second row: background-subtracted images. Third row: contrast-scaled images. Bottom row: vessel-enhanced images.
to the definition of an image corner. Therefore, corner detection was investigated as a method for finding vessel bifurcation and crossover points.

There are a few examples of corner detection being applied to retinal images. Chen, et. al., have used a Harris corner detector to find control points for image registration [35,36]. Figure 9 provides a result from [35], demonstrating their use of a Harris corner detector on a retinal image. The authors claim their result to be more useful than the central line extraction method points because they find a large number of points, uniformly distributed throughout the image. However, close examination of their 'corner points' reveals that many of the bifurcations remain unlabeled, and the majority of points labeled as corners are simply due to noise and/or texture in the image.

Corner detection was further investigated in this thesis as an efficient method for locating retinal vessel bifurcations and crossovers. Figure 10 contains the result of four corner detection algorithms on a preprocessed DRIVE image. These results show that preprocessing of the retinal image increases the quality of corners obtained using corner detection. The (B) Kitchen-Rosenfeld [37], and (C) Moravec [38] results are not
Figure 10: Comparison of corner detection algorithms. Each image is labeled with the 30 strongest corner responses. A - Harris corners; B - Kitchen-Rosenfeld; C - Moravec; D - Shi-Tomasi.

satisfactory. Although these corner detection methods find a few bifurcation/crossover points, many of the points labeled as strong 'corners' are neither. On the other hand, the (A) Harris [39], and (D) Shi-Tomasi [40] corner detection results a show very good selection of vessel bifurcations and crossovers. Not only are the majority of the strongest corners detected by Harris and Shi-Tomasi located at a vessel bifurcation or crossover points, there are few mislabeled points. It is observed that the corner
points found using the Harris method are not uniformly distributed throughout the image, as was observed by Chen in [35,36].

It is not surprising that the Harris and Shi-Tomasi results are similar, as these methods are closely related. Upon comparing several Harris and Shi-Tomasi results, it was observed in this thesis work that the Harris method had more consistently selected bifurcation/crossover points with the fewest mislabeled corners. Additionally, the Shi-Tomasi method is more computationally intensive due to the large number of eigenvalue decompositions involved. Therefore, the Harris corner detector is selected to provide candidate landmarks.

Each corner detection algorithm defines a unique measure of cornerness. The magnitude of a cornerness value defines the strength or sharpness of a corner. A cornerness value is computed at each pixel in image \( I_{VE} \). The image/array of cornerness values is called a feature image. The Harris method [39] begins with calculating a Hessian matrix \( H \) for image \( I_{VE} \):

\[
H = \begin{bmatrix}
G_x^2 & G_x G_y \\
G_y G_x & G_y^2 \\
\end{bmatrix}
\]  

(9)

where, \( G_x \) is the partial derivative of image \( I_{VE} \) in the \( x \)-direction, and \( G_y \) is the partial derivative of image \( I_{VE} \) in the \( y \)-direction. These partial derivatives are computed through the convolution of image \( I_{VE} \) with the appropriate Sobel gradient operator:

\[
G_x = I_{VE} \ast \begin{bmatrix}
1 & 0 & -1 \\
2 & 0 & -2 \\
1 & 0 & -1 \\
\end{bmatrix}
\]

(10)
Here, $\ast$ denotes the 2D convolution operator. The Hessian matrix describes the local curvature at each pixel. By convolving the Hessian matrix with a small 2D Gaussian kernel $h$, a weighted average of local curvature surrounding each pixel is obtained. This result is known as the structure tensor $M$:

$$M = H \ast h$$

where $\ast$ is the 2D convolution operator. The Harris feature image $F$ is found in a final transformation:

$$F = \det(M) - \kappa \text{tr}^2(M)$$

where $\det$ is a determinant operation, $\text{tr}$ is a trace operation, and $\kappa$ is a parameter constant (usually $\kappa = 0.04 \sim 0.06$). The feature image has the same dimensions as image $I_{VE}$. The value of each element in $F$ corresponds to the cornerness of the corresponding pixel in $I_{VE}$. A higher cornerness indicates a high degree of local curvature found at that pixel. The preprocessed image is used as input to the corner detection algorithm. A peak-finding procedure (peak_local_max in the scikit-image Python library) is used to find the local maxima in $F$. The number of corner points obtained from the feature image can be controlled by using a relative cornerness threshold. A relative threshold can be adaptively set so as to consistently provide a minimum number of corners (30).

Figure 11 contains a sample of the results obtained using the Harris corner detector. These results illustrate that the combination of preprocessing and Harris
Figure 11: Harris corner detection under different light and pathology conditions.

corner detection outlined above can find vessel bifurcations and crossovers in images of different lighting conditions, and also in the presence of pathology.

In general, the existing methods for locating vessel bifurcation and crossover points mentioned in Chapter 2 use supervised learning, and other complex procedures. In my opinion, the above corner detection procedure is more efficient and generalizable, without a major loss in empirical accuracy or performance.
3.3 Assessment of Tracking Integrity

One objective of this thesis work is to sort the candidate landmarks in order of decreasing tracking integrity. Here it is proposed that the use of an ideal distortion matrix template can provide a metric to compare candidate landmarks: the similarity between the distortion matrix of each candidate landmark with a benchmark ideal distortion matrix provides a scalar quantity for comparison.

Before discussing this ranking technique, a better understanding of distortion matrices is required. Recall the introduction to distortion matrices presented in Section 2.3. The discussion to follow explores how distortion matrices can be used to infer the degree of a landmark’s salience and tracking integrity.

3.3.1 Understanding Tracking Integrity:

The Distortion Matrix

The overall shape of a distortion matrix surface provides a qualitative indicator of tracking integrity. Consider a landmark window over a straight line or edge in an image. Moving the window along this edge in the search window will produce similar SAD values (Fig 12). This appears as a ridge in the distortion matrix. If this landmark is used for tracking, there will be difficulty estimating true location of this landmark for each frame because there the minimum distortion value is not well-defined. Similarly, a landmark within an area containing nearly constant pixel intensities will produce similar SAD values when moving in any direction (Figure 13). Both of these example landmark windows would not be suitable for use as a trackable landmark because they have poorly defined minima in their distortion matrices (Figures 12 and 13). The linear edge window (Figure 12) is not salient with respect to nearby regions along the edge, the constant intensity window (Figure 13) is not salient with respect to nearby regions in all directions in the immediate neighbourhood. It is important
Figure 12: Distortion matrix surface when a landmark window is placed over a linear edge. SAD values are normalized to be between 0 and 1.

to remember that these distortion matrices are computed using the SAD values of a landmark and search window in the same image. In a real tracking situation, the objective is to find the landmark window in a search window from a new image. The retinal fundus in the new image may appear slightly different due to vessel pulsation, and other dynamic processes. In addition, image noise will further perturb the pixel values from frame to frame. For these reasons, pixels on straight edges or uniform regions are not considered to be reliably recognizable throughout a sequence of images.

Consider a window over an area with a large degree of texture (noisy, no discernable structure). Moving the feature window in any direction (only $1 - 2$ pixels) will produce a window of intensity values which are significantly dissimilar from the window located at the true position of the landmark (Figure 14). There is a high
Figure 13: Distortion matrix surface when a landmark window is placed over an area with (nearly) uniform pixel values. SAD values are normalized to be between 0 and 1.

degree of salience for this landmark window in the static image. However, if the texture observed is not a true part of the image - that is, the texture is mainly a noise artifact involved in the imaging process - the landmark position can change. The true position of the landmark can move several pixels between image frames, making landmarks over highly texturized areas unreliable for tracking.

If the degree of salience of a landmark with respect to the surrounding neighbourhood is thought of as a spectrum from low salience to high salience, the above examples are on the two extremes. A good trackable landmark window will be reasonably salient - somewhere in between the above examples. We must be confident that image noise and dynamic processes in the scene will not affect the ability to find the landmark in other images of the scene.

Now consider a window over a strong edge containing a high degree of curvature
Figure 14: Distortion matrix surface when a landmark window is placed over a highly textured region. SAD values are normalized to be between 0 and 1.

within the landmark window. A considerable difference in the similarity of intensity values is observed when the window is moved in any direction (Figure 15). However, the change in similarity is not as sharp as observed in the case of a highly textured/noisy region. The drop in similarity is gradual. This gradual drop in similarity exemplifies our definition of reasonable salience.

3.3.2 Benchmark Distortion Matrix

The above discussion on distortion matrices provides a qualitative intuition for the shape of a distortion matrix belonging to a landmark with good tracking integrity. This shape can be described in the following way: a single minimum (zero), which pseudo-exponentially increases isotropically up to one. The following function was
Figure 15: Distortion matrix of landmark with reasonable salience. SAD values are normalized to be between 0 and 1.

designed to have these properties:

\[
T = 1 - e^{-\sqrt{\frac{(x-x_0)^2 + (y-y_0)^2}{2\lambda}}}
\] (14)

where \( \lambda \) is a scaling factor which controls the rate of increase. Figure 16 provides a sample benchmark distortion matrix. The symmetry of the benchmark distortion matrix makes the designed metric independent of vessel structure. Selecting a value for \( \lambda \) depends on the size of the distortion matrix - which is determined by the landmark and search window sizes. The \( \lambda \) value tunes how quickly the similarity measure should increase moving away from the minima. In Figure 16, the normalized similarity measure is 0 at the minima, and rises to approximately 0.5 at a 3-pixel radius.
3.3.3 Template Comparison

For each candidate landmark location - i.e. the pixel location of corners detected in the processed image - a distortion matrix is computed using a landmark window and search window from the original unprocessed green channel retinal image. A single benchmark distortion matrix is calculated, and has the same dimensions as the distortion matrices for each candidate landmark. A scalar distance $p$ (similarity) between the distortion matrix $D$ of each candidate landmark and the benchmark distortion matrix $T$ described above is computed. The similarity measure used is the sum of absolute differences:

$$p = \sum_i \sum_j |T_{ij} - D_{ij}|. \quad (15)$$

A low $p$ value indicates a close match between the landmark distortion matrix and the benchmark distortion matrix - from which we can infer that the landmark has good tracking integrity. A large $p$ value suggests that the candidate landmark distortion
matrix is far from our formulated ideal tracking integrity.

A $p$ value is computed for each candidate landmark. The resulting list of $p$ values is sorted from smallest to largest. The same sorting order is applied to the list of candidate landmark pixel coordinates. Landmarks with the highest similarity to the ideal distortion matrix template are found at the top of this list, and are said to have the highest estimated tracking integrity.
Chapter 4

Results

This objective of this chapter is to address the following questions regarding the landmarks obtained using the procedure outlined in Chapter 3:

1. What do the top-ranked distortion matrices and landmarks look like?

2. How do these landmarks perform in a tracking simulation?

4.1 Top-ranked Distortion Matrices and Landmarks

The performance of the landmark detection procedure is to ultimately be evaluated using a measurement of tracking error over an image sequence. However, it is possible to visually verify whether the top-ranked landmarks found using the ranking method appear suitable for tracking. Figures 17 and 18 contain retinal images with labeled landmarks corresponding to the distortion matrices in Figures 19 and 20, respectively. It is observed that, in general, candidate landmarks located at the bifucation or crossing of thinner vessels are selected as good landmarks more frequently than candidate landmarks located near thick vessels.
Figure 17: Example retinal image 1 with top-scoring landmarks. Corresponding to the distortion matrices found in Figure 19.
Figure 18: Example retinal image 2 with top-scoring landmarks. Corresponding to the distortion matrices found in Figure 20.
Figure 19: Top-scoring normalized distortion matrices corresponding to landmarks labeled in Figure 17. SAD values on the z-axes range from 0 to 1.
Figure 20: Top-scoring normalized distortion matrices corresponding to landmarks labeled in Figure 18. SAD values on the z-axes range from 0 to 1.
The distortion matrices in Figures 19 and 20 show a visual similarity to the ideal distortion matrix template. Each of the top-ranked distortion matrices has a distinct global minima, with a steep (but not too sharp) increase in the distortion value moving away from the minimum.

Figure 21 shows the sorted $p$ values for all candidate landmarks found in the retinal image provided in (Figure 17). The lower $p$ values are considered to be good landmarks because their distortion matrices closely match the distortion matrix benchmark. The highest $p$ value candidate landmarks have distortion matrices very different from the ideal benchmark. Figure 22 provides the distortion matrices of the highest $p$ value candidate landmarks. These distortion matrices exhibit poorly defined minima, which would potentially make it difficult to localize the true landmark location in each image frame.

![Figure 21](image.png)

**Figure 21:** Plot of all $p$ metric values corresponding to the sorted candidate landmarks found in the retinal image from Figure 17.
Figure 22: Bottom-scoring normalized distortion matrices corresponding to the poorest landmarks found in Figure 17. SAD values on the z-axes range from 0 to 1.
4.2 Testing the Tracking Performance

Simulation is used to test the performance of automatically selected landmarks. Under simulation conditions the true position of landmarks in each image is known, and this data can be used to calculate a tracking error. Although testing on real SLO image sequences is important, they do not provide information about the true eye motion.

As described in Chapter 2, the ALET project provides a three step simulation procedure. Using this procedure, image sequences were generated using the SLO simulator. For each image, a list of simulated image pixel positions and their corresponding positions in the base retinal map is stored. Upon selecting the landmarks from the first simulated image, the location of each landmark in the base retinal map is obtained. For every subsequent frame, this original base retinal map landmark location corresponds to one of the pixels in the simulated image.

The automatic landmark selection procedure is programmed to select 6 top landmarks from the list of candidate landmarks ordered by decreasing tracking integrity. A top ranking landmark can be rejected if it is too close to the image boundary (in this case, a landmark is rejected if less than 20 pixels from the image boundary). This is to ensure that the landmark does not move outside the boundary of an image frame. Landmarks are then located in subsequent images.

Figure 23 displays the result of automatic landmark detection on the first frame of a simulated SLO image sequence. These landmarks are located on or near the optic disk - this image does not contain any suitable vessel bifurcations/crossings outside of this region.
Figure 23: Landmarks selected from the first image of a simulated sequence.

Figure 24: Horizontal (left) and vertical (right) motion profiles generated by the SLO simulator.

A simulated image sequence was generated having a true motion profile as depicted in Figure 24. The true motion is strictly horizontal and vertical translation, with no
Figure 25: The landmark tracking results. Each group of four subplots contains the following: horizontal detected (red) and true (blue) pixel locations (top left); horizontal pixel tracking error (top right); vertical detected (red) and true (blue) pixel locations (bottom left); vertical pixel tracking error (bottom right).
rotation or scaling. Multiple fast, high-amplitude saccadic motions (near-vertical lines) are visible in the true motion profiles. Additionally, slow, low-amplitude drift motion occurs in between saccades.

Under simulation conditions, the true position of each landmark is known for each frame. The tracking results for each of the six individual landmarks labeled in Figure 23 are shown in Figure 25.

Horizontal and vertical tracking error is calculated as the difference between the true and detected landmark pixel locations in each frame. Tracking error for each landmark has spikes in the image frames where a saccade begins to occur. In the majority of the saccades, only a single pixel of tracking error is observed, and the landmark tracking quickly locks onto the true location in subsequent frames. The vertical tracking for landmark 4 is off by \( \approx 4 \) pixels for approximately 15 frames, and then returns to zero error for the remainder of the simulation. This is not entirely surprising since, in this implementation, the saccade detector is only able to provide horizontal motion prediction and not a vertical motion prediction. A 4 pixel error for this many image frames may appear significant - however, it is the estimated positions of all landmarks that contribute to an overall motion estimation computed using the Umeyama method. The aggregate motion estimation of all 6 landmarks should, intuitively, have a smaller combined error. Figure 26 contains the overall horizontal and vertical translation motion estimation computed using the Umeyama method, compared to the true motion profile.

The overall motion estimation remains reasonably close to the true motion for the entire simulation. The image frames where landmark 4 displayed \( \approx 4 \) pixels of vertical tracking error appears as only \( \approx 1 \) pixel of error in the aggregate motion estimation. The overall horizontal and vertical tracking error is provided in Figure 27.

Tracking error is small, except for the frames coinciding with the onset of fast saccadic motions. When a saccade is occurring simultaneous to the detection of pixels
Figure 26: Horizontal (left) and vertical (right) motion estimation (red) and truth (blue).

Figure 27: Horizontal (left) and vertical (right) motion estimation error.

for the construction of an image, a rotational distortion is observed (Figure 28). What occurs is an apparent rotation of landmark positions, even though the true motion is strictly translational. Figure 28 provides two frames from the simulation where a rotational distortion is observed. It can be seen that the motion of landmarks between these two frames is not strictly translational.
Figure 28: Rotational distortion when a saccade occurs during the collection of pixel values for image construction. Frame 53 on the left, frame 54 on the right.

The Umeyama method computes the transformation parameters between two sets of points. These transformations include translation and rotation. The translation parameter results have been presented as landmark tracking estimates in Figure 25. Figure 29 provides the rotation computed for each frame in the simulation, relative to landmark points in the initial frame. Spikes in the rotation estimate coincide with the occurrence of saccades, and also with the frames where the estimation of translational motion has significant error.
Figure 29: Image rotation, as computed using the Umeyama procedure.
Chapter 5

Conclusions and Future Work

5.1 Conclusions

It is important to compensate for eye motion during laser photocoagulation treatment of the retina so that healthy eye tissue is not damaged. The Automated Laser Eye Treatment project addresses this issue by performing motion estimation using landmark tracking in a scanning laser ophthalmoscope video feed. This thesis work eliminates the requirement for the selection of landmarks from a reference image by an expert user. Instead, landmarks of high tracking integrity are automatically selected.

Candidate landmarks are found using a Harris corner detection procedure. This demonstration of corner detection as a means for locating retinal vessel bifurcation and crossover points is new to the field of retinal image analysis. The tracking integrity of candidate landmarks can be estimated through the comparison of the distortion matrix of a candidate landmark to the distortion matrix of an ideal template. Landmarks having the highest tracking integrity are selected for tracking.

Simulation of landmark-based motion estimation using landmarks found using this method show promising results. Tracking error is minimal during drift motion. During a saccadic motion, however, tracking error increases significantly for 1-2 image frames. After this sharp increase in error, minimal-error tracking resumes. The error
in tracking during a saccade may be acceptable in a real implementation of the ALET system. A feature of the ALET system is the detection of oncoming saccades. When a saccade has been detected, the treatment laser is signaled to shut off until the eye returns to a state of minimal motion. During this time, an error in landmark tracking is acceptable because there is no risk of damaging healthy tissue during this time. Once the saccade has ended, and good landmark tracking has resumed, the laser can resume treatment on the area of interest.

5.2 Future Work

Applications of candidate landmark ranking Selecting top-ranking landmarks by ranking the distortion matrices of a list of candidate landmarks may be useful in other applications. For example, estimation of vehicular ground speed has been performed using a downward facing camera [41]. This is typically performed using an optical flow method. It may be possible to develop a landmark-based speed estimation system, which detects good landmarks near the inward-flowing part of the image, and locates them before they exit the field of view.

Improving the landmark localization method In the ALET landmark localization procedure, a search window around a predicted landmark position is exhaustively examined for a match of the original landmark window. The efficiency of this procedure may be increased if, instead of a brute-force search, a gradient-descent procedure was used. The top-ranking landmarks selected using the method developed in this work typically have minima with surrounding slope/curvature that would work well in a gradient descent procedure.
Investigating Alternative Landmark Selection Schemes  Tracking accuracy could potentially be improved by increasing the number of landmarks used for tracking. Additionally, the distribution of landmarks throughout the image may have an affect on the accuracy of transformation parameter computation. A future study could involve measuring change in tracking accuracy by varying the number of landmarks, and their distribution throughout the image.
List of References


