ABSTRACT

SOLAR ACTIVITY DETECTION AND PREDICTION USING IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

by
Gang Fu

The objective of the research in this dissertation is to develop the methods for automatic detection and prediction of solar activities, including prominence eruptions, emerging flux regions and solar flares. Image processing and machine learning techniques are applied in this study. These methods can be used for automatic observation of solar activities and prediction of space weather that may have great influence on the near earth environment.

The research presented in this dissertation covers the following topics: i) automatic detection of prominence eruptions (PEs), ii) automatic detection of emerging flux regions (EFRs), and iii) automatic prediction of solar flares.

In detection of prominence eruptions, an automated method is developed by combining image processing and pattern recognition techniques. Consecutive Hα solar images are used as the input. The image processing techniques, including image transformation, segmentation and morphological operations are used to extract the limb objects and measure the associated properties. The pattern recognition techniques, such as Support Vector Machine (SVM), are applied to classify all the objects and generate a list of identified the PEs as the output.

In detection of emerging flux regions, an automatic detection method is developed by using multi-scale circular harmonic filters, Kalman filter and SVM. The method takes a sequence of consecutive Michelson Doppler Imager (MDI) magnetograms as the input. The multi-scale circular harmonic filters are applied to detect bipolar regions from the solar disk surface and these regions are traced by Kalman filter until their disappearance. Finally, a SVM classifier is applied to distinguish EFRs from the other regions based on statistical properties.

In solar flare prediction, it is modeled as a conditional density estimation (CDE) problem. A novel method is proposed to solve the CDE problem using kernel-based nonlinear regression and moment-based density function reconstruction techniques. This method involves two main steps. In the first step, kernel-based nonlinear regression techniques are applied to predict the conditional moments of the target variable, such as flare peak intensity or flare index. In the second step, the condition density function is reconstructed based on the estimated moments. The method is compared with the traditional double-kernel density estimator, and the experimental results show that it yields the comparable performance of the double-kernel density estimator. The most important merit of this new method is that it can handle high dimensional data effectively, while the double-kernel density estimator has confined to the bivariate case due to the difficulty of determining optimal bandwidths. The method can be used to predict the conditional density function of either flare peak intensity or flare index, which shows that our method is of practical significance in automated flare forecasting.

SOLAR ACTIVITY DETECTION AND PREDICTION USING IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

by
Gang Fu

A Dissertation
Submitted to the Faculty of
New Jersey Institute of Technology
in Partial Fulfillment of the Requirements for the Degree of
Doctor of Philosophy in Computer Science

Department of Computer Science

August 2007

APPROVAL PAGE

SOLAR ACTIVITY DETECTION AND PREDICTION USING IMAGE PROCESSING AND MACHINE LEARNING TECHNIQUES

Gang Fu

Dr. Frank Y. Shih, Dissertation Advisor — Date
Professor of Computer Science, NJIT

Dr. Haimin Wang, Dissertation Advisor — Date
Distinguished Professor of Physics, NJIT

Dr. Carsten Denker, Committee Member — Date
Assistant Professor of Physics, NJIT

Dr. Alexandros Gerbessiotis, Committee Member — Date
Associate Professor of Computer Science, NJIT

Dr. Qun Ma, Committee Member — Date
Assistant Professor of Computer Science, NJIT

BIOGRAPHICAL SKETCH

Author: Gang Fu

Degree: Doctor of Philosophy

Date: August 2007

Undergraduate and Graduate Education:

- Doctor of Philosophy in Computer Science,
  New Jersey Institute of Technology, Newark, NJ, 2007
- Master of Engineering in Computer Engineering,
  Soochow University, Suzhou, P. R. China, 2001
- Bachelor of Engineering in Computer Engineering,
  Soochow University, Suzhou, P. R. China, 1998

Major: Computer Science

Presentations and Publications:

Fu, G. Shih, F. Y. and Wang, H., "Solar flare prediction using kernel-based regression analysis method," in preparation for *Solar Physics*, 2007.

Fu, G., Shih, F. Y. and Wang, H., "Moment-based conditional density estimation using kernel nonlinear regression," in preparation, 2007.

Shih, F. Y. and Fu, G., "Decision combination of multiple classifiers," *Pattern ecognition and Artificial Intelligence*, accepted for publication, 2007.

Fu, G., Shih, F. Y. and Wang, H., "Automatic detection of emerging flux regions in consecutive Michelson Doppler Imager magnetograms," submitted to *IEEE Trans. on Image Processing*, 2007.

Fu, G., Shih, F. Y. and Wang, H., "Automatic detection of prominence eruption using consecutive solar images," *IEEE Trans. on CSVT*, vol. 17, no. 1, pp. 79-85, January 2007.

Fu, G., Jing, J., Song, H., Shih, F. Y. and Wang, H., "Solar flare prediction using kernel-based regression analysis," *AAS, SPD meeting,* Honolulu, HI, May 2007.

Fu, G., Shih, F. Y. and Wang, H., "Automatic detection of Emerging Flux Regions in consecutive Michelson Doppler Imager Magnetograms" *AAS, SPD meeting,* Honolulu, HI, May 2007.

Fu, G., Shih, F. Y. and Wang, H., "Solar image processing: detection of prominence eruptions," *2006 Computer Vision, Graphics, and Image Processing Conference,* Taiwan, August 2006.

Fu, G., Shih, F. Y. and Wang, H., "Automatic detection of emergence flux regions," *AAS, SPD meeting,* Durham, NH, June 2006.

Fu, G., Qu, M., Shih, F. Y. and Wang, H., "Automatic detection of prominence eruptions," *AAS, SPD meeting*, New Orleans, LA, May, 2005.

Thomasian, A., Fu, G. and Han, C., "Performance of two disk failure tolerant disk arrays," *IEEE Trans. on Computers*, vol. 56, no. 6, pp. 799-814, June 2007.

Thomasian, A., Fu, G. and Ng, S. W., "Analysis of rebuild processing in RAID5 disk arrays," *Computer Journal*, vol. 50, no, 2, pp. 217-231, March 2007.

Fu G. and Thomasian, A., "Anticipatory disk arm placement to reduce seek time," *Computer Systems: Science and Engineering*, vol. 21, no. 9, September 2006.

Thomasian, A., Han, C., Fu, G. and Liu, C., "A performance tool for RAID disk arrays," *Proc. Quantitative Evaluation of Systems* (QEST'04), pp. 8-17, Enschede, Netherlands, September 2004.

Fu, G., Thomasian, A., Han, C. and Ng, S. W. "Rebuild strategies for clustered redundant disk arrays," *Proc. Quantitative Evaluation of Systems* (SPECTS'04), pp. 598-607, San Jose, CA, July 2004.

Fu, G., Thomasian, A., Han, C. and Ng, S. W. "Rebuild strategies for clustered RAID," *Proc. Int'l Symp. on Performance Evaluation of Computer and Telecommunication Systems*, San Jose, CA, July 2004.

Fu, G., Thomasian, A., Han, C. and Ng, S. W., "Rebuild strategies for redundant disk arrays," *Conference on Mass Storage Systems and Technologies* (MSST'04), College Park, MD, April 2004.

Thomasian, A., Spirollari, J., Liu, C., Hun, C. and Fu, G., "Mirrored disk scheduling," Proc. Int'l Symp. Performance Evaluation of Computer and Telecommunication Systems'03 (SPECTS'03), Montreal, Canada, July 2003.

To my family,

for their endless love and support.

ACKNOWLEDGMENT

The tremendous effort and good will of many people have enabled completion of this dissertation. I am very thankful for all the help I have received throughout my PhD studies. I really appreciate that I have such a chance to express my appreciation and respect to these people.

First of all, I am full of gratitude to my dissertation advisors Dr. Frank Y. Shih and Dr. Haimin Wang for their great help, encouragement and guidance in all aspects of my research. I really appreciate that they offered me such a wonderful opportunity to study solar image processing, and no words could fully encompass the amount of gratitude I have for their supervision and support. Their broad knowledge, great expertise in research and kind personality has affected me deeply, which have benefited me greatly in the past and will be of important value in my future career. I really feel honored to be their student.

I am heartily grateful to Dr. Carsten Denker for his generous help and valuable advice in my research. I also thank him for his effort for serving as a committee member of my dissertation defense. His rigorous approach, attitude and conscientiousness in scientific research has impressed me deeply and benefit me continually in the future.

I would like to thank Dr. Alexandros Gerbessiotis, Dr. Qun Ma for their advice and help effort in my graduate study and dissertation defense.

I would like to thank my friends in this solar physics group. They are Jeongwoo Lee, Guo Yang, Yan Xu, Ming Qu, Ju Jing, Chang Liu, Deng Na, Jun Ma, Zhiwei Liu,

Hui Song, Changyi Tan, Angelo Verdoni, Samuel Tun and all other people in the group. I appreciate their help and enjoy the time with them.

Special thanks should be given to Christine A. Oertel, the administrative assistant in our research center, for assisting me in many ways so as to make my study easier.

Finally, and most importantly, I would like to thank my parents, Yaoliang Fu and Qianqian Wu, for their endless love, encouragement and support during my life, even in those hard days. I will never feel lonely, even I am far away from them. My gratitude to them can never be overstated.

The research work presented in this dissertation is supported by the National Science Foundation (NSF) under grants IIS 03-24816, ATM 05-48952, ATM 05-36921, and ATM 03-13591.

TABLE OF CONTENTS

TABLE OF CONTENTS
(Continued)

TABLE OF CONTENTS
(Continued)

Chapter **Page**

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES
(Continued)

LIST OF FIGURES
(Continued)

Figure **Page**

CHAPTER 1

INTRODUCTION

1.1 Overview

Space weather describes the geo-magnetic and particle conditions in the near-earth environment. Usually, space weather is of little concern in people's daily lives, but it has a great influence in many technological systems both in orbit and on the ground, such as satellites, ground-based power systems, oil pipelines, wireless communication systems, navigation systems, and also it affects human space exploration and development activities.

Most effects of space weather can ultimately be traced to solar activities, such as flares and coronal mass ejections (CMEs). A solar flare is a violent explosion in the chromosphere releasing up to a total energy of $6*10^{25}$ J. A coronal mass ejection is the result of a large-scale rearrangement of solar magnetic structure which leads to large amounts of ejected materials from solar corona (Wang 2005). These solar activities cause variations in both the solar electromagnetic radiation and the production of solar wind plasma, and energetic particles, which strongly influence space weather conditions in the near-earth space environment (Jing *et al.* 2004; Wang *et al.* 2004; Webb *et al.* 2000). Besides, solar flares and CMEs, other solar activities, such as filament (prominence) eruptions, sunspots, emerging flux regions and so on, are also of scientists' research interests.

The main purpose of this study is to develop automated tools that can detect and predict solar activities automatically by using image processing and machine learning techniques. With the great advance in solar observations, higher quality and large

quantity of solar images, obtained by various space-based and ground-based observatories, become available. In parallel, the great progress that has been made in information technologies during the past decades making it possible and essential to develop automated methods that can process large amounts of solar images efficiently and effectively. These methods can detect and predict solar activities automatically and thus can be used for automatic space weather monitoring and prediction in the future.

Much previous work has been done in developing automated methods to detect solar activities, such as filament (prominence) eruptions, flares, sunspots and CMEs. The methods of detecting filaments based on Hα full-disk solar images were previously presented by the NJIT group (Gao *et al.* 2002; Shih and Kowalski, 2003; Qu *et al.* 2004). The methods to detect solar flares from Hα full-disk solar images were discussed by Veronig *et al.* (2000), Borda *et al.* (2001) and Qu *et al.* (2003, 2005). Turmon et al. (2002) presented a method based on Bayesian image segmentation to detect sunspots from MDI magnetograms. The methods to detect CMEs can be found in Wilson (2002,), Qu *et al.* (2005) and Shih *et al.* (2007).

In this dissertation, the methods to detect prominence eruptions and emerging flux regions are developed by using image processing and pattern recognition techniques. Also, the method to predict solar flares based on magnetic parameters of active regions is developed by using statistical machine learning techniques. These methods are reviewed briefly in the following sections.

1.2 Prominence Eruptions

Prominences are clouds of relatively cool and dense gas embedded in the chromosphere and/or in the corona. Usually they are observed as bright features above the solar limb; while the same objects are observed on the solar disk, they show up in absorption as dark features, named as filaments. So in fact prominences and filaments refer to the same kind of physical objects, but they are named differently when observed in different locations.

Prominences are not always quiescent. When the magnetic support of prominences becomes unstable, the materials would erupt into corona. It has been recognized that prominence eruptions are usually associated with other solar activities, such as CMEs, and the relationship between prominence eruptions and CMEs is still an interesting topic in solar physics. Automated detection of prominence eruptions helps to understand their relationship and improve the accuracy of space weather prediction.

In this dissertation, a new automated method is developed to detect and characterize the prominence eruptions. Typically, prominence eruptions are observed in Hα spectral lines, so Hα full-disk images obtained from Big Bear Solar Observatory (BBSO) are used as data set. The input to the method is a sequence of consecutive Hα solar images, of which the time cadence is one minute, and the output is a list of prominence eruption events detected. The method involves several steps. The first step is pre-processing, in which each frame is processed individually and the outlier of all the limb objects in each frame forms a front vector. In the second step, the front vectors of all the frames are piled up in a time sequence to construct a front-time map, based on which the prominence detection and tracking can be quickly carried out. In the third step, the properties of each limb object, such as brightness, angular width, height and rising

velocity, are measured. Finally, a SVM classifier is applied to classify the limb objects and identify real prominence eruptions. The characteristics of prominence eruptions, such as brightness, angular width, radial height and velocity are measured. Experimental results show that the detection rates for eruptive prominences and non-eruptive prominence are 93.3% and 93.6% respectively. The method presented in this chapter can be used in automatic real-time solar activity monitoring and detection.

1.3 Emerging Flux Regions

Emerging flux regions form the first stage of active regions and they usually appear within or at the borders of existing active regions. A fraction of them evolve to flare- and CME-productive active regions, while most of the other EFRs evolve through their life-cycle without producing any major activity. The flare productive active regions are the most important source of space weather effect. The role of EFRs in triggering flares was clearly described (Zirin, 1983). Jing *et al.* (2004) found that in a sample of about 100 CMEs, 70% of them were associated with EFRs, confirming earlier results of Feynman and Martin (1995) that was based on a smaller sample. Therefore, prediction of flare production of an active region before the emergence and at early stage of its evolution is a key topic in the space weather prediction. So it is vitally important to detect EFRs automatically and in real time.

Usually, the full-disk magnetograms, taken by the Michelson Doppler Imager (MDI) (Sherrer *et al.* 1995) at the Solar and Heliospheric Observatory (SOHO), are used as the data source to study EFRs. In the MDI magnetograms, initially an EFR appears as a pair of small opposite-polarity magnetic dipoles, and then the dipoles move apart from

each other. More fluxes are added gradually after their initial emergence. Most regions stop growing in less than a day, but sometimes the magnetic flux continues to emerge and a simple bipolar region may grow into a fully-fledged active region (Strous *et al.* 1996, 1999; Kubo *et al.* 2003).

In this dissertation, a novel method is presented to detect Emerging Flux Regions (EFRs) in consecutive Michelson Doppler Imager (MDI) magnetograms. To the best knowledge of the author, this is the first developed technique for automatically detecting EFRs. The method includes several steps. First, the projection distortion on the MDI magnetograms is corrected. Second, the bipolar regions are extracted by applying multi-scale circular harmonic filters. Third, the extracted bipolar regions are traced in consecutive MDI frames by Kalman filter as candidate EFRs. Fourth, the properties, such as positive and negative magnetic fluxes and distance between two polarities, are measured in each frame. Finally, a feature vector is constructed for each bipolar region using the measured properties, and the Support Vector Machine (SVM) classifier is applied to distinguish EFRs from other regions. Experimental results show that the detection rate of EFRs is 96.4% and of non-EFRs is 98.0%, and the false alarm rate is 25.7%, based on all the available MDI magnetograms in 2001 and 2002.

1.4 Solar Flares

Solar flares are known as short periods of explosive energy release mainly in the chromosphere and corona. The solar flare is one of the most important solar explosive phenomena as far as space weather effects are concerned. Although physical mechanisms of energy buildup and release are not yet fully understood, from the observational point

of view, flares originate preferentially in magnetic active regions with strong field gradients and magnetic shear, long polarity-inversion lines and complex polarity patterns (Falconer 2001; Falconer *et al.* 2003; Song *et al.* 2006; Wang *et al.* 2006; Schrijver 2007).

Forecasting the occurrence of solar flares from the analysis of solar photoshperic magnetogram data is an important and challenging task in the solar physics research. The earlier predictions were primarily based on McIntosh classification of sunspots (McIntosh, 1990) or relying on observer's experience (Zirin and Marquette, 1990). In recent years, the statistical correlation between a variety of photospheric magnetic parameters and flares productivity of source regions have been investigated in many studies, and some parameters are found to be positively correlated with the probability of a region to produce a major flare (Falconer *et al.* 2002; Leka & Barnes 2003a, 2003b, 2007; Jing *et al.* 2006; Schrijver *et al.* 2005; Schrijver 2007; Song *et al.* 2007).

The full-disk MDI magnetograms are used as the primary data source to measure the magnetic parameters of active regions. In this study, the following parameters are considered: (1) total unsigned magnetic flux and (2) total magnetic energy dissipation in a unit layer per unit time. These parameters are chosen primarily because they can be derived directly from the line-of-sight magnetograms and have historically shown their potential in flare forecasting (Jing et al. 2006; Song et al. 2007).

The main goal of this study is to develop a method to predict the occurrence probability of each class flare, such as C-, M- and X-class flares based on the magnetic parameters. This problem is modeled as a conditional density estimation problem, and a novel two-step conditional density estimator is proposed in this dissertation by combining

kernel-based nonlinear regression and moment-based density function reconstruction techniques. The kernel-based nonlinear regression technique is applied first to model the relationship between magnetic parameters and flare peak intensity or flare index and then used for predicting the conditional moments. After that, the probability density function is reconstructed based on the estimated moments.

1.5 Outline of Dissertation

The goal of this study is to develop a suite of methods for automatic solar activity observation and prediction, including prominence eruptions, emerging flux regions and solar flares. This work adopts advance image processing, pattern recognition and machine learning techniques. These methods can be used for processing large-scale solar data efficiently and effectively, and will also lead to real-time space monitoring and prediction.

In Chapter 2, a method for automatic detection of prominence eruptions is presented. Image segmentation, morphological operations and pattern recognition techniques are applied together. In Chapter 3, the details of a method for automatic detection of emerging flux regions (EFRs) is presented. In this method, multi-scale circular harmonic filter is used for detect candidate EFRs, and Kalman filter is used for tracing each candidate EFR in consecutive frames, while support vector machine (SVM) is applied finally to identify true EFR objects. In Chapter 4, a novel method for conditional density estimation is proposed to solve solar flare prediction problem. In this method, kernel-based nonlinear regression techniques is applied to estimate conditional moments of the response variable, and moment-based density function reconstruction

method is applied then to recover the conditional density function of the response variable. Compared with traditional methods, the proposed method is capable of handling high dimensional data efficiently, which is essential for solar flare prediction, since scientists are looking forward to predicting solar flares based on multiple magnetic parameters.

CHAPTER 2

AUTOMATIC DETECTION OF PROMINENCE ERUPTIONS

2.1 Introduction

Prominence eruptions, one of the major solar activities, have received considerable attention since late 1800s. Prominences are cool, dense objects that are embedded in the hot corona. They are held above the Sun's surface by certain magnetic field topology. Their lifetime could be as long as weeks or even months. Prominences are observed above the solar limb, while the same physical structures observed on the solar disk are named filaments. Therefore, prominences and filaments are referred to the same structures, and in this chapter, limb prominences are of interests. The methods of detecting filaments were previously studied by the NJIT group (Gao *et al.* 2002; Shih and Kowalski, 2003; Qu *et al.* 2004).

Prominences are not always quiescent. When the magnetic support of prominences becomes unstable, their material would erupt into corona. During a prominence eruption, the materials are ejected outward rapidly, and either stem from a part of solar surface into a nearby region or leave the Sun completely or partially. Typically, the events would last for a few minutes or hours. Prominence eruptions are usually observed in Hα lines. In Hα solar full-disk images, prominences are bright features above the solar limb against the dark background. As shown in Figure 2.1, a bright prominence is observed clearly above the right side of the solar limb. It has been recognized that prominence eruptions are usually associated with other solar activities, such as flares or coronal mass ejections (CMEs). The relationship between prominence

eruptions and CMEs has been explored (Gopalswamy *et al.*, 2003; Wang *et al.*, 1998; Gilbert *et al.*, 2000), but it is still not fully understood yet. Since flares and CMEs have great influences on space weather, their relationship helps improve the accuracy of space weather prediction.

It is desirable to develop an automatic detection algorithm, which can detect and characterize prominence eruptions with little human intervention. Our research is based on statistics study, so it is essential to collect a large amount of events and measure the associated properties. The data collection is traditionally conducted by human visual inspection that is quite labor-consuming. More importantly, it is usually subjective in selecting events and measuring parameters. On the contrary, the automatic detection algorithm can work more efficiently, objectively, and accurately.

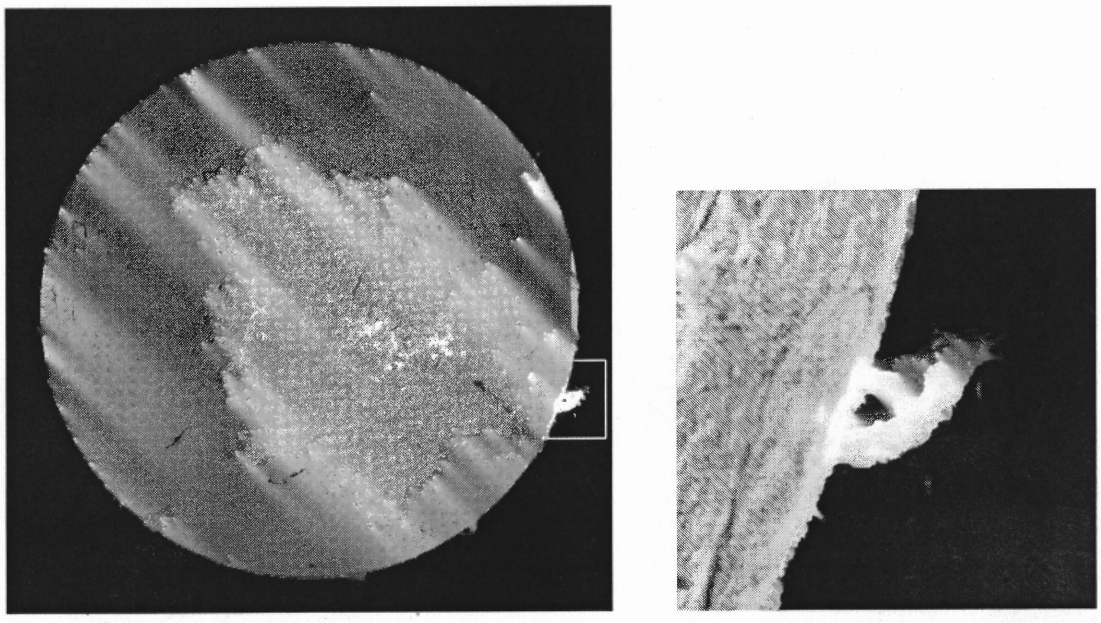


Figure 2.1 A full-disk Hα solar image and its enlarged prominence observed at the Big Bear Solar Observatory on April 15, 2001, 22:15:25 UT.

Gopalswamy *et al.* (2003) developed an automatic prominence detection algorithm using microwave images from Nobeyama radioheliograph. Shimojo *et al.* (2006) presented an improved version of the algorithm. The major difference between these two algorithms is the time interval of the images used for detection. In Gopalswamy's algorithm it is ten minutes, while in Shimojo's algorithm it is three minutes. Both algorithms detect the enhanced pixels whose values are greater than six times of their average values of the day, and then trace the center of all the enhanced pixels in time sequences. If the center location changes persistently over 30 minutes, the algorithms would report it as a candidate limb event. Finally, all the candidate events are inspected visually to obtain true prominence eruptions and measure the properties. The time interval in Shimojo's algorithm is three minutes, which is enough to detect fast prominence eruption, since the life time of most prominence eruptions should be greater than this time interval although it is still insufficient to detect the eruptions of the velocity greater than 400km/s. The algorithms can be used to detect the appearance of prominence eruptions, but still there are some disadvantages in both algorithms. First, the algorithms cannot detect slowly eruptive prominences because they would increase the average pixel values of the day and cause the prominence pixels to be enhanced. Second, since the algorithms trace the centers of all the enhanced pixels, they cannot detect the prominence eruptions if they occur simultaneously above the opposite hemispheres. Third, the algorithms can only detect the existence of a prominence event, but do not check the direction and speed of the prominence motion, so they cannot characterize the features automatically.

In this chapter, a new algorithm is developed to detect and characterize prominence eruptions automatically. It consists of several steps. The first step is pre-processing, in which each frame is processed individually and the outlier of all the limb objects in each frame forms a front vector. In the second step, the front vectors of all the frames are piled up in a time sequence to construct a front-time map, and based on which the prominence detection and tracking can be quickly carried out. In the third step, the properties of each limb object, such as brightness, angular width, height and rising velocity, are measured. Finally, a support vector machine (SVM) classifier is applied to classify the limb objects.

The rest of this chapter is organized as follows. The pre-processing techniques are presented in Section 2.2. The feature extraction and prominence classification are described in Section 2.3. Experimental results are provided in Section 2.4. Finally, the conclusions are presented in Section 2.5.

2.2 Pre-processing

Image pre-processing intends to process an image, so the result is more suitable than the original image for a specific application. First, a 3×3 median filter is applied to remove the noise in the captured solar image. Then, the polar transformation is applied onto the region surrounding the solar disk limb where the prominences reside, and perform image segmentation based on local contrast to extract bright pixels from dark background. Finally, the limb structure front is extracted.

2.2.1 Polar Transformation

The full-disk Hα solar images are observed from the Big Bear Solar Observatory (BBSO). The images are acquired by a large-format, 2032×2032 pixels, 14-bit Apogee KX4 CCD camera, manufactured by Apogee Instruments (Denker *et al.*, 1999; Steinegger *et al.*, 2000), and the time cadence is one image frame per minute. Prior to further processing, the images have been calibrated to remove the limb darkening, and the basic image parameters, such as the position and radius of the solar disk, have been obtained. The technical details pertaining to calibration and parameter measurement were discussed by Denker *et al.* (1999).

Since the prominences are observed above the solar limb, the polar transformation is then applied onto this surrounding region of the solar disk, as shown in Figure 2.2. Let O be the center of the solar disk, R_0 be the radius of the solar disk, and R_1 be the radius of the outer circle, which is centered at O and tangent to image boundaries. Since the solar disk is centered during calibration, the two circles are co-centered at O. The circular region between the two concentric circles is then transformed into a rectangular (or angular) image, as shown in Figure 2.3. The radius of the solar disk is typically 900 pixels, and the width of the original image is 2032 pixels. Therefore, the width of the angular image is approximately 5655 pixels and the height of the angular image is 116 pixels. Let w_I and h_I denote the width and height of the angular image, respectively. In order to speed up processes, the angular image size is reduced to 2000×1000 pixels.

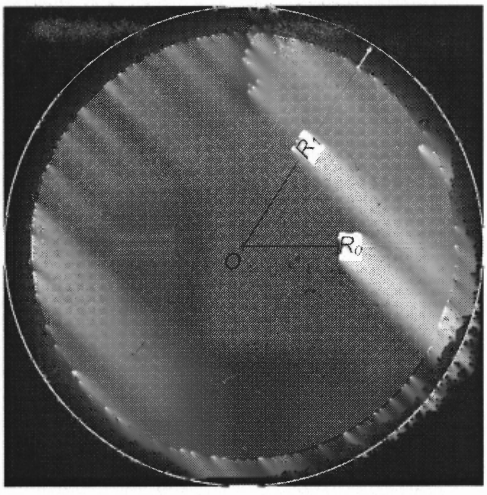


Figure 2.2 The circular region to be transformed between two co-centric circles centered at O with radius R_0 and R_1.

The relationship of the coordinates (x, y) in the angular image and (x', y') in the original image can be represented as follows:

$$\begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} 0 & \sin x & x_c \\ 0 & \cos y & y_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \tag{2.1}$$

where (x_c, y_c) denote the coordinates of the disk center in the full-disk image . The bilinear interpolation is used to calculate the gray values in the angular image, as shown in Figure 2.3.

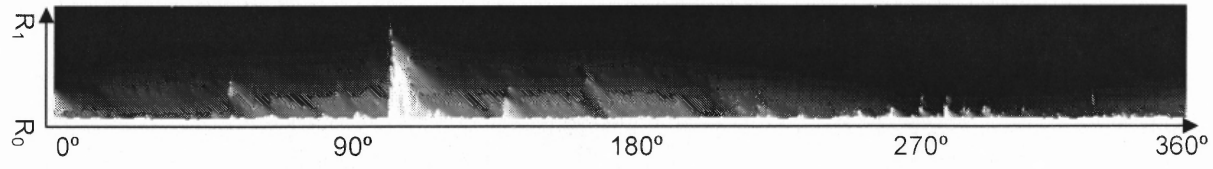


Figure 2.3 The transformed angular image of the Hα solar image observed at the Big Bear Solar Observatory on April 15, 2001, 22:28:25 UT.

2.2.2 Image Segmentation

Because the light from the bright limb objects is scattered, their surrounding background would look brighter than the faint limb objects. The global thresholding method may miss the faint limb objects or select the bright background pixels. Therefore, the local contrast method is developed. Let the contrast image $C(x, y)$ be defined as

$$C(x, y) = \ln \frac{I(x, y)}{I_N(x, y)}, \tag{2.2}$$

where $I(x, y)$ is the intensity of pixel (x, y) and $I_N(x, y)$ is the average intensity of its neighborhood. In order to calculate the contrast image, a linear diffusion filter is applied iteratively onto the angular image. The number of iteration t defines the scale of resolution at which the image is observed. A small t corresponds to a fine scale and a large t corresponds to a coarse scale.

Let I_0 denote the original image, and $I_1, I_2, \ldots, I_t$ denote the successive iteration of the coarse images. In each step, the pixel is coupled to its four neighbors by a force function $F(u)$. The linear diffusion filter is defined as

$$\begin{aligned} \Delta I_t(x, y) &= I_{t+1}(x, y) - I_t(x, y) \\ &= G(I_t(x-1, y) - I_t(x, y)) + G(I_t(x, y-1) - I_t(x, y)) \\ &\quad + G(I_t(x+1, y) - I_t(x, y)) + G(I_t(x, y+1) - I_t(x, y)) \end{aligned} \tag{2.3}$$

where $\Delta I_t(x, y)$ is the derivative representing the evolution in the number of iterations. For rapid processing, a linear force function is used as follows:

$$G(u) = u. \tag{2.4}$$

In each iteration the pixel value of the angular image is updated by adding the derivative. The number of iterations is determined based on the comparison between segmented results and visual effects. The value chosen can enable them to be very close.

If more iterations are applied, the image $I_N(x,y)$ would be more blurred and too many weak features would be picked up. On the contrary, if less iterations are applied, the image $I_N(x,y)$ would be less blurred, but some strong features would not be distinguishable from background. Based on experimental results, the appropriate iteration number is set to 20. Figure 2.4 shows the blurred image after applying the diffusion filter 20 times.

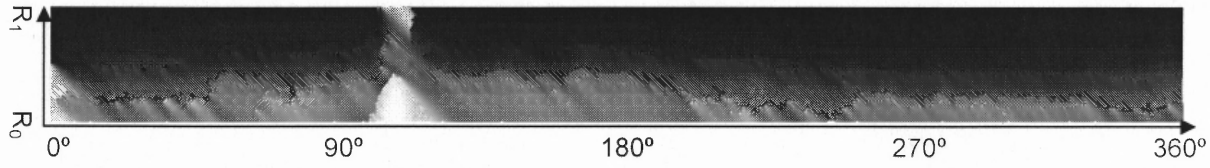


Figure 2.4 The blurred angular image of the Hα solar image observed at the Big Bear Solar Observatory on April 15, 2001, 22:28:25 UT. after applying the diffusion filter 20 times.

The contrast image, as shown in Figure 2.5, is obtained by using eq. (2.2) with the angular image and the average local brightness image. It can be observed that the bright background around the large middle-left prominence is removed. If a pixel is brighter than its local background, its value is positive; otherwise, its value becomes negative.

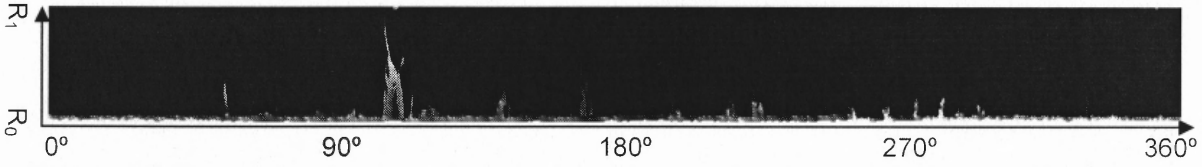


Figure 2.5 The contrast angular image of the Hα solar image observed at the Big Bear Solar Observatory on April 15, 2001, 22:28:25 UT.

Next, a threshold function $F(x,y)$ is applied to segment the contrast image as follows:

$$F(x,y)=\begin{cases}1, & \text{if } C(\mathrm{x},\mathrm{y})\geq T_f.\\ 0, & \text{otherwise.}\end{cases} \tag{2.5}$$

where T_f is a threshold. The T_f is set to be slightly greater than 0 (say, 0.045) to avoid picking up some noisy pixels in dark areas. The value is determined based on the comparison between segmented results with visual effects. Next, morphological closing and opening are applied to fill in small gaps and remove small noisy regions (Shih & Mitchell 1989), respectively. The resulting image is shown in Figure 2.6. By comparing Figures 2.3 and 2.6, it can be noticed that most bright object pixels are picked up while a small number of faint objects are missed. It is acceptable since the eruptive prominences are usually bright and will be seldom missed.

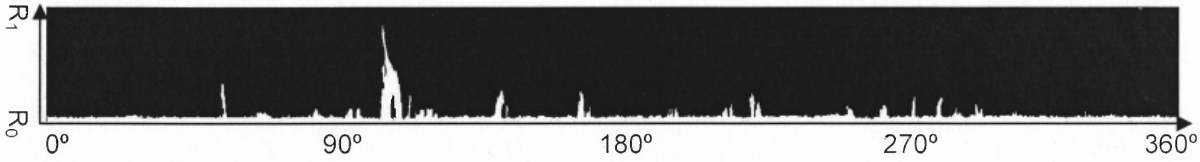


Figure 2.6 The segmented angular image of the Hα solar image observed at Big Bear Solar Observatory on April 15, 2001, 22:28:25 UT.

2.2.3 Structure Front Extraction

The front of limb objects are of interests, since the movement of the front is a vital indicator to detect the eruptive prominences. A vector, named *front vector*, is used to record the profile of limb objects. By scanning all the columns in the thresholded image, The front vector $V(x)$ is defined as follows:

$$V(x) = \max\{y \mid F(x,y) = 1\}, \tag{2.6}$$

where $F(x,y)$ denotes the thresholded image. Figure 2.7 shows the front vector.

Since the radius of the solar disk may vary in different frames, the front vector is normalized in order for comparison, which is defined as follows:

$$V^*(x) = V(x) \times \frac{(W/2 - R_0)}{h_l R_0}, \tag{2.7}$$

where W denotes the width of the original full-disk Hα image and h_l denotes the height of the angular image.

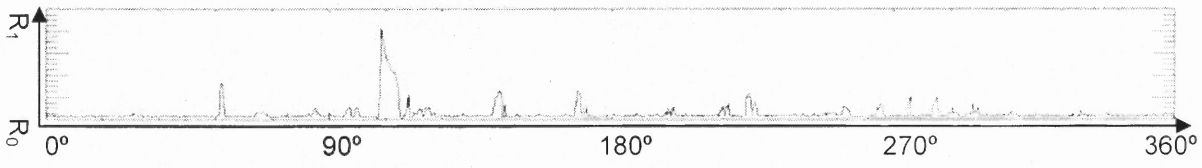


Figure 2.7 The front vector of the Hα solar image observed at Big Bear Solar Observatory on April 15, 2001, 22:28:25 UT.

2.3 Feature Extraction and Prominence Classification

In order to trace the moving limb objects, a front-time image is defined. The major limb objects are extracted by thresholding. Then, the properties, such as time span, position angle, angular width, radial height and brightness, are measured. Finally, the pattern classification is performed.

2.3.1 Limb Object Tracing

The new eruptive prominences may appear at any time, and the existing eruptive prominences may change the size and shape successively. A front-time image, denoted as $I_f(x, y)$, is constructed by combining the front vectors of all the frames to detect the appearance and trace the movement. Its size is $w_f \times h_f$, where w_f denotes the width of the angular image and h_f denotes the number of all the frames taken within one day. Each row corresponds to a front vector at a time instance, and the pixel value indicates

the height of the corresponding front vector at a certain angular position. The front-time image $I_f(x, y)$ can be represented as

$$I_f(x, y) = V_y^*(x), \tag{2.8}$$

where $V_y^*(x)$ is the front vector of yth frame. Figure 2.8 shows the front-time image obtained from all the Hα images taken on April 15, 2001. Note that it is rescaled to the range of [0, 255] for the display purpose.

Since the positions of the limb objects will change very little in the successive frames, the same limb object should appear as a connected component in the front-time image in spite of different shape and topological characteristics. A stable limb object corresponds to a stripe-like component, and an eruptive object corresponds to a spot-like component. Figure 2.8 shows several bright stripes and a spot-like component. Assume that each pixel in the front-time image corresponds to at most one limb object. Some exceptional case may have two or more limb objects overlaid, but it is very rare.

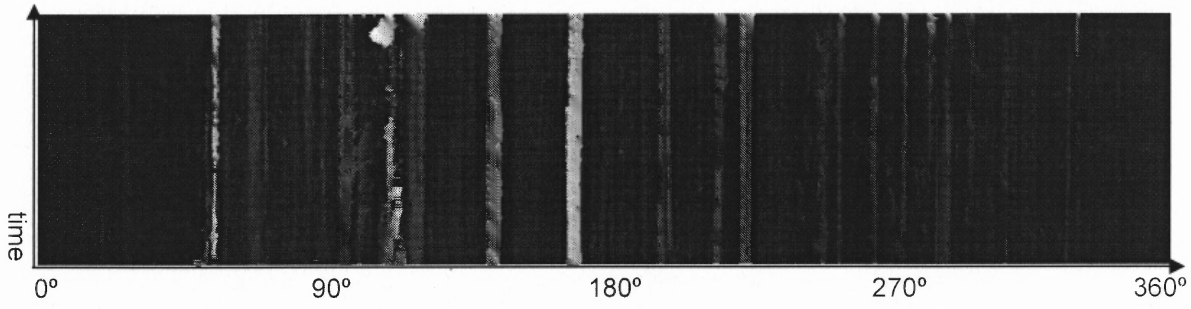


Figure 2.8 The front-time image representing the solar limb activities from April 15, 2001, 20:59:25 UT. to April 15, 2001, 22:56:25 UT.

To extract eruptive limb objects, the image is segmented based on thresholding. The threshold is defined as $H_{med}(x) + T_f^1$, where $H_{med}(x)$ is the median gray level of column x and T_f^1 is a threshold. In the experiment, T_f^1 is 0.013, which corresponds to

8,970 km. In each column, any pixel whose gray level exceeds the threshold is accepted. Since the variance of radial height of stable objects will be quite small, the radial height of stable objects would seldom exceed the threshold. Thus, the stable limb objects will be eliminated almost completely, and only a few stable objects will be preserved together with eruptive objects after the segmentation. The eruptive objects will be preserved because during the eruption the radial height of eruptive objects will be significantly greater than the threshold. Then the morphological closing by the structuring element of 2×5 is applied to merge the disconnected components, and the morphological opening is performed to remove the small components by using the structuring element of 4×5. Figure 2.9 shows the resulting image where many stripe-like components are removed since they correspond to stable limb objects, and the spot-like component is preserved since it corresponds to the eruptive object.

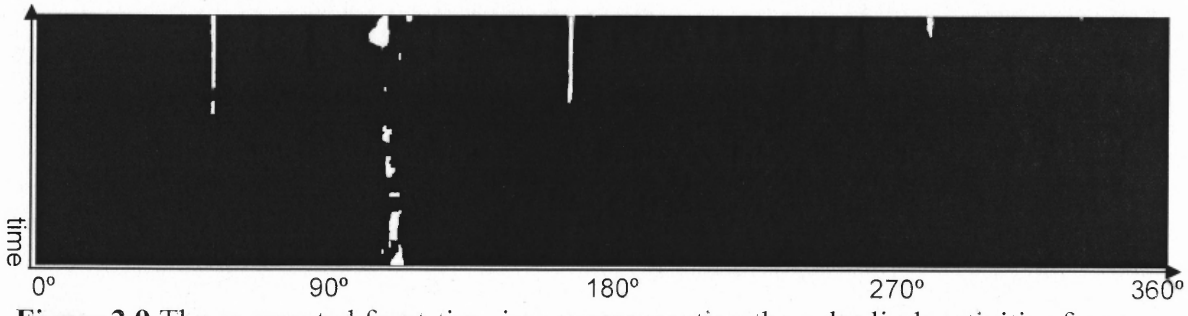


Figure 2.9 The segmented front-time image representing the solar limb activities from April 15, 2001, 20:59:25 UT. to April 15, 2001, 22:56:25 UT.

The segmented front-time image is used as a reference to extract eruptive limb objects. Each connected component corresponds to a limb object in this image because the angular position of limb objects should not change significantly during the observation hours. Thus, in the consecutive rows the line segments of the same limb object would be attached to each other. Both the angular width and time span of the limb

object can be obtained from the segmented image, since the x-coordinate of the component corresponds to the angular position axis and the y-coordinate corresponds to the time axis. After that, the corresponding image segment, which contains the limb object, are extracted. The limb objects are extracted based on the segmented front-time image and measure the associated properties from the extracted image segment.

2.3.2 Property Measurement

According to the physical nature of the limb objects, each object is represented by nine features and use them for classification. The nine features are computed from the angular image and the segmented image as follows:

a. The time span, denoted as t. It is the lifetime of the object which is determined by the first time it appears and the last time when it is still detected.

b. The maximum radial height, denoted as h_r. The radial height of the object is computed in each angular image. After processing all the frames, a radial-height-time function is obtained. It is then convolved by a Gaussian low-pass filter with the standard deviation 2.0. The maximum value is taken as the feature.

c. The maximum of the median angular widths, denoted as w_m. From the segmented image $F(x, y)$, the angular width of the object is measured from bottom to top, and then an angular-width-height function is obtained. The median value is calculated, so a median-angular-width-time function is also obtained after processing all the frames. Then, it is convolved with the same Gaussian low-pass filter, and the maximum value is taken as the feature.

d. The shape feature is defined as $r = \frac{w_m}{h_r}$.

e. The maximum size of the limb object, denoted as s. The total number of pixels of the limb object in $F(x, y)$ is computed to form the size-time function which is then convolved with the Gaussian low-pass filter. The maximum value is taken as the feature.

f. The maximum average brightness of the object, denoted as b. In each frame, all the pixels of the object are extracted from the segmented image $F(x, y)$, and the average of the corresponding pixel values in the angular image are calculated to

form the average-brightness-time function. It is then convolved with the Gaussian low pass filter, and the maximum value is taken as the feature.

g. The average brightness of the object in the key frame, denoted as b_k. The key frame is the frame where the size of the object reaches the maximum. Then, the average brightness of the key frame is calculated as the feature.

h. The standard deviation of object brightness in the key frame, denoted as δ_k.

i. The maximum rising velocity, denoted as v_r. The first derivative of the smoothed radial-height-time function with respect to time is computed to form the velocity-time function. After convolving with the Gaussian low-pass filter, the maximum is taken as the feature. The rising velocity, which is the component projected onto the sky of the actual velocity vector, provides a distinguishable feature for classification.

Finally, for each detected limb object, a feature vector v_f is constructed and defined as $v_f = (t, h_r, w_m, r, s, b, b_k, \delta_k, v_r)$.

2.3.3 Eruptive Prominence Detection

The support vector machine (SVM) is applied to classify the limb objects. The SVM, introduced by Vapnik (1998), is based on statistics learning theory for the two-class classification problem. The idea is to map the input patterns to a high dimensional feature space and construct an optimal hyper-plane to separate all the patterns (Guyon *et al.* 2000). Different kernel functions can be used, such as linear, polynomial, sigmoid and radial basis function. The polynomial of degree 2 is used as the kernel function in the experiment.

The classifier requires training prior to testing. Each pattern in the training set is represented as a feature vector and associated with a class label. For eruptive prominence classification, the training samples are first classified by visual inspection. The class label d is assigned to be 1 for eruptive prominences and -1 otherwise. Then, the SVM

classifier takes the input patterns and the associated class labels to compute the optimal decision hyper-plane. After that, the SVM classifier is ready to test an unknown limb object.

2.4 Experimental Results

The proposed method is implemented in the Interactive Data Language (IDL), developed by Research Systems Inc., together with the Solar SoftWare (SSW) IDL library (Freeland *et al.* 1998). The experimental results are available to the public through the web site http://filament.njit.edu/.

The BBSO full-disk Hα solar images are used as the data set for training and testing. Due to the image availability, The images observed in 475 days from 2001 to 2005 at BBSO are investigated, in which there are 21 days in 2001, 10 days in 2002, 63 days in 2003, 208 days in 2004 and 173 days in 2005. The time cadence is one minute and the observation hours are eight hours a day. Therefore, there are up to 480 images to be observed in one day although fewer images in certain days would be obtained due to bad weather or other reasons. 26 eruptive prominences are identified by visual inspection. Some ambiguous candidate objects, which are very difficult to classify even by human eyes, are excluded. By applying our program, 926 limb objects are detected among which all 26 predefined eruptive prominences are included.

The leave-one-out strategy (Hoffbeck *et al.* 1996) is adopted to train and test the SVM. In order to measure the classification rate accurately, the experiment contains 1000 iterations. There are two steps in each iteration. First, 25 eruptive and 500 non-eruptive prominences are picked up randomly to train the SVM. Second, the remaining eruptive

prominence and a random non-eruptive prominence are used for testing. After 1000 iterations, the average classification rate is calculated. The experiment is performed five times and the results are listed in Table 2.1. According to the experimental results, the detection rates of prominence and non-prominence eruptions are 93.3% and 93.6% respectively. Since totally there are 26 prominence eruptions and 900 non-prominence eruptions, it can be expected that 24.3 prominence eruptions would be classified as prominence eruptions on the average, though another 57.6 non-prominence eruptions would be also classified as prominence eruptions.

Table 2.1 Experimental Results of Automatic Prominence Eruption Detection

Exp. No.	P. E.	Non-P. E.	Total
1	93.2%	93.9 %	93.6%
2	94.0%	93.9%	94.0%
3	91.9%	92.7%	92.3%
4	94.8%	93.0%	93.9%
5	92.8%	94.5%	93.7%
Average	**93.3%**	**93.6%**	**93.5%**

Most of the misclassification on eruptive prominences happens to be faint eruptive prominences. Because of their low brightness, they may not be segmented properly, and the properties, such as angular with and radial height, may not be correctly measured. Most of the misclassification on non-eruptive prominences happens in two conditions. One is that their features, such as rising velocity and brightness, are close to the eruptive prominences. The other is that the measured rising velocity is much higher than the actual value. The inaccurate measurement may be due to the faint object being lost in some frames and then coming back in the subsequent frames, though the radial height is smoothed before the rising velocity is calculated

From Table 2.1, it is noted that the true positive detection rate of eruptive prominence detection is 93.3% and the false positive detection rate is 6.4%, which is a small number. However, if there were a large amount of non-eruptive prominences detected by the algorithm, there would be a large fraction of detected eruptive prominences that are not true eruptions. Therefore, it is critical to improve the performance of the algorithm. There are two aspects can be considered. One is to improve the image processing techniques; for example, image enhancement techniques can be applied to enhance faint objects to avoid loss and obtain accurate measurement. The other is to combine the detection of filament and prominence eruption since the prominences are the filaments observed above the solar limb. The detection of filament on the solar disk will determine whether there is a prominence above the solar limb by taking the rotation of the Sun into consideration. The existence of a prominence and an eruption detected above the solar limb would be a strong evidence for prominence eruption.

The IDL program runs on a Dell Dimension 4600 with an Intel Pentium 4 processor (2.8 Ghz) and 1.0 GB memory under Fedora Core Linux 4.0. It takes less than 3 seconds to process a single frame and extract the front vector. The time to process the front-time image and measure the associated properties depends on the number of frames taken and the number of limb objects detected. For instance, it took 5 seconds to process 118 images taken between 20:59:25 UT and 22:56:25 UT, April 15, 2001 where seven limb objects were detected. This means that it spent about 0.04 seconds per frame. The SVM training is performed off-line. The time to classifying an object takes less than 1

second. Therefore, it takes approximately less than 5 seconds in each frame, which is quite efficient.

2.5 Discussion and Conclusions

In this chapter, a method to automatically detect eruptive prominences is presented by using continuous full-disk H α solar images. The experimental results show that the method works successfully on the eruptive prominences. Only a few insignificant eruptive prominences are misclassified.

Currently, the method can only work offline because the front-time image is defined on all the frames, which is obtained after all the frames are taken. In the future, it is intended to make improvement in four aspects. First, the alternative properties will be investigated to improve the detection rate. Second, the image segmentation technique will be improved to avoid losing faint objects by applying image enhancement techniques and considering consecutive frames instead of a single frame. Third, it will be interesting to consider the probability of combining the detection of filament and prominence eruption to improve the detection rate. Fourth, an online version for real-time detection is going to be developed. Furthermore, extending the current method to detect other limb objects will be considered, such as jets, macro-spicules or spiclues, all of which are limb ejections with visible rising motion.

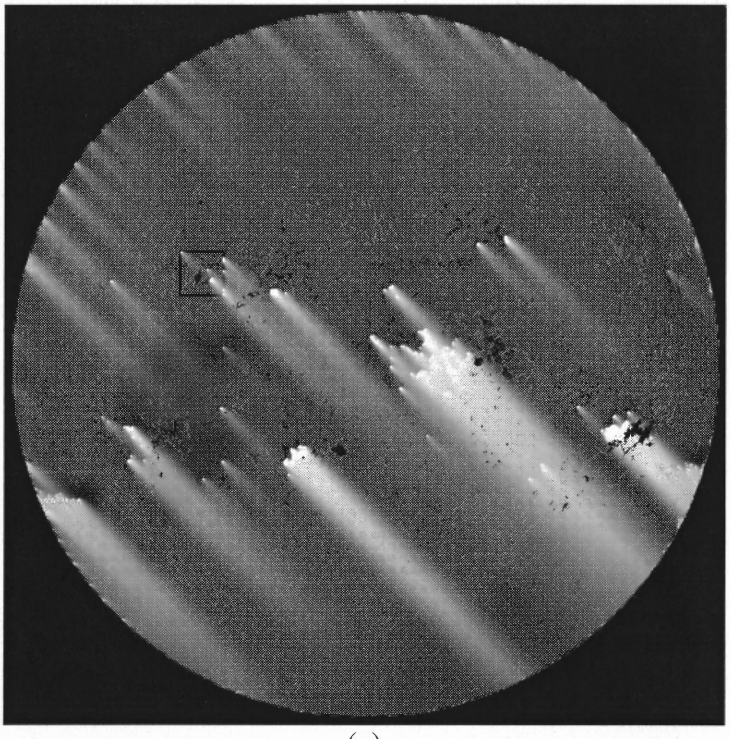
CHAPTER 3

AUTOMATIC DETECTION OF EMERGING FLUX REGIONS

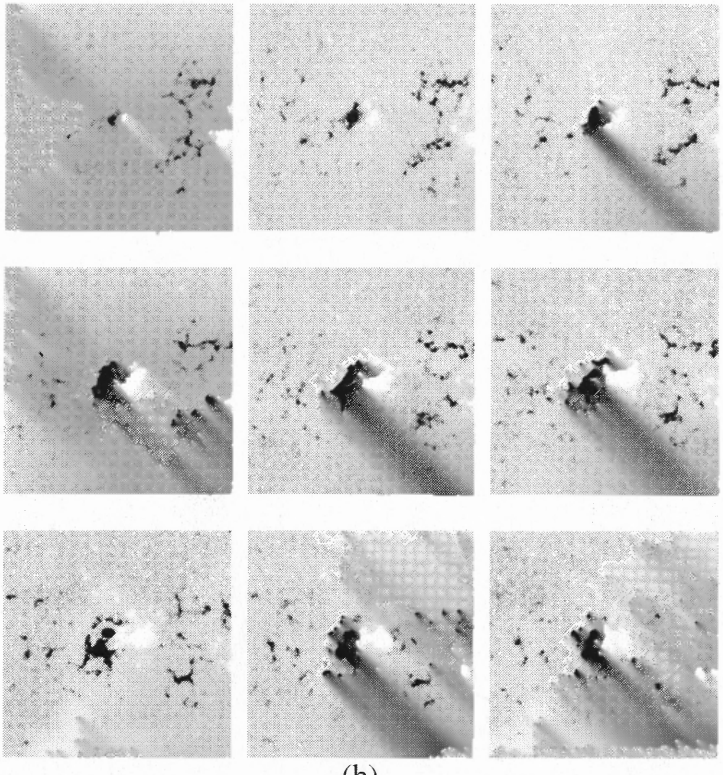
3.1 Introduction

Emerging Flux Regions (EFRs), which form the first stage of active regions, could appear within active regions and everywhere in the quiet solar surface. Some EFRs evolve through their life cycle without producing a major activity, while a fraction of them evolve to flare and Coronal Mass Ejections (CME) productive active regions. It is believed that flares and CMEs have great impact on space weather, and EFRs play an important role in trigging them (Jing *et al.* 2006). Since the prediction of flare producing of active regions before emergence is a key topic in space weather prediction, it is vitally important to detect EFRs automatically and in real time.

The EFR is a bipolar magnetic emergence (Bruzek 1977). Initially, it appears as a pair of small opposite-polarity magnetic dipoles, and then the dipoles move apart from each other. More fluxes are added gradually after their initial emergence. Most regions stop growing in less than a day, but sometimes the magnetic flux continues to emerge and a simple bipolar region may grow into a fully-fledged active region (Strous *et al.* 1996, 1999; Kubo *et al.* 2003). Figure 3.1 (a) shows a full-disk MDI magnetogram on April 18, 2002, with an EFR marked, and (b) shows its evolution within a week from its initial emergence.



(a)



(b)

Figure 3.1 (a) Full-disk MDI magnetogram on April, 18, 2002, 22:24:01 UT, and (b) evolution of the marked EFR from April 18, 2002, 00:00:01UT to April 25, 2002, 19:11:01UT.

It is considered that there is no existing automatic EFR detection algorithm available. However, there is an automatic algorithm to detect Ephemeral Regions (ERs) presented by Hagenaar (2001). The ERs, similar to EFRs, are short-lived, small emerging flux regions (Bruzek 1977; Harvey 1993). Though ERs and EFRs may occur at different origins, there is no clear cut between them (Harvey *et al.* 1973; Wang *et al.* 1992). The most notable difference appears that EFRs have sunspots, while ERs do not. They can be also be distinguished based on their properties, such as life time, size, flux, and magnetic field strength. In the ER detection algorithm (Hagenaar 2001), all the major magnetic flux concentrations are first extracted by a global threshold. Then, two concentrations of opposite polarities are determined as a dipole based on the distance between their centers of gravity and the fluxes in two concentrations. After that, the coherent concentrations are traced in the consecutive frames.

The basic problem to detect EFRs is to extract the pairs of coherent peaks with opposite polarities. If applying the ER detection method discussed by Hagennar (2001) to detect EFRs, there would be a major problem. Because the magnetic concentration may be scattered within a neighborhood, each magnetic concentration of EFRs will be divided into a group of discrete fragments in the MDI magnetogram after thresholding. Therefore, it is difficult to decide the group size and re-group these discrete fragments to form a complete magnetic concentration. Note that the fragments of the same magnetic concentration should be considered as a whole instead of individual ones.

In this chapter, a new method is proposed to detect EFRs using pre-processing, the multi-scale circular harmonic filters, Kalman filter, feature vector, and the SVM classifier. The method involves several steps. First, the projection distortion on the MDI

magnetograms is corrected. Second, the multi-scale circular harmonic filters are applied to the MDI magnetograms to extract bipolar regions. The circular harmonic filter is a powerful technique and has been widely used in signal detection (Lugt 1964; Hsu *et al.* 1982). It allows rotation invariance, which is important since the EFRs could be in any orientation. Third, the extracted bipolar regions are traced as candidate EFRs in the consecutive MDI magnetograms based on Kalman filter. Fourth, after the complete observation is available, the candidate EFR is classified based on its time profiles of properties by the pattern recognition technique.

The full-disk MDI magnetograms are used as the data source in this study. The bright area on the solar surface indicates the fields with positive magnetic signals, while the dark area indicates the fields with negative magnetic signals. The magnetograms can measure the magnetic flux density distribution. The size of each MDI magnetogram is 1024×1024 and the time cadence is 96 minutes.

The remainder of this chapter is organized as follows. The pre-processing method is presented in Section 3.2 and the processing of single frames in Section 3.3. The tracing and classification of candidate EFRs are discussed in Section 3.4 and Section 3.5, respectively. The experimental results are shown in Section 3.6. Finally, the conclusions are presented in Section 3.7.

3.2 Pre-processing Method

The Sun is a sphere, while the observed MDI magnetograms are two dimensional. Due to projection effect, the pixels close to the solar limb correspond to a larger area than those near the disk center in the MDI magnetogram. According to Hagenaar (2001), a pixel at

the disk center corresponds to an area of $S = 2'' \times 2''$. The scaling factor at a point (x, y), whose coordinates are in the Cartesian coordinate system with the origin at the solar disk center, depends on its position angle α. As illustrated in Figure 3.2, the position angle α is the angle between z-axis and the line from the origin to the point (x, y, z), in which $z = \sqrt{R^2 - x^2 - y^2}$, and R is the radius of the solar disk in pixels.

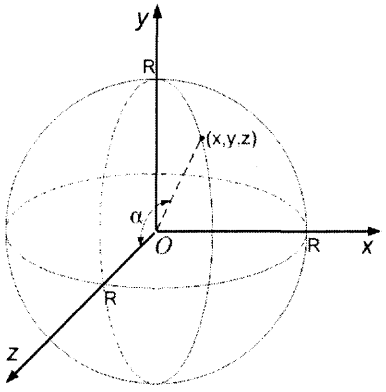


Figure 3.2 The position angle α at the point (x, y).

Therefore, given a pixel (x, y), the angle α is calculated as

$$\alpha = \cos^{-1} \frac{\sqrt{R^2 - x^2 - y^2}}{R}. \tag{3.1}$$

Since a pixel at angle α corresponds to an area of $S_{\perp}(\alpha) = S / \cos\alpha$, the size of each pixel is multiplied by the factor of $1/\cos\alpha$ to obtain the actual corresponding area.

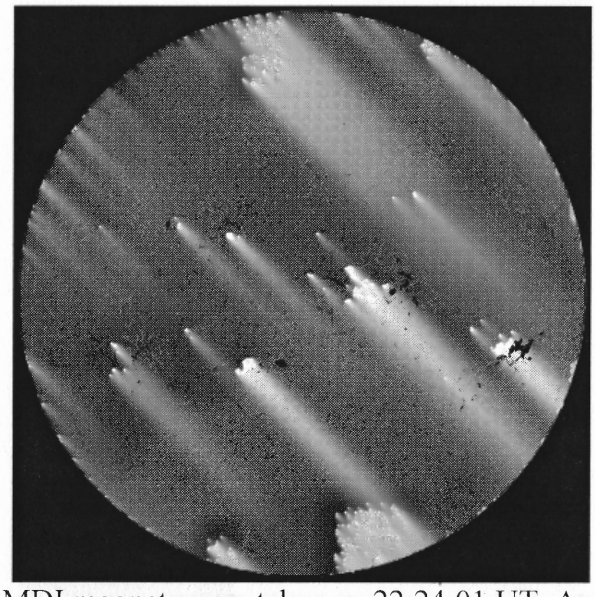
The MDI magnetogram measures the magnetic flux density at each pixel, and the observed magnetic flux density is the line-of-sight component of the actual flux density. The magnetic flux density perpendicular to the solar surface is needed to evaluate at each pixel. However, in practice it is impossible to estimate the density, since the orientation of the flux is unavailable. According to Murray (1992), it is assumed that the magnetic flux density perpendicular to the solar surface, denoted as $B_{\perp}(x, y)$, is related to the

observed flux density, denoted as $B(x,y)$, in such a way that $B_{\perp}(x,y) = B(x,y)/\cos\alpha$, where α is the position angle at (x,y). Thus, the flux $\Phi_{\perp}(x,y)$ is computed as

$$\Phi_{\perp}(x,y) = B_{\perp}(x,y)S_{\perp}(\alpha) = \frac{B(x,y)S}{\cos^2\alpha}. \tag{3.2}$$

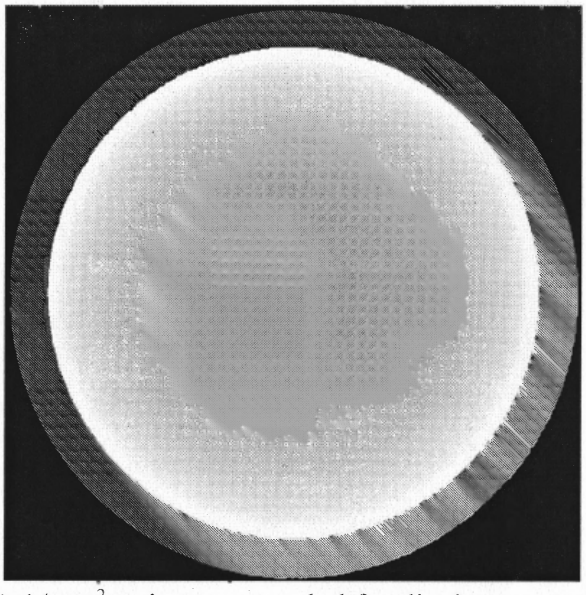
Note that the correction factor used is $1/\cos^2\alpha$, which is different from the one used by Hagenaar (2001). The estimated flux $\Phi_{\perp}(x,y)$ will be used for further processing to replace the observed flux density.

Similar to Hagenaar (2001), only $0 \le \alpha \le 60^{\circ}$ is considered, because when $\alpha \ge 60^{\circ}$, the correction $1/\cos^2\alpha \ge 4$, and when $\alpha \to 90^{\circ}$, $1/\cos^2\alpha \to \infty$, both producing severe geometric distortion of EFRs. Figure 3.3 illustrates the correction of an MDI magnetogram, in which (a) is the original MDI magnetogram, (b) is the image of $1/\cos^2\alpha$, which is rescaled for display purpose, and (c) is the corrected magnetogram whose pixel value is multiplied by $1/\cos^2\alpha$ for $0 \le \alpha \le 60^{\circ}$.

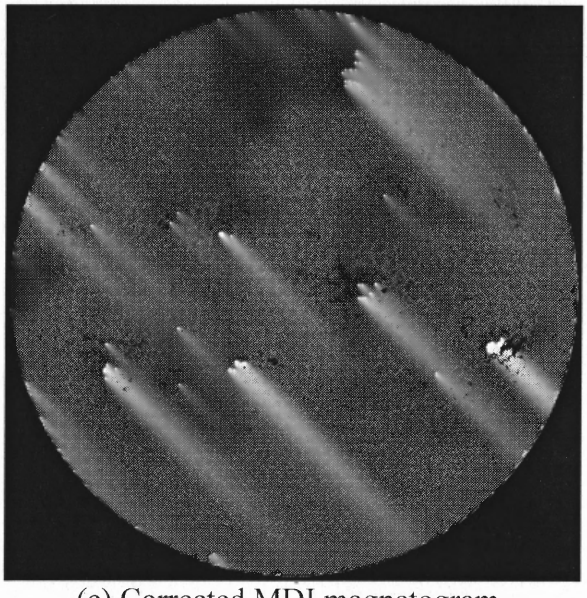


(a) Original MDI magnetogram taken on 22:24:01 UT, April, 18, 2002.

Figure 3.3 Correction of the MDI image, taken on April, 18, 2002, 22:24:01UT.



(b) $1/\cos^2 \alpha$ image, rescaled for display purpose.



(c) Corrected MDI magnetogram.

Figure 3.3 Correction of the MDI image, taken on April, 18, 2002, 22:24:01UT. (Continued)

3.3 Processing of Single Frames

In this section, the extraction of bipolar regions from the MDI magnetograms in single frames is presented. The circular harmonic filters are applied to correlate with the MDI magnetogram. The peaks in correlation indicate the locations where the filter closely

matches the enclosed object. To match the EFRs of different sizes, multi-scale filters are applied, all of which are derived from the same base filter. To ensure high discrimination, the shape of the filter should be close to the EFRs. The strategy aims to derive the filter from previously observed EFRs. Therefore, the filter is designed using the circular harmonic decomposition of EFR samples in consideration of rotation invariance.

Some correlation peaks may appear in positive- or negative- flux dominant regions. Such peaks are false peaks, because the EFR is a bipolar region, which should contain both positive and negative magnetic fluxes. Therefore, the *local positive to negative flux ratio* (LPNFR) at each pixel within the region covered by the filter is calculated. If the positive and negative fluxes are approximately balanced, the ratio would be close to 1; while in the positive or negative flux dominant regions, the ratio approaches to 0.

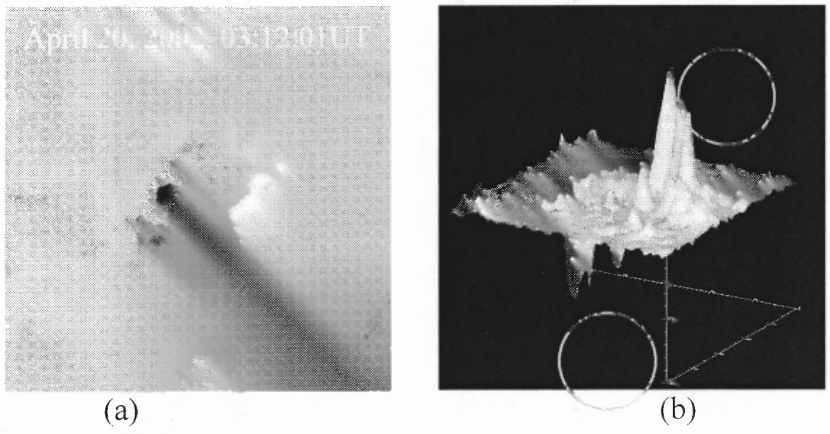
Note that some high correlation peaks generated by the filter do not cover the object completely. In other words, the correlation response is generated based on the incomplete object, though the correlation intensity could be high. The correlation peaks generated on the complete object are of interest, instead of the partial object. Therefore, the *relative energy ratio* (RER) is defined to evaluate how completely the filter covers the object. If the filter covers the object completely, the ratio is 1; otherwise, it is smaller than 1.

The product of correlation, LPNFR and RER is calculated at each pixel. In the combined response, the peaks located in positive- or negative- flux dominant regions or based on the incomplete objects are suppressed. After applying the filters of all sizes, a set of combined responses are generated. The peaks in each combined response are

extracted by thresholding. They will be traced in the consecutive MDI frames by Kalman filter as candidate EFRs.

3.3.1 Multi-scale Circular Harmonic Filter

As aforementioned, the EFR is a bipolar magnetic region containing both positive and negative concentrations. Figure 3.4(a) shows a sample EFR image, and (b) shows the 3-D structure of the EFR, in which two concentrations enclosed by two circles can be observed clearly.



(a) (b)

Figure 3.4 (a) A sample EFR image and (b) the 3-D structure of the EFR.

To allow rotation invariance, the filter is designed based on the *circular harmonic* (CH) expansion of EFRs. Let $f(r,\theta)$ be a sample image, which can be expressed by the CH components (Premont & Sheng 1993) as

$$f(r,\theta) = \sum_{m=-\infty}^{\infty} f_m(r,\theta) = \sum_{m=-\infty}^{\infty} f_m(r)\exp(jm\theta), \tag{3.3}$$

where the CH function $f_m(r)$ is

$$f_m(r) = \frac{1}{2\pi}\int_0^{2\pi} f(r,\theta)\exp(-jm\theta)d\theta\,. \tag{3.4}$$

The function $f_m(r,\theta)$ is the m-th order CH component. The correlation between the image $g(x,y)$ and the filter $f_m(r,\theta)$ is defined as

$$C(x,y) = g \circ f_m^*(x,y) = 2\pi\int_0^{\infty} g_m(r;x,y)f_m^*(r)rdr\,, \tag{3.5}$$

where '*' denotes complex conjugate, and $g_m(r;x,y)$ denotes the m-th order CH function of the image $g(x,y)$ expanded at (x,y). Rotation invariance is verified because the correlation intensity only depends on the CH functions of the image $g(x,y)$ and the filter $f_m(r,\theta)$.

In the EFR detection, the first order CH component $f_1(r,\theta)$ is chosen as the filter based on the analysis of energy distribution. The energy function $f(r,\theta)$ is defined as

$$E_f = \int_0^{2\pi}\int_0^{\infty}|f(r,\theta)|^2 rdrd\theta = \sum_{m=-\infty}^{\infty} 2\pi\int_0^{\infty}|f_m(r)|^2 rdr = \sum_{m=-\infty}^{\infty} E_{f_m}\,. \tag{3.6}$$

By investigating EFR sample images, it is found that more than 50% of image energy is concentrated on $f_1(r,\theta)$ and $f_{-1}(r,\theta)$, indicating that $f_1(r,\theta)$ and $f_{-1}(r,\theta)$ have the most discrimination capability. While the MDI magnetogram is real valued, the real part of $f_1(r,\theta)$ is identical to that of $f_{-1}(r,\theta)$, and the imaginary part of $f_1(r,\theta)$ is equal to the negative imaginary part of $f_{-1}(r,\theta)$. This means the shapes of $f_i(r,\theta)$ and $f_{-1}(r,\theta)$ are same, and the only difference is the phase. Therefore, $f_1(r,\theta)$ is used instead of $f_{-1}(r,\theta)$.

To achieve the highest correlation peak and rotation invariance, a proper expansion center is chosen for CH decomposition. The center correlation of the CH filter

is determined by a correlation peak. A proper expansion center is chosen to ensure the center correlation to be a peak (Premont & Sheng 1993). In most cases, the proper expansion center is different from the geometric center of the reference image, and there are several ways of determining it (Premont & Sheng 1993; Garcia-Martinez *et al.* 1995). The proper expansion center is chosen through the CH energy map with the peak to the correlation energy map. The CH energy map is defined as

$$E(x,y) = f_m \circ f_m^*(x,y) = 2\pi \int_0^\infty f_m(r;x,y) f_m^*(r;x,y) r dr \,. \tag{3.7}$$

The pixel value at each location is the center auto-correlation intensity. The maximum peak is chosen as the expansion center.

To simplify the computational complexity, the magnitude of $f_1(r)$ is taken at each radial location. The phase information is discarded because it is assumed that the phase of $f_1(r)$ at each radial location is identical. The next step is to combine the CH functions of all the EFR sample images. They can not be combined directly because the flux magnitude and the size of the CH functions are in different scales. Figure 3.5 (a) shows the CH functions computed from the EFR sample images. The normalization of CH functions is performed by Mellin transformation (Ren *et al.* 2003) before combining them. It is assumed that all the computed functions are derived from the same base function, and the differences between these functions and the base function are the magnetic flux magnitude and size scales. Each computed CH function is defined as

$$f_m(r) = k_2 f_m'(r/k_1), \tag{3.8}$$

where $f_m'(r)$ is the base function, k_1 is the size scaling factor, and k_2 is the flux magnitude scaling factor. Then, the transformation is computed as follows:

$$T_0 = \int_0^{\infty} f_m(r)dr = k_1 k_2 \int_0^{\infty} f_m'(r)dr \ , \tag{3.9}$$

$$T_1 = \int_0^{\infty} f_m(r)rdr = k_1^2 k_2 \int_0^{\infty} f_m'(r)rdr \ . \tag{3.10}$$

The size scaling factor k_1 can be normalized by T_1/T_0, and the flux magnitude scaling factor k_2 can be normalized by T_0^2/T_1. The normalized CH functions are shown in Figure 3.5(b). The average value is taken at each r instance, which is shown in Figure 3.5(c). Figure 3.5(d) shows the averaged CH function computed from 6000 real EFR samples, of which the maximum is normalized to 1.

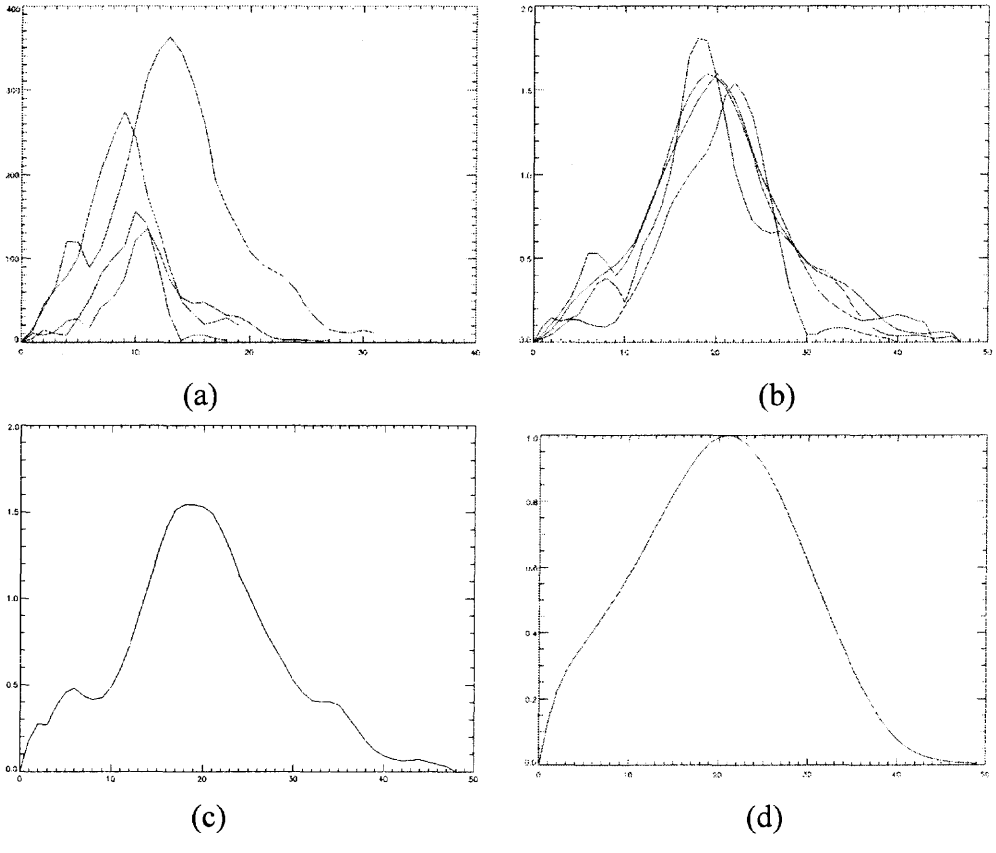
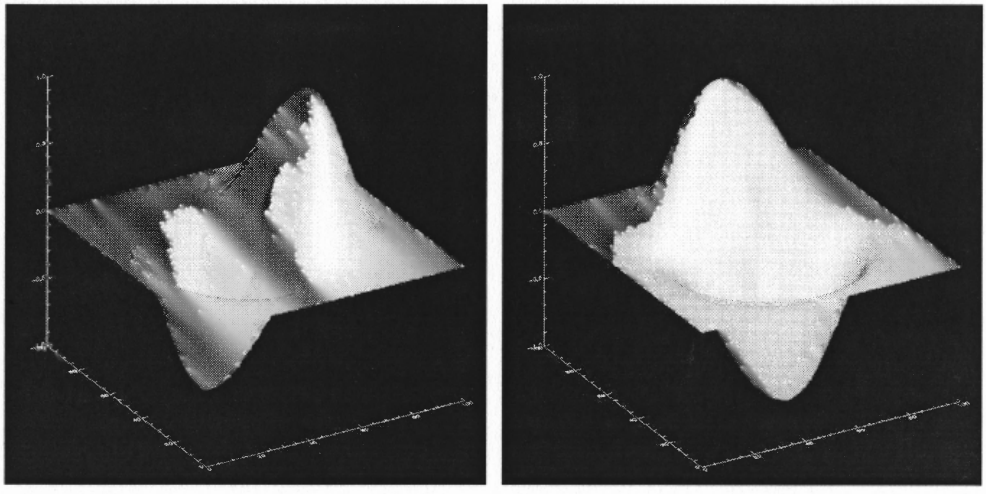


Figure 3.5 (a) Original, (b) normalized CH functions, (c) averaged CH function, and (d) normalized CH function computed from 6000 real EFR samples.

The filter is defined as: $f_r(r,\theta) = \overline{f}_1(r)\exp(j\theta)$, where $\overline{f}_1(r)$ is the resulting averaged CH function obtained in the last step as shown in Figure 3.5 (d). The real and imaginary parts of the resulting filter are shown in Figure 3.6 (a) and (b), respectively.



(a) Real part of the filter (b) Imaginary part of the filter

Figure 3.6 The resulting filter derived from the sample EFRs.

Since the EFRs could be of different sizes, they are matched with multi-scale similar filters. The sizes of the filters are defined as

$$s_i = 16 \times 2^{i/4}, \tag{3.11}$$

where $i = 0,1,...,10$. The sizes of the filters are from 16×16, 19×19 until 90×90. The filter of size s_i is defined as

$$f_r(r,\theta;s_i) = \frac{1}{s_i^2} f_r(\frac{r}{s_i},\theta), \tag{3.12}$$

where $f_r(r,\theta)$ is the filter obtained in the previous step and serves as the base filter, and $f_r(r,\theta;s_i)$ is the rescaled version of size s_i.

The designed filters of all the sizes are applied onto the MDI magnetogram and obtain a set of the correlation response images. Figure 3.7 shows the magnitude of the correlation between the MDI magnetogram shown in Figure 3.1 (a) and the filter of size 32×32. The phase of the correlation, which is not shown in the figure, indicates the rotation angle between the object and the filter at each pixel location.

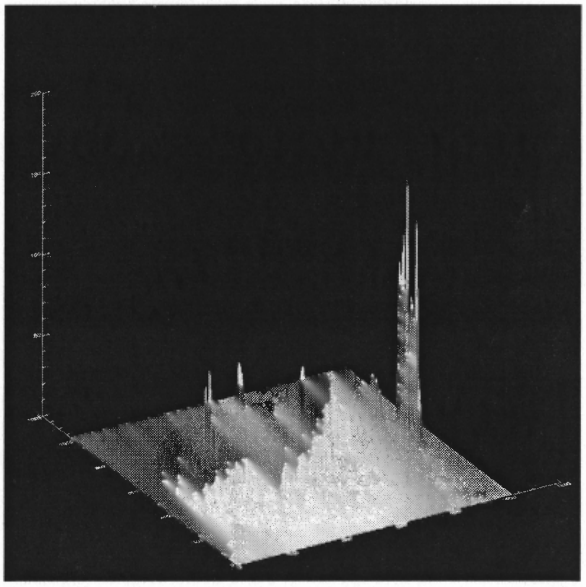


Figure 3.7 The magnitude of the correlation between the MDI magnetogram shown in Figure 3.1(a) and the filter of size 32×32.

3.3.2 Local Positive-to-Negative Flux Ratio

Some correlation peaks may happen in the positive or negative flux dominant regions. Such peaks should be suppressed since each EFR contains both positive and negative magnetic flux concentrations. Therefore, the *local positive-to-negative-flux ratio* (LPNFR), denoted as $r_{\Phi}(x, y; s_i)$, is defined at each pixel as

$$r_{\Phi}(x, y; s_i) = 1 - \frac{\left|\Phi_{+}(x, y; s_i) - \Phi_{-}(x, y; s_i)\right|}{\left|\Phi_{+}(x, y; s_i) + \Phi_{-}(x, y; s_i)\right|}, \tag{3.13}$$

where $\Phi_+(x,y;s_i)$ and $\Phi_-(x,y;s_i)$ are the total positive and negative fluxes in the neighboring area represented by $\{(x',y') \mid \sqrt{(x'-x)^2+(y'-y)^2} \le s_i\}$. The calculation can be conducted in the frequency domain for speedup. A sample of the LPNFR map is shown in Figure 3.8, where the bright parts correspond to the positive and negative flux balanced regions based on the filter of size 32×32.

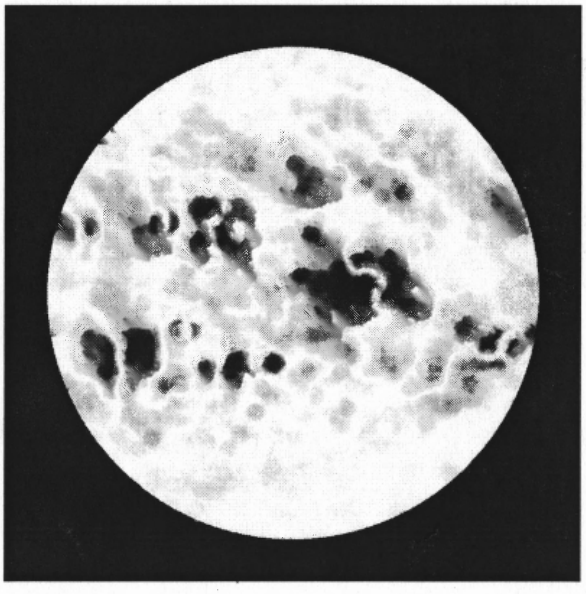


Figure 3.8 Positive-and-negative-flux ratio map of the MDI magnetogram shown in Figure 3.1 (a), with the filter of size 32×32.

3.3.3 Relative Energy Ratio

The peaks that the correlation is evaluated based on the complete object in the correlation response image are of interests. However, some peaks are generated even if the filter covers the object partially. To suppress such peaks, the *relative energy ratio* (RER), denoted as $r_E(x,y;s_i)$, is defined as

$$r_E(x,y;s_i) = \exp\left(-\left(1-\frac{E(x,y;s_i)}{E(x,y;s_{i+1})}\right)^2 / 2\delta^2\right), \tag{3.14}$$

where $E(x, y; s_i)$ is the energy of the MDI magnetogram in the neighboring area of (x, y) with size s_i, and δ is a controlling parameter. If a larger δ is chosen, the more correlation peaks are preserved; otherwise more suppressed. According to the experiments, δ is set to 0.1. The $E(x, y; s_i)$ is defined as

$$E(x, y; s_i) = \sum_{(x'-x)^2+(y'-y)^2 \le s_i^2} |g(x', y')|, \tag{3.15}$$

where g denotes the MDI magnetogram. Figure 3.9 shows the RER map of the MDI magnetogram shown in Figure 3.1 (a), in which $s_i = 32$.

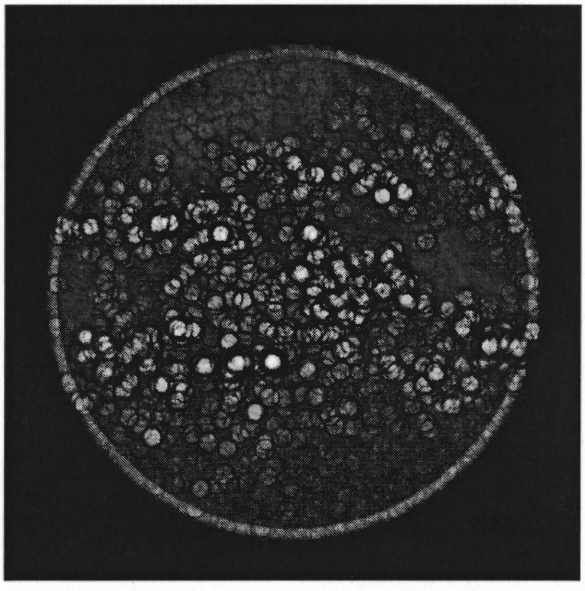


Figure 3.9 Relative energy ratio map of the MDI magnetogram shown in Figure 3.1 (a), with the filter of size 32×32.

3.3.4 Combined Response

After applying the filter of size s_i, $C(x, y; s_i)$, $r_\Phi(x, y; s_i)$, and $r_E(x, y; s_i)$ are determined. Then, they are multiplied together to be the combined response $D(x, y; s_i)$ corresponding to the filter as follows:

$$D(x, y; s_i) = C(x, y; s_i) r_{\Phi}(x, y; s_i) r_E(x, y; s_i). \quad (3.16)$$

Figure 3.10 shows the combined response image of the filter of size 32×32, where quite a lot peaks found in Figure 3.7 have been suppressed because of the low LPNFR and RER. After applying the filters of all sizes, a set of combined responses are obtained.

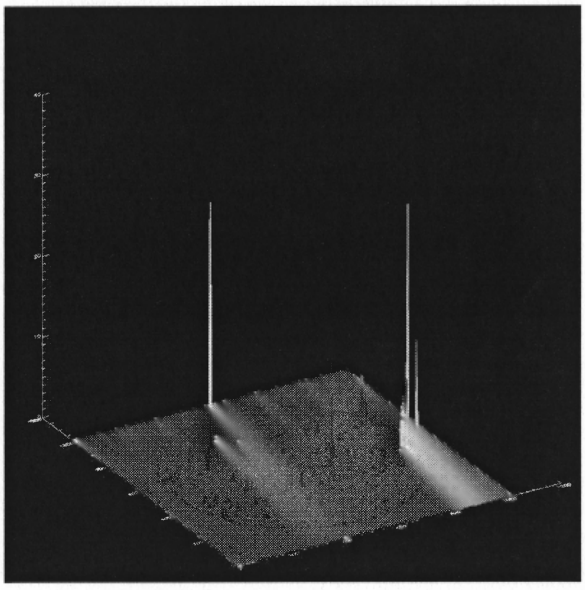


Figure 3.10 The combined response image of the filter of size 32×32.

3.3.5 Bipolar Region Extraction

Bipolar regions are extracted from each combined response image by the local maxima filter, whose size is equal to the one of the filter generating the combined response, and a global threshold η_1. The peaks, of which the magnitude is greater than η_1, are selected. In the experiments, η_1 is set to 6. The peaks picked up from the combined response image are shown in Figure 3.11. A larger threshold would lead to lose small bipolar regions which could be EFRs in the initial status, while a smaller threshold would pick up more small noisy bipolar regions. After combining all the responses, a set of extracted peaks are obtained. The corresponding bipolar regions are then determined based on these

peaks, because their centers are located at these peaks and their sizes are equal to the sizes of the filters generating these peaks. The bipolar regions are further investigated to trace the candidate EFRs.

Figure 3.11 The peaks picked up from the combined response image in Figure 3.10.

3.4 Tracing Candidate Objects

There are many popular tracking algorithms that can be used in tracking multiple targets (Bar-Shalom 1978; Bar-Shalom & Fortmann 1988; Musicki *et al.* 1994; Musicki & Evans 2002). The Global Nearest Neighbor method is combined with Kalman filter to trace the candidate EFRs. The global nearest neighbor method is carried out in the state space, and the bipolar region closest to the predicted state of the candidate EFR is picked up to update the track. The Kalman filter is used to estimate and predict the state of the candidate EFR.

Kalman filter was developed by Kalman (1960). It is basically a recursive filter to estimate the future variation of a dynamic system from a sequence of incomplete and

noisy measures. It has been applied in research and applications in signal processing, automatic control, and radar tracking extensively. More references could be found in (Welch & Bishop 2001; Maybeck 1979; Sorenson 1970; Grewal & Andrews 1993).

Each candidate EFR is assigned with an individual Kalman filter to evaluate its state. The candidate EFR either evolves to one of newly extracted bipolar regions or disappears. For each candidate EFR, the probability that a certain bipolar region is its new observation is evaluated by Kalman filter. The bipolar region with the highest probability is picked up as the new observation of the candidate EFR. If a bipolar region, not picked up for candidate EFRs, is distinguished enough from others, a new tracking will be initiated for it.

3.4.1 Background of Kalman Filter

Kalman filter, based on linear dynamical systems discredited in the time domain, is used to estimate the state of an object. The state is represented by a vector of real numbers. Let $x_{i,j}$ denote the state vector, in which i is the global index of the object and j is the time step. The evolution of the object is governed by the following linear stochastic difference equation (Welch & Bishop 2001):

$$x_{i,j} = A_{i,j}x_{i,j-1} + B_{i,j}u_{i,j} + w_{i,j}. \tag{3.17}$$

Here, $A_{i,j}$ is a transformation matrix, which relates the state at the previous time step j-1 to the state at the current step j, and it may vary with time steps or objects. The matrix $B_{i,j}$ relates the optional control input $u_{i,j}$ to the state $x_{i,j}$. $w_{i,j}$ is a zero-mean Gaussian noise with the covariance $Q_{i,j}$. Let $z_{k,j}$ denote the k-th measurement at time step j. If

$z_{k,j}$ is the observation of the *i*-th object at time step *j*, it is supposed to be the sum of a linear projection of $x_{i,j}$ and a Gaussian noise $v_{i,j}$, defined as:

$$z_{k,j} = Hx_{i,j} + v_{i,j}, \tag{3.18}$$

where H is the projection matrix and $v_{i,j}$ is a zero-mean Gaussian noise with the covariance $R_{i,j}$.

In practice, usually it is impossible to obtain the exact value of the state $x_{i,j}$, so it has to been estimated based on the observations. Let $\widetilde{x}_{i,j}$ be the state estimate of the *i*-th object at the time step *j*. It is assumed that the difference between $x_{i,j}$ and $\widetilde{x}_{i,j}$ is a zero-mean Gaussian distribution, and let $P_{i,j}$ denote the error covariance matrix. Thus, $p(x_{i,j}) \sim N(\widetilde{x}_{i,j}, P_{i,j})$.

There are two distinct phases while applying the Kalman filter: predict and update. In predict phase, $\widetilde{x}_{i,j}$ and $P_{i,j}$ are predicted based on the priori knowledge, and in update phase, they are updated based on the new observations.

In the predict phase, the priori state estimate is denoted as $\widetilde{x}_{i,j}^{-}$, and it is estimated as follows:

$$\widetilde{x}_{i,j}^{-} = A_{i,j}\widetilde{x}_{i,j-1} + B_{i,j}u_{i,j}. \tag{3.19}$$

Let $P_{i,j}^{-}$ denote the priori error covariance estimate, and it is calculated as follows:

$$P_{i,j}^{-} = A_{i,j}P_{i,j-1}A_{i,j}^{T} + Q_{i,j}. \tag{3.20}$$

In the update phase, as the new observations are available, the state estimate $\widetilde{x}_{i,j}$ and the error covariance estimate $P_{i,j}$ should be updated, as follows:

$$\tilde{x}_{i,j} = \tilde{x}_{i,j}^{-} + K_{i,j}(z_{k,j} - H\tilde{x}_{i,j}^{-}), \tag{3.22}$$

$$P_{i,j} = (I - K_{i,j}H)P_{i,j}^{-}, \tag{3.23}$$

where $K_{i,j} = P_{i,j}^{-}H^{T}(HP_{i,j}^{-}H^{T} + R_{i,j})^{-1}$.

3.4.2 Applying Kalman Filter

To apply Kalman filter, it is necessary to define the following variables: $x_{i,j}$, $A_{i,j}$, $B_{i,j}$, $u_{i,j}$, $Q_{i,j}$, $z_{k,j}$, H and $R_{i,j}$. The state vector $x_{i,j}$ is defined as:

$$x_{i,j} = (\theta, \beta, \Phi^{+}, \Phi^{-}, v^{\theta}, v^{\beta}, g^{+}, g^{-})^{T}, \tag{3.24}$$

where θ and β are the latitude and longitude of the bipolar region center in the spherical coordinate system, Φ^{+} and Φ^{-} are the sum of the positive and negative pixel values within the bipolar region in the MDI magnetogram, v^{θ}, v^{β}, g^{+} and g^{-} are the first derivatives of θ, β, Φ^{+} and Φ^{-} with respect to time, respectively. The transformation matrix $A_{i,j}$ is defined as a 8×8 matrix defined as follows:

$$A_{i,j} = \begin{bmatrix} 1 & 0 & 0 & 0 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & \Delta t & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & \Delta t \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{3.25}$$

where Δt is the time interval between two consecutive MDI magnetograms in seconds. In this application, the matrix $B_{i,j}$ and the input $u_{i,j}$ vector are not necessary, so they are

ignored. Note here it is not necessary to calculate the solar differential rotation angle (Song & Wang 2005) explicitly, instead it is evaluated by Kalman filter implicitly.

In eq. (3.17), the noise $w_{i,j}$ represents the accuracy of the linear dynamics model, and the larger covariance means that the model is less accurate, while smaller covariance means that it is more trustable. For simplification, it is assumed the distribution of $w_{i,j}$ is identical for all the candidate EFRs at any the time step. So the subscripts of $w_{i,j}$ and the covariance matrix $Q_{i,j}$ are dropped in the following discussion. In the experiments Q is set to

$$\begin{bmatrix}
7.01e-6 & -5.15e-9 & -2.90e-1 & -2.06e-1 & 6.05e-10 & -2.05e-12 & -1.32e-5 & -3.17e-5 \\
-5.15e-9 & 1.03e-8 & -1.16e-2 & -5.81e-3 & -1.79e-12 & 8.73e-13 & -5.46e-7 & -3.13e-7 \\
-2.90e-1 & -1.16e-2 & 7.14e+5 & 3.77e+5 & -7.77e-6 & -3.41e-7 & 5.33e+1 & 3.35e+1 \\
-2.06e-1 & -5.81e-3 & 3.77e+5 & 6.19e+5 & -2.62e-5 & -2.83e-7 & 3.28e+1 & 5.76e+1 \\
6.05e-10 & -1.79e-12 & -7.77e-6 & -2.62e-5 & 7.70e-14 & -2.66e-16 & -1.31e-9 & -4.91e-9 \\
-2.05e-12 & 8.73e-13 & -3.41e-7 & -2.83e-7 & -2.66e-16 & 1.10e-16 & -5.52e-11 & -4.43e-11 \\
-1.32e-5 & -5.46e-7 & 5.33e+1 & 3.28e+1 & -1.31e-9 & -5.52e-11 & 6.41e-3 & 4.28e-3 \\
-3.17e-5 & -3.13e-7 & 3.35e+1 & 5.76e+1 & -4.91e-9 & -4.43e-11 & 4.28e-3 & 7.55e-3
\end{bmatrix}$$

.

Like the state vector $x_{i,j}$, the observation vector $z_{k,j}$ is also defined as

$$z_{k,j} = (\theta, \beta, \Phi^+, \Phi^-, v^\theta, v^\beta, g^+, g^-)^T, \tag{3.26}$$

where θ, β Φ^+ and Φ^- are measured from the k-th bipolar region directly, while v^θ, v^β, g^+ and g^- can not be determined in the absence of the previous observation. But when the i-th candidate EFR is concerned, suppose $z_{i,j-1}$ is its previous observation at time step j-1, and then the values of v^θ, v^β, g^+ and g^- are computed as follows:

$$z_{k,j}(4:7) = \frac{1}{\Delta t}\left(z_{k,j}(0:3) - z_{i,j-1}(0:3)\right), \tag{3.27}$$

where $z_{k,j}(4:7)$ represents the last four components of $z_{k,j}$, while $z_{k,j}(0:3)$ and $z_{l,j-1}(0:3)$ are the first four components of $z_{k,j}$ and $z_{l,j-1}$ respectively.

The projection matrix H is simply set to a 8×8 identify matrix in this application. As for $R_{i,j}$, it is also assumed the distribution of observation noise $v_{i,j}$ is identical for all the candidate EFRs at any the time step, so the subscripts of $v_{i,j}$ and its covariance matrix $R_{i,j}$ are also dropped. In this experiments, the matrix R is set to

$$\begin{bmatrix}
3.93e\text{-}6 & 7.47e\text{-}8 & 8.31e{+}1 & 4.88e{+}1 & 3.14e\text{-}11 & 1.40e\text{-}11 & 4.51e\text{-}5 & 5.91e\text{-}5 \\
7.47e\text{-}8 & 3.88e\text{-}6 & -7.66e\text{-}1 & -2.35e\text{-}1 & 1.31e\text{-}11 & 3.41e\text{-}10 & -7.29e\text{-}5 & -4.31e\text{-}5 \\
8.31e\text{-}1 & -7.66e\text{-}1 & 1.18e{+}8 & 7.31e{+}7 & 8.81e\text{-}5 & -5.59e\text{-}5 & 6.51e{+}3 & 3.69e{+}3 \\
4.88e\text{-}1 & -2.35e\text{-}1 & 7.31e{+}7 & 1.24e{+}8 & 9.98e\text{-}5 & -3.59e\text{-}5 & 3.64e{+}3 & 6.80e{+}3 \\
3.14e\text{-}11 & 1.31e\text{-}11 & 8.81e\text{-}5 & 9.98e\text{-}5 & 1.16e\text{-}13 & 3.64e\text{-}15 & -1.73e\text{-}8 & -1.97e\text{-}8 \\
1.40e\text{-}11 & 3.41e\text{-}10 & -5.59e\text{-}5 & -3.59e\text{-}5 & 3.64e\text{-}15 & 1.19e\text{-}13 & -2.11e\text{-}8 & -1.33e\text{-}8 \\
4.51e\text{-}5 & -7.29e\text{-}5 & 6.51e{+}3 & 3.64e{+}3 & -1.73e\text{-}8 & -2.11e\text{-}8 & 2.24 & 1.25 \\
5.91e\text{-}5 & -4.31e\text{-}5 & 3.69e{+}3 & 6.80e{+}3 & -1.97e\text{-}8 & -1.33e\text{-}8 & 1.25 & 2.33
\end{bmatrix}.$$

The values of matrices Q and R are both determined based on the statistics.

3.4.3 Updating State of Candidate EFRs

To determine the new state of a certain candidate EFR, its state is first predicted by applying Kalman filter. Then the Mahalanobis distance between it and all the newly extracted bipolar regions is calculated, and then pick up the one closest to it as its new observation at the current time step. Finally, the state for the candidate EFR is updated based on the new observation.

There are two assumptions: 1) a newly extracted bipolar region can be assigned to at most one candidate EFR; 2) a candidate EFR either evolves to a newly extracted bipolar region, or disappears forever. As for assumption 1, since the EFRs are distributed

over the solar disk sparsely, it should hold in most cases. As for the assumption 2, it is found that once the track of an EFR is initiated, it will seldom be lost until the disappearance. Therefore, the "lost-and-return" phenomena are not considered.

As mentioned above, first the next observation of current candidate EFRs should be determined. It is done based on the Mahalanobis distance between the candidate EFR and each newly extracted bipolar region. The distance is denoted as $M_{i,j}(k)$, where k and i are the indexes of the bipolar region and candidate EFR respectively, and j is the time step. Based on $\tilde{x}_{i,j}^{-}$ and $z_{k,j}$, the distance $M_{i,j}(k)$ is calculated as follows:

$$M_{i,j}(k) = \sqrt{(\tilde{x}_{i,j}^{-}(0:3) - z_{k,j}(0:3))^T W^{-1} (\tilde{x}_{i,j}^{-}(0:3) - z_{k,j}(0:3))}, \tag{3.31}$$

where W is the covariance matrix. Based on statistics, in the experiments, it is set to

$$\begin{bmatrix} 1.09e-5 & 6.95e-8 & 5.41e+1 & 2.82e+1 \\ 6.95e-8 & 3.89e-6 & -7.77e-1 & -2.40e-1 \\ 5.41e-1 & -7.77e-1 & 1.19e+8 & 7.34e+7 \\ 2.82e-1 & -2.40e-1 & 7.34e+7 & 1.25e+8 \end{bmatrix}.$$

The bipolar region closest to the candidate EFR will be assigned to the candidate EFR as the new observation. But if the distance is greater than the threshold η_2, which is set to 50 in the experiments, the assignment will be invalid and the disappearance of the candidate EFR will be reported; otherwise, the state vector and error covariance matrix of the candidate EFR will be updated based on the new observation by using eq. (3.22) and (3.23).

It is possible that two or more candidate EFRs get closer and closer and even evolve to the same bipolar region. If so, only one candidate EFR will be preserved, and

all the others will be dropped and not traced any more. Similar to eq. (3.31), the distance between two candidate EFRs is defined as follows:

$$D_j(i_1, i_2) = \sqrt{(\tilde{x}_{i_1,j}(0:3) - \tilde{x}_{i_2,j}(0:3))^T W^{-1} (\tilde{x}_{i_1,j}(0:3) - \tilde{x}_{i_2,j}(0:3))}\,. \tag{3.32}$$

If the distance is less than the threshold η_2, the candidate EFRs are too close. Then one of them should be selected for further tracking and dropped the others. The selection criterion is based on the average drift distance of the *i*-th candidate EFR, denoted as $\tilde{M}_{i,j}$, and it is defined as

$$\tilde{M}_{i,j} = p\tilde{M}_{i,j-1} + (1-p)M_{i,j}\,, \tag{3.33}$$

where $M_{i,j} = \min_k(M_{i,j}(k))$ and p is set to 0.5 in the experiments. If p is smaller, $\tilde{M}_{i,j}$ is more sensitive to $M_{i,j}$; otherwise $\tilde{M}_{i,j}$ trusts its previous values more. The candidate EFR with the smallest $\tilde{M}_{i,j}$ will be preserved.

3.4.4 Identifying New Candidate EFRs

After the states of all candidate EFRs have been updated, some bipolar regions are picked up as new observations of canididate EFRs. All the remaining bipolar regions are checked for new tracking. Based on the covariance W, the distance between a bipolar region and all the other regions, including the existing candidate EFRs and other bipolar regions, can be calculated similar as eq. (31) or (32). If the minimum distance is greater than the threshold η_3, which is set to 500 in the experiments, a new track will be initiated and the bipolar region will be traced as a new candidate EFR. The first four components of its $\tilde{x}_{i,j}$ are measured from the bipolar region and the last four components

are all set to zeros. The initial error covariance $P_{i,j}$ can be set to an arbitrary large value, and it is set to the sum of the matrix Q and R in the experiments. The average drift distance $\widetilde{M}_{i,j}$ can be set to an arbitrary large value, and it is set to η_3 in the experiments.

3.5 Classification of Candidate Objects

After the complete observation of a candidate EFR is available, a feature vector is defined based on the properties measured during the observation time. A Support Vector Machine (Vapnik 1998) classifier is then applied to distinguish the EFR from all the other bipolar regions.

3.5.1 Property Measurement

Each candidate EFR is traced in consecutive frames from its initial appearance until it disappears from the field of view. In each frame, the properties are measured, such as location, orientation, distance between two polarities, and the positive and negative magnetic fluxes. Thus, the corresponding time profiles of proprieties can be constructed for further investigation.

The time span of growing period is critical to distinguish the EFR from other regions, as well as the property profiles during the growing period. Therefore, the growing time should be determined from the time profile of total magnetic flux. Typically, the total magnetic flux increases from the beginning to the maximum and then decreases. Its time profile is convolved with a Gaussian function with the standard deviation of 1.5, and then the maximum is picked up as the end of the growing period.

Thus, the growing period is determined. The following properties of each candidate EFR are measured during the growing period:

a. Time span of the growing period, denoted as t.

b. Growing rate of total flux magnitude, denoted as k_f.

c. Initial relative energy ratio, denoted as r_E^0. It is the mean value of relative energy ratio measured from the first three frames.

d. Initial total flux magnitude, denoted as f_0. It is the mean value of the total flux magnitude measured from the first three frames.

e. Initial distance between two polarities, denoted as d_0. It is the mean value of the distance between tow polarities measured from the first three frames.

f. Correlation between the positive and negative magnetic fluxes, denoted as ρ_-^+. It is the correlation between the positive and negative flux magnitudes during the growing period.

g. Mean local positive to negative flux ratio, denoted as $\overline{r_\Phi}$.

h. Positive to negative flux growing rate ratio, denoted as r_k. It is defined as $r_k = \frac{2k_+k_-}{k_+^2 + k_-^2}$, where k_+ and k_- are the growing rates of positive and negative flux magnitude respectively.

i. Growing rate of the distance between two polarities, denoted as k_d.

j. Maximum total flux magnitude, denoted as $f_{\max}$

k. Mean relative energy ratio, denoted as $\overline{r_E}$.

Hence, an eleven-dimensional vector $v = (t, k_f, r_E^0, f_0, d_0, \rho_-^+, \overline{r_\Phi}, r_k, k_d, f_{\max}, \overline{r_E})$ is defined for the candidate EFR.

3.5.2 Object Classification

The support vector machine (SVM) is applied to classify the candidate EFRs. The SVM, introduced by Vapnik (1998), is based on statistics learning theory for the two-class classification problem. The idea is to map the input patterns to a high-dimensional feature space and construct an optimal hyper-plane to separate all the patterns (Guyon & Stork 2000). Different kernel functions can be used, such as linear, polynomial, sigmoid, and radial basis function. The polynomial of degree 2 is chosen as the kernel function in the experiments.

The classifier requires training prior to testing. Each candidate EFR in the training set is represented as the feature vector defined above and is classified by human inspection. A label is associated to each candidate EFR. Then, the SVM classifier takes the input patterns and the associated class labels to compute the optimal decision hyper-plane. After that, the SVM classifier is ready to test an unknown bipolar region.

3.6 Experimental Results

The method is implemented in the RSI Interactive Data Language (IDL) together with the Solar SoftWare (SSW) IDL library (Freeland & Handy 1998). The IDL program runs on a Dell Dimension 4600 with an Intel Pentium 4 processor (2.8 GHz) and 1.0 GB memory under Fedora Core Linux 4.0.

The MDI magnetograms in 2001 and 2002 are used as the data set in the experiments. The program is first applied on all the magnetograms, resulting 3234 candidate EFRs detected, including 186 EFRs and 3058 non-EFRs. This list does not include the objects that exist in less than 9 hours or of which the maximum total magnetic

flux is less than 3.15×10^{21} Mx, which is verified visually. All the included objects are classified by human inspection. A feature vector is assigned to each detected object.

Next, half of samples are picked up randomly for training and the remaining for testing. That is, 93 EFRs and 1529 non-EFRs are picked up randomly for training, and the others are used for testing. The experiment is repeated 100 times and the average classification rates are computed. The results show that the detection rates of EFRs and non-EFRs are 96.4% and 98.0%, respectively. But since the number of non-EFRs is much larger than that of EFRs, among the EFR events reported by the program, only 74.3% of them are true EFRs, which means the false alarm rate is 25.7%.

The computational time can be divided into several parts. First, it takes 0.015 seconds to conduct the pre-processing and 40 seconds to process a single frame and extract bipolar regions. The time to trace the candidate EFRs and measure the associated properties depends on the number of candidate EFRs detected. Typically, it takes 1.5 seconds to process the objects in a frame, if the disk I/O is not considered, which can be avoided by implementing the detection method in on-the-fly way. The SVM training is performed off-line. The time to classifying an object takes less than 1 second. Therefore, it takes approximately about 45 seconds to process each frame, indicating that 68 hours are needed to process all the magnetograms in a year.

3.7 Discussion and Conclusions

In this chapter, a method is presented to automatically detect emerging flux regions by using continuous MDI magnetograms. The experimental results show that the method can be used to automatically detect EFRs based on consecutive MDI magnetograms

effectively. In most cases, the circular filters captured the most EFRs' centers and sizes correctly, and Kalman filter worked effectively to estimate and predicate the states of candidate EFRs.

By analyzing the misclassified objects, it is found that the main reason is that the diversity of the EFRs' shape caused the circular harmonic filters of the same shape to be less effective to detect the EFRs of irregular shape. It means if the shape of an EFR is irregular, its center would be either lost or mis-determined, and thus the associated properties could not be measured correctly. To address this problem, the joint composite filter will be considered to match EFRs of diverse shapes more closely. The composite filter may be more important for detecting active regions, of which the shape diversity is much more than that of EFRs.

Another reason is that it was noticed that the "lost-and-return" happened in the experiments, though it did not happen frequently, especially when the significant EFRs are traced. When the "lost-and-return" happens, the track of the candidate EFR would be interrupted and the property time profiles could not be continuous and accurate, which also leads to misclassification. To address this problem, Joint Probability Data Association (JPDA) rule should be considered, which is more robust than the Global Nearest Neighbor method used currently.

Besides, the complex background around the object may also degrade the classification rate. Because if the object is located in complex background and there are other magnetic concentrations around it, the false correlation peaks may be generated based on un-matched concentrations and the true peaks could then be overwhelmed.

In the future, the improvement will be made in the following aspects. First, the joint composite filter would be considered to match object more closely. This method will be also extended to detect active region and monitor the relationship between the EFR and its nearest active region, since it is of interest for space weather predication study. Second, other tracking algorithm, such as IPDA, JPDA and JIPDA, will be investigated to improve the track performance. Third, alternative features will be evaluated to continue improving the classification rates.

CHAPTER 4

SOLAR FLARE PREDICTION

4.1 Introduction

Solar flares and coronal mass ejections (CMEs) are the most important solar explosive phenomena as far as space weather effects are concerned. Although physical mechanisms of energy buildup and release are not yet fully understood, from the observational point of view, flares and CMEs originate preferentially in magnetic regions with strong field gradients and magnetic shear, long polarity-inversion lines and complex polarity patterns (e.g., Falconer 2001; Falconer et al. 2003; Song et al. 2006; Wang et al. 2006; Schrijver 2007, etc.).

Forecasting the occurrence of solar flares from the analysis of solar photospheric magnetogram data is an important and challenging task in the solar physics research. The earlier predictions were more based on McIntosh classification of sunspots (McIntosh 1990) or relying on observer's experience (e.g., Zirin and Marquette, 1991). In recent years, the statistical correlation between a variety of photospheric magnetic parameters and flare productivity of source regions have been investigated in many studies, and some parameters are found to be positively correlated with the probability of a region to produce a major flare (Falconer et al. 2002; Leka & Barnes 2003a, 2003b, 2007; Jing et al. 2006; Schrijver et al. 2005; Schrijver 2007; Song et al. 2007).

The purpose of this study is to develop a mathematical tool to predict the density function of flare peak intensity or soft X-ray flare index based on a set of magnetic parameters. It is proposed to divide this problem into two parts: the prediction of the occurrence probability of significant flares, such as X-, M- and C-class flares, and the

prediction of probability density function of flare peak intensity or flare index under the flare occurrence assumption. The first part of the problem can be solved by using ordinal logistic regression (Song *et al.* 2007) or kernel logistic regression (Wahba *et al.* 1995), since the output is dichotomous (0 or 1). The other part is more challenging, since its output should be the whole probability density function of flare peak intensity or flare index. This chapter is focused on the second part of the problem.

A novel two-step conditional density estimator is proposed in this chapter by combining kernel-based nonlinear regression and moment-based density function reconstruction techniques to solve the problem. Compared with traditional double-kernel density estimator, the new method can easily handle high dimensional data set, while it would be quite difficult to apply the double-kernel density estimator in this case. The capability of handling high dimensional data is essential in this problem, since scientists are looking for various magnetic parameters to predict solar flares, therefore, the parameter vector is multi dimensional, and it would be extended further when new data sources become available.

4.2 Dataset and Magnetic Parameters

The dataset includes 2972 active regions (NOAA 07961 to 10932) observed from 1996 to 2006. To predict the density function of solar flare peak flux (in W/m^2), 1017 C-, M- and X-class solar flare events were selected, if they are located in an active region, whose position angle is less than 30 degrees from the solar disk center. To predict the density function of solar flare index, 530 active regions close to the disk center and associated with at least one significant flare were selected and analyzed.

For each active region, three magnetic parameters, including total unsigned magnetic flux T_{flux}, energy dissipation E_{diss} and length of strong-gradient neutral line L_{gnl}. These parameters are chosen primarily because they can be derived directly from the MDI line-of-sight magnetograms and have historically shown their potential in flare forecasting (Jing *et al.* 2006; Song *et al.* 2007). The procedures of deriving three magnetic parameters are well described in Jing *et al.* (2006) and Song *et al.* (2007).

The flare peak intensity, denoted as *I*, was read from Solar Geophysical Data (SGD) flare reports. The flare index (in $10^{-6}\,\mathrm{W/m^2}$), denoted as *FI*, is defined as

$$FI = (100\sum_{\tau} I_x + 10\sum_{\tau} I_M + \sum_{\tau} I_C + 0.1\sum_{\tau} I_B)/\tau \tag{4.1}$$

by weighting the flare peak intensities of X-, M-, C- and B-class flares within the specified time window as 100, 10, 1 and 0.1, respectively (Abramenko 2005).

4.3 Mathematical Modeling

The solar flare prediction of the density function of either *I* or *FI* is modeled as the conditional density estimation (CDE) problem. Given a set of magnetic parameters, either *I* or *FI* is treated as a random variable, and then try estimating the associated probability density function.

The task of CDE is to estimate the probability density function of a random variable *y*, denoted as $f_{Y|X}(y \mid x)$, give a specific value of *x*, where $y \in R$ and $x \in R^d$. In this application, *y* is either *I* or *FI* and *x* is the magnetic parameter vector. Usually, start with a sequence of observations $(X_1, Y_1), \ldots, (X_n, Y_n)$, and these observations will be used to train the model. The CDE can be considered as the generalization of regular

regression, whose goal is to estimate conditional mean, and quantile regression, which aims to model conditional quantiles. The conditional density function plays a critical role in various prediction problems, since it provides a complete description of the stochastic behavior of the response variable based on its covariates.

The CDE problem was first introduced by Rosenblatt (1969), and typically this problem is solved by non-parametric methods, and the best well-known conditional density estimators are named as the double-kernel estimators, including Nadaraya-Watson estimator and its extensions (Nadaraya 1964; Watson 1964; Hyndman *et al.* 1996; Hall *et al.* 1999; Fan *et al.* 1996; Fan & Yao 2003). In these kernel-based conditional density estimators, the kernel function plays a critical role, which is a symmetric weight function and serves as a smoothing factor. The choice of specific type function is not crucial, and usually a Gaussian function is selected to serve the purpose. On the other hand, the bandwidth is much more important, and there are several bandwidth selection rules that have been proposed in the literature to determine optimal bandwidths (Bashtannyk & Hyndman 2001; Fan & Yim 2004; Hall *et al.* 1999; Hall *et al.* 2004). These double-kernel estimators are used most widely and applied to a broad range of application problems, the success is due to the fact that little assumptions is necessary to be made about the parametric form of $f_{Y|X}(y \mid x)$, and thus these methods are quite flexible and capable of approximating any functions. It is very desirable, especially when handling complex data and little knowledge is available about the underlying density function. But the main problem of these methods is that their performance depends heavily on the selection of bandwidths. Unfortunately, the selection of good bandwidths is not a trivial task and usually it's difficult to determine the optimal values of bandwidth

based on finite samples (Pritsker 1998). It becomes worse when high dimensional covariates are involved (Pritsker 1998). So due to this fact, currently these methods seem to have confined to bivariate datasets, and are difficult to be applied in multivariate cases.

To address the problem of the double-kernel estimators, an alternative conditional density estimator is proposed to solve the problem. The method involves two main steps. First, the conditional moments are estimated based on nonlinear regression analysis, and then the conditional density function of the response variable is recovered based on the moments. To achieve the high accuracy of the estimation of conditional moments of response variable, machine learning techniques are adopted in the nonlinear regression, which can handle high dimensional data easily. It should be noted that the moments can uniquely determine its probability density function. Therefore, if a sufficient number of moments are available, its density function can be approximated based on these moments, which is known as a moment problem (Shohat & Tamarkin 1943; Akhiezer 1965).

4.4 Double-kernel Methods

Suppose that the observations (X_1, Y_1),..., (X_n, Y_n), where $Y_i \in R$ and $X_i \in R^d$, are available, and they are generated by a underlying joint density function $f_{X,Y}(x,y)$. It is assumed that they are strictly stationary, independent and identically distributed. The goal is to estimate the conditional density function $f_{Y|X}(y|x)$ based on the available observations. The scalar case of X is presented in this section, and the functions used in multivariate case should be similar and be easily deduced.

The conditional density function $f_{Y|X}(y|x)$ can be defined as:

$$f_{Y|X}(y \mid x) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \tag{4.2}$$

so if the marginal density function $f_X(x)$ and joint density function $f_{X,Y}(x,y)$ are estimated first, the conditional density function $f_{Y|X}(y \mid x)$ can be estimated as:

$$\hat{f}_{Y|X}(y \mid x) = \frac{\hat{f}_{X,Y}(x,y)}{\hat{f}_X(x)}, \tag{4.3}$$

where $\hat{f}_{X,Y}(x,y)$ and $\hat{f}_X(x)$ are the kernel estimators of $f_{X,Y}(x,y)$ and $f_X(x)$, respectively. By applying the kernel function, the $\hat{f}_{Y|X}(y \mid x)$ and $\hat{f}_X(x)$ can be estimated as follows:

$$\hat{f}_{X,Y}(x,y) = \frac{1}{n}\sum_{i=1}^{n} W_{h_1}(y - y_i) W_{h_2}(x - x_i), \tag{4.4}$$

$$\hat{f}_X(x) = \frac{1}{n}\sum_{i=1}^{n} W_{h_2}(x - x_i), \tag{4.5}$$

where h_1 and h_2 are bandwidths, W is a kernel function, and $W_h(x) = \frac{1}{h} W(\frac{x}{h})$. In general, the kernel function K should be a symmetric weight function and possess the following properties:

$$\int W(x)dx = 1, \tag{4.6}$$

$$\int xW(x)dx = 0. \tag{4.7}$$

The specific choice of the kernel function is not critical (Mays & Gamlin 1995), and usually a Gaussian density function is selected. By plugging the eq. (4.4) and (4.5) into eq. (4.3), it is changed to

$$\hat{f}_{Y|X}(y \mid x) = \frac{\sum_{i=1}^{n} W_{h_1}(y - y_i) W_{h_2}(x - x_i)}{\sum_{i=1}^{n} W_{h_2}(x - x_i)}, \tag{4.8}$$

which is known as Nadaraya-Watson estimator (Nadaraya 1964; Watson 1964), and is a classical kernel density estimator.

Fan *et al.* proposed a direct estimator based on local polynomial regression, considered as the extension of Nadaraya-Watson estimator (Fan *et al.* 1996). It is noted that as $h_1 \to 0$,

$$E\{W_{h_1}(Y - y) \mid X = x\} \approx f_{Y|X}(y \mid x), \tag{4.9}$$

and thus $f_{Y|X}(y \mid x)$ can be estimated by regressing $W_{h_1}(Y - y)$ on X. To apply local polynomial regression, $g(x;\alpha)$ is defined as a polynomial of degree p, in which α is a parameter vector $(\alpha_0, \alpha_1 ..., \alpha_p)$, such that

$$g(x;\alpha) = a_0 + a_1 x + a_2 x^2 + ... + a_p x^p. \tag{4.10}$$

The optimal value of α is defined as

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{n} \{W_{h_1}(Y_i - y) - g(X_i - x;\alpha)\}^2 W_{h_2}(X_i - x). \tag{4.11}$$

The details of compute $\hat{\alpha}$ are presented by Fan *et al.* (1996). Once the optimal value $\hat{\alpha}$ is obtained, $\hat{f}_{Y|X}(x, y) = \hat{\alpha}_0$. Thus, $f_{Y|X}(x, y)$ can be estimated by regressing $W_{h_1}(Y - y)$ on X.

Let $g(x;\alpha)$ be a linear function, which means $p = 1$ in eq. (4.10), and the alternative method based on local polynomial regression would be reduced to Nadaraya-Watson estimator, which is the reason that this method is considered as the extension of

Nadaraya-Watson estimator. In both of these methods, the kernel function is used twice, in both x and y direction, respectively, they are named as double-kernel estimators.

In double-kernel density estimation, the bandwidths are critical for the performance of the estimators, and it is trickier to select good bandwidths than kernel functions. There are several methods proposed in the literature to select good bandwidths, such as bootstrap, cross-validation or plug-in (Bashtannyk & Hyndman 2001; Fan & Yim 2004; Hall *et al.* 1999; Hall *et al.* 2004).

The nonparametric estimators are quite flexible, since they make little assumptions on the target conditional density function. The freedom from parametric assumptions makes it able to handle complicated density functions, so the double-kernel estimators have been applied in a very wide class of problems successfully.

The main limitation of double-kernel estimators is that the double-kernel estimators seem to have confined to the bivariate case up to now. The main reason for this limitation is the difficulty of selecting optimal bandwidths from finite data samples especially in the presence of high dimensional data. Although there are a couple of methods proposed to select bandwidths, it is still a tough task in many cases, and it becomes even worse in multivariate case. But, in many application problems, high dimensional covariates are essential, since they are more informative than one-dimensional covariates and help to reduce the uncertainty of the conditional density estimation. Due to the limitation of double-kernel estimators, they can not handle such multivariate data very well, and thus the estimation can not be done based on high dimensional covariates.

4.5 Moment-based Conditional Density Estimator

In this section, a new method is described to solve the CDE problem and overcome the main limitation of double-kernel estimators reviewed in the last section. This method involves two main steps. First, conditional moments of the response variable Y are estimated based on the given covariates, and then the condition density function of Y is constructed based on the estimated moments. To handle high dimensional covariates and achieve high accuracy, the machine learning techniques, which was introduced in Support Vector Machine (SVM) (Vapnik 1995; Vapnik 1998), and the kernel trick (Müller *et al.* 2001) are adopted to do nonlinear regression and estimate conditional moments. After that, the probability density function is constructed based the estimated moments (Shohat & Tamarkin 1943; Akhiezer 1965). There are a couple of available methods making use of Pearson Curves (Solomon & Stephens 1978), saddelpoint approximations (Reid 1988) or inverse Mellin transform (Mathai & Saxena 1978). In these methods usually exact moments are required; otherwise the reconstructed density function could be easily ruined by the estimation errors. Instead, a unified method is presented by Provost (2005), and approximated moments are required, which is more feasible in this application.

4.5.1 Normalization

Before proceeding to the next section, there are two suggestions for the implementation of conditional density estimation. First, it is suggested to preprocess the original observations as follows:

$$X_i' = \left[\frac{2(X_{i,1} - \min(X_{*,1}))}{\max(X_{*,1}) - \min(X_{*,1})} + 1, \ldots, \frac{2(X_{i,J} - \min(X_{*,J}))}{\max(X_{*,J}) - \min(X_{*,J})} + 1 \right]^T, \tag{4.12}$$

$$Y_i' = \frac{2(Y_i - \min(Y_*))}{\max(Y_*) - \min(Y_*)} + 1, \tag{4.13}$$

where $X_{i,j}$ is the j-th component of Y_i, $X_{*,j} = \{X_{1,j}, \ldots, X_{n,j}\}$ and $Y_* = \{Y_1, \ldots, Y_n\}$. Thus, each component of X and Y are mapped into [-1, 1]. Such mapping is not necessary theatrically but very helpful in practice, because the higher moments of Y could easily become intractable if we do not do the mapping.

4.5.2 Conditional Moment Estimation

As mentioned above, the first step is to estimate the conditional moments of the response variable Y, denoted as $E(Y^i)$, where i=0, 1,..., m, and clearly, $E(Y^0) = 1$. To handle the high dimensional covariates and achieve high accuracy, the nonlinear regression is applied based on kernel trick and estimate the condition moments of Y. Here, the kernel trick is a method for converting a linear regression algorithm into a non-linear one by using a non-linear function to map the original covariates into a higher-dimensional space, and this makes a linear regression in the new space equivalent to non-linear regression in the original space. The non-linear regression is more powerful than linear regression and capable of capturing complex relationship between the response variable and the covariates. Besides, please note that the kernel function in machine learning is different from the counterpart in statistics, such that it is a function used to calculate the dot product of two vectors.

Given a specific covariate X, it is proposed to estimate its conditional moments in the following two steps:

(1) Compute the feature vector $\Omega(X)$ by projecting X onto the principal components of the training dataset in the higher dimensional feature space, which are extracted by using Kernel Principal Component Analysis (KPCA) (Schölkopf *et al.* 1998).

(2) Apply the linear transform to determine the output. The weight vector used in the linear transform is determined by applying linear regression analysis on the training dataset. Since the linear transform is conducted in the higher-dimensional feature space, it is equivalent to the non-linear transform in the original space.

First of all, the KPCA should be applied to extract the principal components of the training dataset. The KPCA is the kernel version of the traditional Principal Component Analysis (Schölkopf *et al.* 1998). Suppose a sequence of observations, $(X_1, Y_1), \ldots, (X_n, Y_n)$, are avaiable. Consider a mapping function ϕ transforming the original vector X to a higher-dimensional vector $\phi(X)$, and thus construct a new space spanned on $\phi(X_1), \ldots, \phi(X_n)$. Then a kernel matrix K of dot products can be defined as

$$K = \Phi^T \Phi, \tag{4.14}$$

where $\Phi = [\phi(X_1), \phi(X_2), \ldots, \phi(X_n)]$. Let $K_{i,j}$ denote the element in the i-th row and j-th column of the matrix K, and $K_{i,j} = \langle \phi(X_i), \phi(X_j) \rangle$. Since the feature space is spanned on n sample vectors, the dimension of all the data lying in this space should be lower than n and equal to the rank of the kernel matrix K.

Next, a mapping function ϕ should be chosen to make the kernel matrix K fully defined. After a specific mapping function ϕ is chosen, the original X can be mapped into $\phi(X)$ and then element of K is computable. But the explicit mapping X into $\phi(X)$ should be avoided, because it is inefficient and unnecessary. Instead, a function should be applied to compute $\langle \phi(X_i), \phi(X_j) \rangle$ directly, which is far more efficient and flexible than using the explicit mapping, and such function is named as kernel function, denoted as k. Note that kernel function used here is totally different from that mentioned in the last section, which is a symmetric weight function. The implementation of the kernel function determines the exact form of the mapping function implicitly, and thus the inner products can be computed in the kernel space efficiently without explicit mapping of Xs.

Typically, there are many choices for kernel functions, and common kernel functions are listed as follows (Müller *et al.* 2001):

Table 4.1 Common Kernel Functions

Gaussian RBF	$\exp\left(-\frac{\|u-v\|^2}{c}\right)$
Polynomial	$\left(\langle u,v \rangle + \theta\right)^d$
Sigmoidal	$\tanh\left(k\left(\langle u,v \rangle\right)+\theta\right)$
Inv. Multiquadric	$\frac{1}{\sqrt{\|u-v\|^2 + c^2}}$

Any of the functions can be used as the kernel function to compute the inner products. But an interesting function is developed as follows:

$$k(u,v)=\begin{cases}\dfrac{1-\langle u,v\rangle^{d+1}}{1-\langle u,v\rangle}, & if\ \langle u,v\rangle \neq 1\\ d+1, & otherwise.\end{cases} \tag{4.15}$$

where d is the degree of the polynomial that will be used. The details of the inference of this kernel function can be found in the appendix. With larger value of d, more complicated function can be approximated, but usually more training samples are needed. The developed kernel function is of interest, because it implies to use the Taylor expansion as the mapping function. The motivation to develop this kernel function is explained using the following basic example. Suppose X is a scalar, and then model the relationship between X and Y, denoted as $f_Y(x)$. The underlying relationship is not available and it has been estimated from the available observations. The function $f_Y(x)$ could be quite complex, but usually it could be approximated by using a polynomial, according to the Taylor's theorem. The following Taylor expansion at the origin is used to approximate the unknown function $f_Y(x)$:

$$f_Y(x)=f(0)+\frac{f'(0)}{1!}x+\frac{f''(0)}{2!}x^2+\ldots+\frac{f^{(d)}(0)}{d!}x^d+R_d, \tag{4.16}$$

where d is the arbitrary degree of the polynomial, and R_d is the remainder term and would be quite small if x is close enough to 0. Ignore R_d and rewrite eq. (4.16) in vector notation:

$$f_Y(x)\approx\left\langle\left(1,x,x^2,\ldots,x^d\right)^T,\left(f(0),\frac{f'(0)}{1!},\frac{f''(0)}{2!},\ldots.\frac{f^{(d)}(0)}{d!}\right)^T\right\rangle. \tag{4.17}$$

The right-hand side of eq. (4.15) is the inner product of a vector $\left(1,x,x^2,\ldots,x^d\right)^T$ and a weight vector, which can be determined based on linear regression analysis. So let the

mapping function ϕ be the function which maps the original x into a vector $\left(1, x, x^2, ..., x^d\right)^T$. It is believed that the mapped vectors $\phi(x)$ should carry sufficient information for modeling the relationship between Y and original X. Then calculate the corresponding kernel function k as follows

$$k(u,v) = \left\langle (1,u,...,u^d),(1,v,...,v^d) \right\rangle = \begin{cases} \dfrac{1-(uv)^{d+1}}{1-uv}, & if\ uv \neq 1 \\ d+1, & otherwise. \end{cases}, \tag{4.18}$$

which is a special case of eq. (4.15) such that u and v are both scalars. In case that u and v are both vectors, the kernel functions can be found in the appendix.

After the kernel function k is selected, the kernel matrix K is then determined. The kernel function defined in eq (4.14) is applied to construct the kernel matrix K. Note that the degree d in the kernel function has not been determined yet. The selection of d will be discussed later in this section, and now assume it has been determined temporarily. Then, extract the principal components of the feature space, which are the eigen vectors of the kernel matrix K. Eigen decomposition is applied to the kernel matrix K, and

$$K = U\Sigma U^T, \tag{4.19}$$

where Σ is a diagonal matrix, of which the diagonal elements are non-zero eigen values of K and U is the matrix of corresponding eigen vectors.

The kernel matrix K is a $n \times n$ matrix, and suppose there are s non-zero eigen values. Clearly, n should be the upper bound of s. In fact, a tighter upper bound of s can be obtained, if the kernel function defined in eq. (4.15) is applied. Let N_d denote the upper bound of s in case that can eq. (4.15) is applied, and then N_d can be defined as:

$$N_d = \sum_{i=0}^{d} \binom{i+l-1}{l-1}, \tag{4.20}$$

where l is the length of the vector X. So there should be at most N_d non-zero eigen values of the kernel matrix K. If other kernel function is applied, the upper bound of the dimensionality of K should be different correspondingly.

After obtaining the principal components of the feature space, the feature factor of a given original sample $X = x$ can be computed as follows:

$$\Omega(x) = U^T \Phi^T \phi(x). \tag{4.21}$$

A new dataset can then be constructed, $(\Omega(X_1), Y_1)$,..., $(\Omega(X_n), Y_n)$, based on the original observations, and then the linear regression is applied to model the relationship between $\Omega(X)$ and the moments of Y, and the modeled relationship will be used for predicting conditional moments of Y.

The following linear equation is used to model the relationship between $\Omega(X)$ and Y^i:

$$W_i^T \Omega(X) + \varepsilon = Y^i, \tag{4.22}$$

where W_i is a weight vector and ε is a zero-mean Gaussian error. By applying minimum squared error (MSE) method, the following result of W_i is obtained

$$W_i = (A^T A)^{-1} A^T B_i, \tag{4.23}$$

where $A = [\Omega(X_1), \Omega(X_2), ..., \Omega(X_n)]^T$ and $B_i = [Y_1^i, Y_2^i, ..., Y_n^i]^T$. Now given any specific $X = x$, the i-th moment of Y can be predicted as follows:

$$\hat{Y}^i |_{X=x} = W^T \Omega(x). \tag{4.24}$$

Now let us solve the remaining problem discussed previously in this section about the selection of d in the kernel function. Clearly, with larger d, the corresponding $\phi(x)$

would contain more terms, and thus it is capable of approximating more complex functions. But the overfitting problem should be aware of here. The overfitting means that there are too many parameters in the model in presence of limited number of training samples, and it would make the performance worse on the unseen data. But if the value of d is too small, the corresponding $\phi(x)$ would not contain enough high-order terms to approximate complex functions. The d's values are usually different while predicting different moments of Y, and usually to predict higher order moment of Y a larger value of d should be chosen.

To predict the i-th moment of Y, the suitable d should be selected based on Akaike information criterion (AIC) (Akaike 1974) to balance the model's capability and simplicity. Follow the following steps to determine the value of d:

(1) Set d=0.

(2) Apply the kernel function defined in eq. (4.13) with the current d, and compute the feature vectors $\Omega(X_1),\ldots, \Omega(X_n)$.

(3) Regress Y^i on $\Omega(X)$ and determine the weight vector W_i.

(4) Evaluate the impact of each principle component on the output and select the optimal principle components based on AIC score.

(5) Increase d by 1 and repeat the steps from (2) to (4) until d exceeds the maximum value allowed.

(6) Pick up the value of d with the maximum AIC score as the determined value of d.

All the steps mentioned above except Step 4 has been described and can be implemented.

Now let us specify how to implement Step 4. Let $U=(\alpha_1,...,\alpha_s)$, $\Sigma=\mathrm{diag}(\lambda_1,...,\lambda_s)$ and $W_i=(w_1,...,w_s)^T$. The impact of the j-th principal component on the output, denoted as IMP_j, can be evaluated as follows:

$$IMP_j = w_j^2 \sum_{t=1}^{n} (\phi_j(X_t))^2, \tag{4.25}$$

where $\phi_j(X)$ denotes the j-th component of $\phi(X)$. IMP_j can be considered as the weighted energy of the j-th principal component. Thus, the impact of each principal component can be evaluated and sort all the principal components in the descending order of their impact. Suppose the order is r_1, ..., and r_s.

Then to select the optimal principal components, the AIC scores can be computed as follows:

$$\mathrm{AIC}_t = 2t + (n-1)\ln(RSS_t/(n-1)),\ t=1,\ldots,s, \tag{4.26}$$

where RSS_t is the residual sum of squares and evaluated by using leave-one-out cross validation. The RSS_t is defined as follows:

$$RSS_t = \sum_{j=1}^{n} (Y_j^i - \hat{Y}_{-j}^i|_{X=X_j})^2, \tag{4.27}$$

where $\hat{Y}_{-j}^i|_{X=X_j}$ is the predicted $\hat{Y}^i|_{X=X_j}$ using all the observations except (X_j,Y_j) as the training data set. Then the maximum is picked up from $\mathrm{AIC}_0,\ldots,\mathrm{AIC}_s$. Suppose AIC_t is the maximum, and then the optimal principal components are $\alpha_{r_1},\ldots,\alpha_{r_t}$. Thus, the maximum AIC score and corresponding principal component set are obtained.

The degree of the kernel function can be searched within a certain range, e.g. from 1 to 10, etc. After all the candidate kernel functions have been investigated, the

value of d with the maximum AIC score is selected, and the corresponding principal component set is selected as the optimal principal component set. So a new eigen vector matrix $U' = (\alpha_{r_1}, ..., \alpha_{r_t})$ is defined based on the optimal principal component set. Then, the U in eq. (19) could be replaced with U' to compute the feature vectors and predict the moment of Y. Please note that the value of d and the corresponding optimal principal component set are usually different while predicting different moments of Y. So the optimal d and U' should be evaluated for each moment of Y respectively.

It is better to predict central moments instead of ordinary moments of Y. Though theoretically they are equivalent, predicting central moments usually yield better performance in practice. It can be done by predicting the first moment of Y firstly, and then predicting the moments of $(Y - \hat{Y}^1)$. The techniques to reconstruct the density function based on central moments are identical to those based on ordinary moments, except that the conditional density function reconstructed based on central moments at X are shifted toward the origin and the offset is $-\hat{Y}^1|_{X=x}$, which can be easily adjusted.

4.5.3 Moment-based Density Function Reconstruction

In this step, the probability density function will be reconstructed based on the moments obtained in the last step, which is known as a moment problem (Shohat & Tamarkin 1943; Akhiezer 1965). There are a couple of methods proposed in the literature, which makes use of Pearson Curves (Solomon & Stephens 1978), saddelpoint approximations (Reid 1988) or inverse Mellin transform (Mathai & Saxena 1978). Alternatively, the density function could be represented in terms of orthogonal polynomials, such as Laguerre, Legendre, Jacobi, and Hermite polynomials (Provost & Rudiuk 1995; Provost

2005). Typically, the exact moments are required in above methods, and the reconstructed function is very sensitive to the estimation error, meaning that the reconstructed functions could be easily ruined by the inaccurately estimated moments. Since only the approximated moments estimated based on regression are available and the exact values are never available, the methods mentioned above are not usable in this application.

A unified semi-parametric method is proposed by Provost (2005) and only approximated moments are required, which is feasible in this application. This method is reviewed briefly and then explain its application in this application in the remaining of this section.

The basic idea of this method is that if the first t moments of a density function are known, the density function can be approximated by means of the product of a base density function, whose parameters are determined by matching available moments, and a polynomial of degree t, whose efficients are obtained by making use of the moment method as well. Generally speaking, with more moments the density functions with more details could be recovered. But in practice due to the difficulty of estimating higher order moments, the estimation of higher order moments is not as reliable as that of lower ones, so adopting more moments does not always yield better performance. The problem that how many moments should be used for density function reconstruction will be discussed later.

Suppose the first $2m$ moments of Y have been obtained given a specific X in the last step. First, construct a $(m+1)\times(m+1)$ matrix M, and $M_{i,j} = \hat{Y}^{i+j}|_{X=x}$, of which $M_{i,j}$ is the element in i-th row and j-th column. Clearly, $M_{0,0} = 1$ in all the cases.

Second, a base function should be chosen as the initial approximation of the target density function. Usually, one of Gaussian, Gamma and Beta functions can be selected as the base function. Different base functions possess the different tail behavior, so we could pick up a suitable base function based on the prediction of the tail behavior of the underlying density function. Generally speaking, if there is not any available knowledge about the underlying density function, Gaussian function is a good starting point. After the type of base function is selected, its parameters can be determined by matching the first few available moments.

Third, the first m moments of the base function are computed on the support [-1, 1], denoted as $\mu_0, \ldots, \mu_m$. Then the underlying density function can be approximated as:

$$\hat{f}_{Y|X}(y \mid X = x) = \psi(y) \sum_{i=0}^{m} \xi_i y^i , \tag{4.28}$$

where $\psi(y)$ is the base function and $(\xi_0,\ldots,\xi_m)^T = M^{-1}(\mu_0,\ldots,\mu_m)^T$. For more details of this method, please refer to Provost (2005).

Thus, the conditional density function of Y has been reconstructed based on the estimated moments. If we perform the preprocessing step, we could recover the density function of original data by performing simple scaling and shifting easily.

4.5.4 Performance Measurement and Optimization

A cross-validation method was proposed to evaluate the performance of conditional density estimators (Fan & Yim 2004). The cross-validation method is applied to evaluate the performance and solve the remaining problem about determining the number of moments to be predicted left previously.

For an estimator $\hat{f}_{Y|X}(y|x)$ of $f_{Y|X}(y|x)$, an integrated squared error (*ISE*) is defined as follows:

$$I = \iint\left(\hat{f}_{Y|X}(y|x) - f_{Y|X}(y|x)\right)^2 f(x)dydx$$

$$= \iint\hat{f}_{Y|X}(y|x)^2 f(x)dydx - 2\iint\hat{f}_{Y|X}(y|x)f_{Y|X}(y|x)f(x)dydx + \iint f_{Y|X}(y|x)^2 f(x)dydx$$

$$= I_1 - 2I_2 + I_3. \tag{4.29}$$

I is not computable, because $f_{Y|X}(y|x)$ is unknown. Note that I_3 does not depend on the estimator $\hat{f}_{Y|X}(y|x)$, and thus can be ignored. Let $I' = I_1 - 2I_2$. So the minimization of I is equivalent to the minimization of I'. Fortunately, based on the available observations I_1 and I_2 can be estimated as:

$$\hat{I}_1 = \frac{1}{n}\sum_{i=1}^{n}\int\hat{f}_{-i}(y|X_i)^2 dy, \tag{4.30}$$

$$\hat{I}_2 = \frac{1}{n}\sum_{i=1}^{n}\hat{f}_{-i}(Y_i|X_i). \tag{4.31}$$

both of which are available, and then $\hat{I}' = \hat{I}_1 - 2\hat{I}_2$. In eq. (4.30) and (4.31) the leave-one-out cross-validation method is applied, and $\hat{f}_{-i}(y|X_i)$ is the estimator $\hat{f}_{Y|X}(y|X)$ with the observation (X_i, Y_i) omitted. Thus, I' is estimated instead of original I. If the $\hat{I}'$ is smaller, the performance is better. Therefore, the value of $\hat{I}'$ is used as the *ISE* to evaluate the performance of a conditional density estimator.

Now, let us solve the remaining problem, i.e. how to determine the number of moments to be used for density function reconstruction. In fact, the performance of the estimator should be evaluated with different number of the moments, and then the one

with the best performance should be picked up; the number of the moments used in this best estimator is treated as the optimal value of number of the moments to be used.

4.6 Experimental Results

First, the proposed method discussed in Section 4.5 is compared with the double-kernel density estimator discussed in Section 4.4 through a simulated model for demonstration. Then, the proposed method is applied to the real solar flare dataset and the density functions of flare peak intensity and flare index are predicted, respectively. In each experiment, the *ISE*s of the proposed method and double-kernel density estimator are measured for comparison.

4.6.1 A Simulated Model

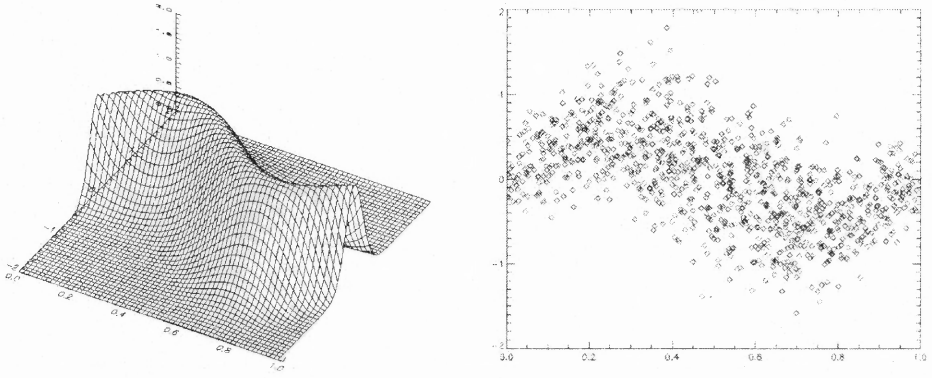
To compare the performance of the proposed method and double-kernel density estimator, the following simple simulated model is considered:

$$f_{Y|X}(y \mid x) \sim N\left(\frac{\sin(2\pi x)}{2}, 1.2x(1-x)+0.2\right), \tag{4.32}$$

where x is uniformly distributed over [0, 1]. We generate 1000 random samples for training and the same number of samples for testing based on the $f_{Y|X}(y \mid x)$. The Figure 4.1 (a) shows the conditional density function $f_{Y|X}(y \mid x)$, and (b) shows the generated training samples.

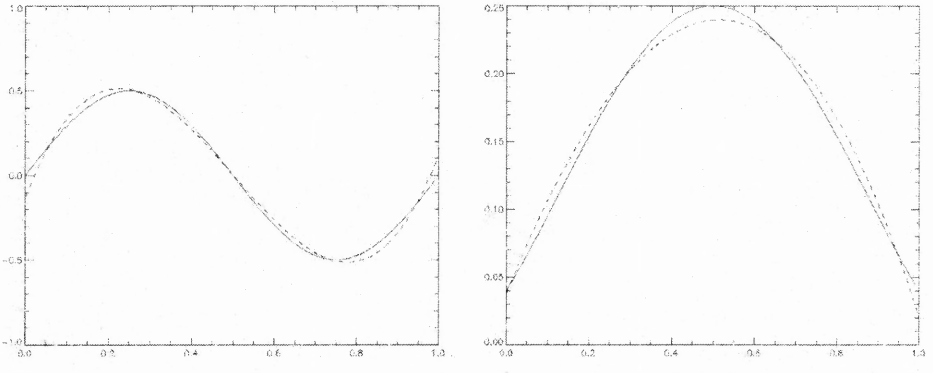
The proposed method is applied to estimate the conditional moments of Y and the first two moments are plotted in Figure 4.2 (a) and (b), respectively. From the figure, it is be observed that the conditional moments estimated by the proposed method are very

close to the true moments. Then the proposed method and double-kernel method are applied to estimate the conditional density functions, and the conditional density functions estimated at $x = 0.2$, $x = 0.4$, $x = 0.6$ and $x = 0.8$ are plotted in Figure 4.3 (a), (b), (c) and (d), respectively. From Figure 4.3, the conditional density functions generated by the proposed method are smoother and closer to the true densities. The *ISE* scores is computed to compare the performance of two methods, and *ISE*s of the proposed method and double-kernel method are -2.887 and -2.828, respectively. Therefore, in this experiment, the proposed method outperforms the double-kernel method.



(a) (b)

Figure 4.1 (a) 3-D view of the conditional density function $f_{Y|X}(y \mid x)$ and (b) 1000 samples randomly generated from $f_{Y|X}(y \mid x)$.



(a) (b)

Figure 4.2 (a) The true and estimated first moment of Y, and (b) the true and estimated second central moment of Y, where the true functions are plotted using solid curves and the estimated functions are plotted using dotted curves.

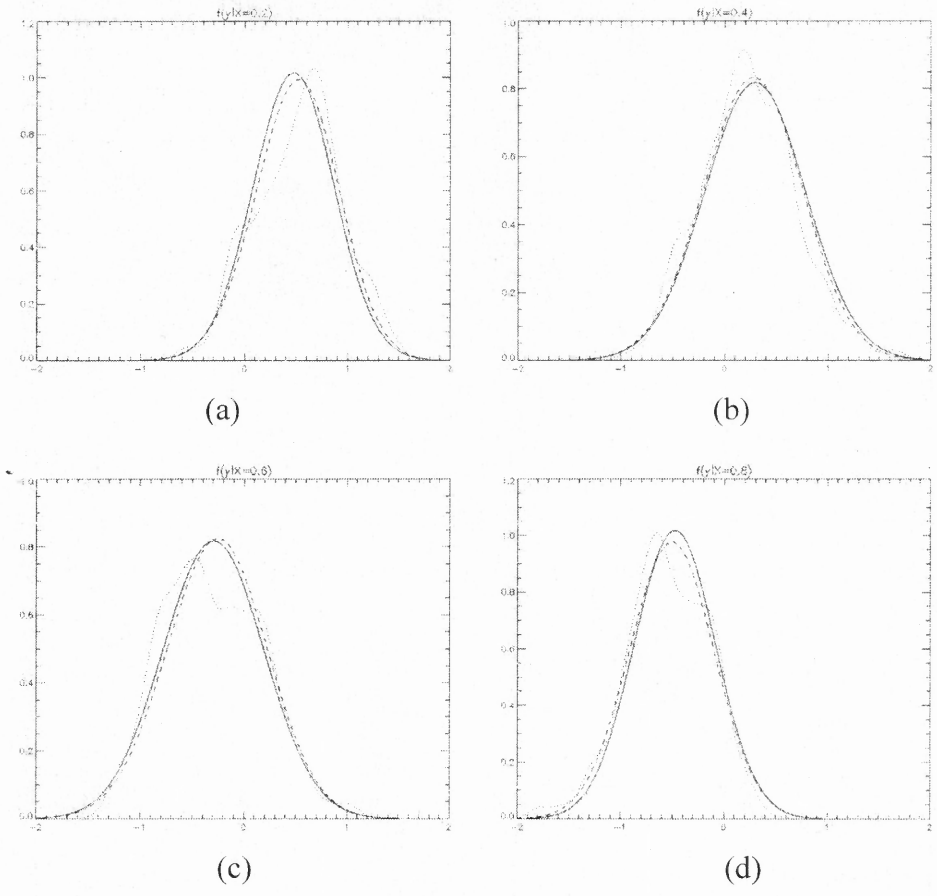


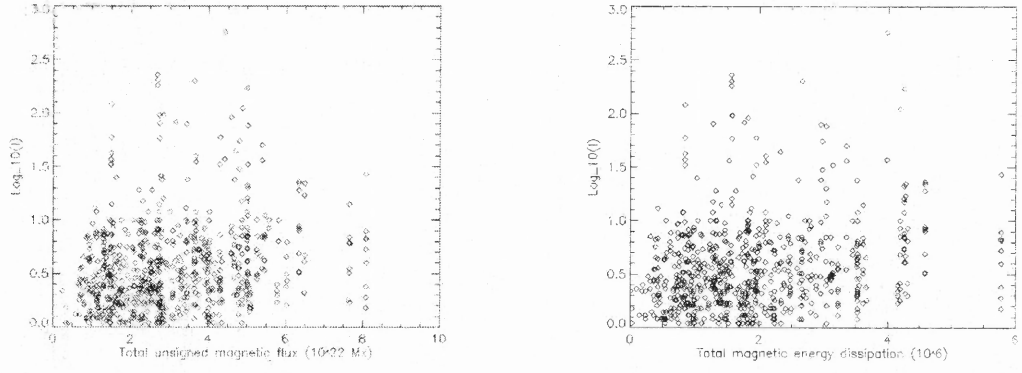
Figure 4.3 Estimated conditional densities for model (4.31) for (a) $x = 0.2$, (b) $x = 0.4$, (3) $x = 0.6$ and (4) $x = 0.8$, using the proposed method (dashed curve) and double-kernel method (dotted curve), compared with the true densities (solid curve).

4.6.2 Application to Flare Peak Intensity Prediction

The proposed method is applied to predict the conditional density function of the peak intensity of solar fares that will be generated within next 24 hours based on the magnetic parameters of the active region. After checking the flare events associated with the active regions (NOAA 07961 to 10932) from 1996 to 2006, 711 significant solar flares (C-, M-, and X-class) are selected for investigation and they are associated with 299 active

regions, and all the active regions are near the disk center, of which the position angles are less than 30 degrees, to facilitate magnetic parameter measurement.

The magnetic parameters of the associated active region, including total unsigned magnetic flux T_{flux}, energy dissipation E_{diss}, were measured from MDI line-of-sight magnetograms. The response variable is the flare peak intensity I. Instead of using the original I, let $Y = \log_{10}(I)$ and predict the conditional density function of Y. Thus, for a C-calss flare, the Y should be [0, 1); for a M-class flare, the Y should be between [1, 2), and for a X-class flare, the Y should be [2, $+\infty$). Let X denote the covariate, and the task then is to predict the conditional density function $f_{Y|X}(y \mid x)$ based on a specific magnetic covariate X. The relationship between T_{flux} and $\log_{10}(I)$ is shown in Figure 4.4 (a) and the relationship between E_{diss} and $\log_{10}(I)$ is shown in Figure 4.4(b).



(a) (b)

Figure 4.4 (a) The relationship between T_{flux} and $\log_{10}(I)$ and (b) the relationship between E_{diss} and $\log_{10}(I)$.

Both the proposed method and double-kernel method are applied to predict the conditional density function $f_{Y|X}(y \mid x)$ based on $X = T_{flux}$. To measure the performance accurately, the experiment contains 50 iterations. In each iteration, 356 samples are

picked up randomly for training and the remaining samples are used for testing, and then measure the performance of two methods. After 50 iterations, the average *ISE* of each method is obtained. The average *ISE*s of the proposed method and the double-kernel method are -2.43 and -2.47, respectively, meaning that the proposed method performed slightly worse than the double-kernel method. In Figure 4.5, the conditional density functions estimated by two methods at different values of T_{flux} are plotted, and it can be observed that two curves are closer to each other with larger T_{flux}.

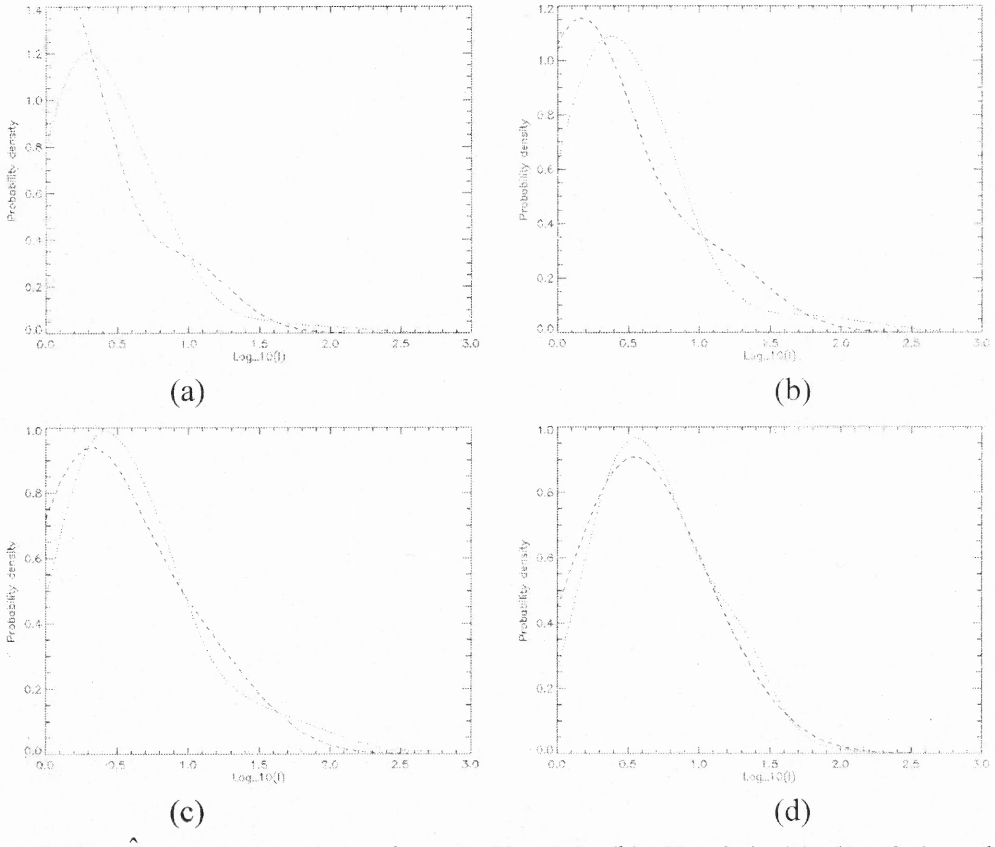


Figure 4.5 The $\hat{f}_{I|T_{flux}}(y \mid x)$ estimated at (a) $X = 0.2$, (b) $X = 0.4$, (c) $X = 0.6$, and (d) $X = 0.8$, using the proposed method (solid curve) and double-kernel method (dashed curve).

Then two methods are applied to predict the conditional density function $f_{Y|X}(y|x)$ based on $X = E_{diss}$ and did the similar experiment. The average *ISE*s of the proposed method and the double-kernel method are -2.41 and -2.47, respectively. This result also shows the proposed method performed slightly worse than the double-kernel method and confirmed the result of previous experiment. In Figure 4.6, the conditional density functions estimated by two methods at different values of E_{flux} are plotted, and it can be observed that two curves are closer to each other with larger E_{flux}.

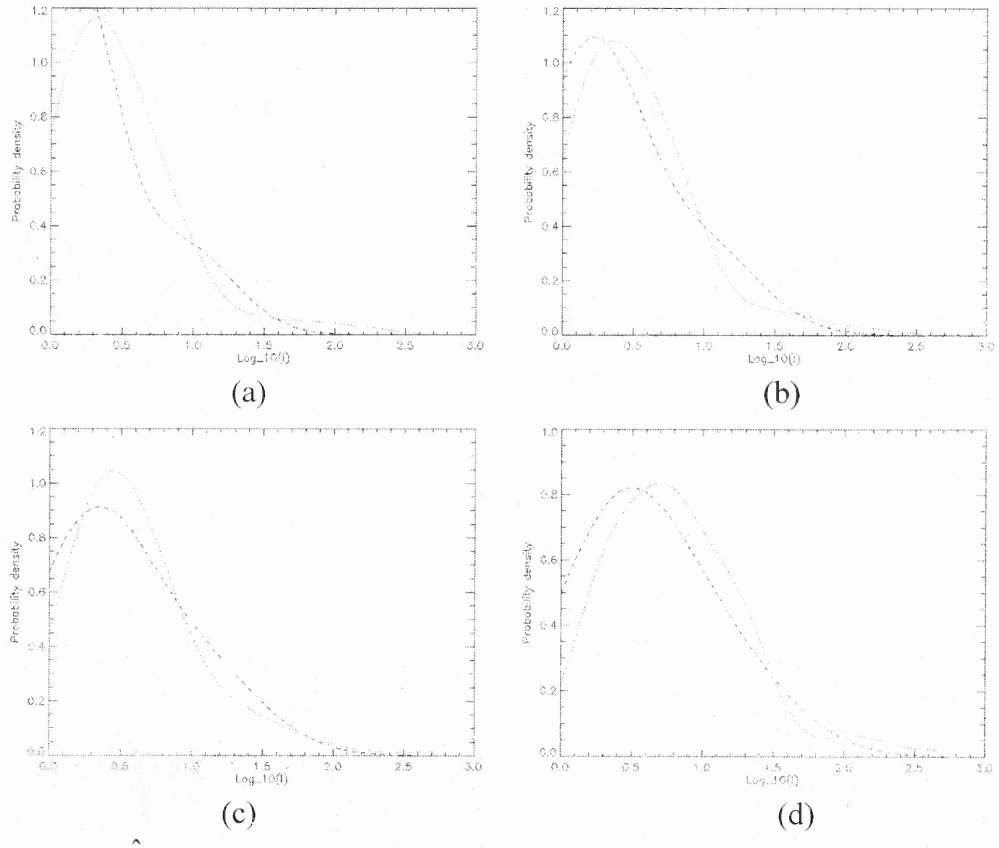


Figure 4.6 The $\hat{f}_{Y|E_{diss}}(y|x)$ estimated at (a) $X = 0.2$, (b) $X = 0.4$, (c) $X = 0.6$, and (d) $X = 0.8$, using the proposed method (solid curve) and double-kernel method (dashed curve).

We also predict the density function of flare peak intensity based on both T_{flux} and E_{diss}, and the *ISE* is -2.43, which is identical to the result using T_{flux}. The reason could be that T_{flux} and E_{diss} are correlated to each other closely and the combination of them does not provide much more information compared with single one of them.

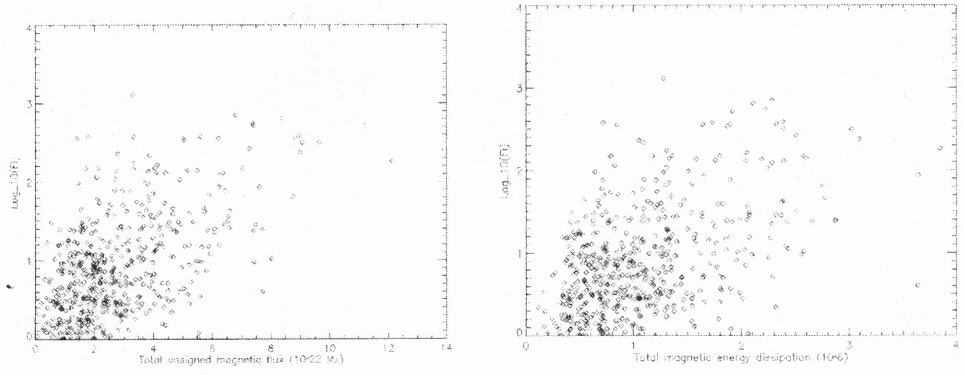
4.6.3 Application to Flare Index Prediction

The two methods are applied to predict the condition density function of flare index based on magnetic parameters. The magnetic parameters used here include total unsigned magnetic flux T_{flux}, energy dissipation E_{diss}. The flare index is calculated using eq. (4.1), and the length of the time window is 3 days. The active regions (NOAA 07961 to 10932) are checked and 530 active regions with at least one C- or higher class flare are picked up for investigation, and all the active regions are selected near the disk center, of which the position angles are less than 30 degrees, to facilitate magnetic parameter measurement.

Similarly, the magnetic parameters used here include total unsigned magnetic flux T_{flux} and energy dissipation E_{diss}. Let $Y = \log_{10}(FI)$ and X denote the covariate. The task then is to predict the conditional density function $f_{Y|X}(y \mid x)$ based on a specific magnetic covariate X. The relationship between T_{flux} and $\log_{10}(FI)$ is shown in Figure 4.7 (a) and the relationship between E_{diss} and $\log_{10}(I)$ is shown in Figure 4.7(b). From Figure 4.7, It confirms that there does exist correlation between flare index and magnetic parameters.

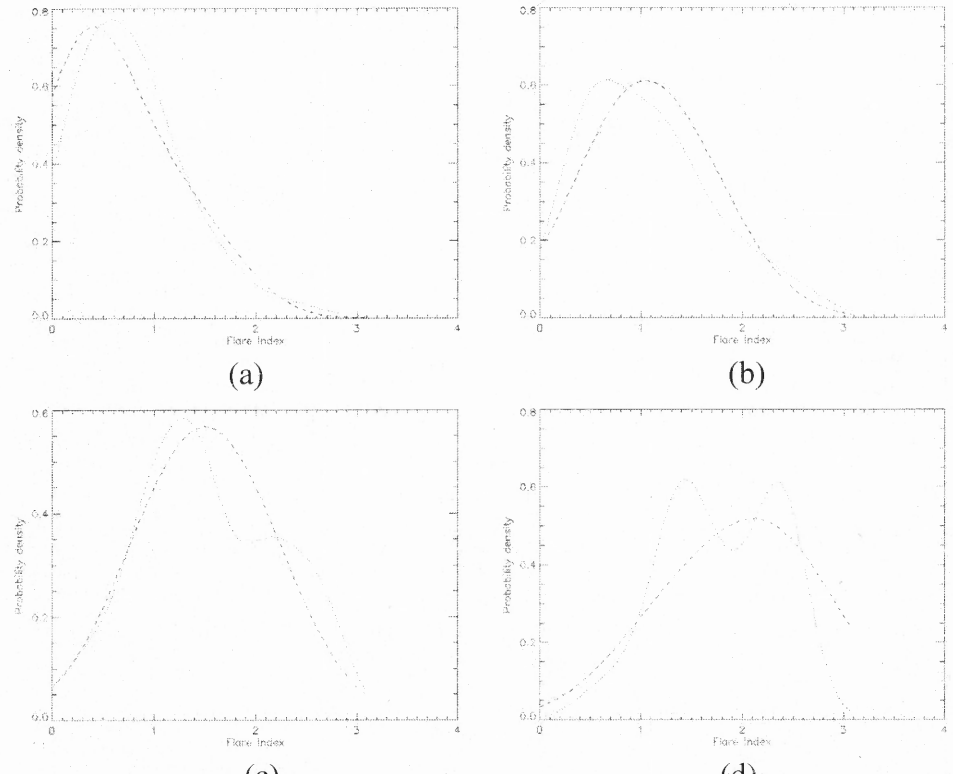
The proposed method and double-kernel method are applied to predict the conditional density function $f_{Y|X}(y \mid x)$ based on $X = T_{flux}$. The same experiments are done and the performance of two methods are measured, repsectively. The experimental

results shows that the average *ISE*s of the proposed method and the double-kernel method are -1.78 and -1.85, respectively, which means the proposed method performed slightly worse than the double-kernel method.



(a) (b)

Figure 4.7 (a) The relationship between T_{flux} and $\log_{10}(FI)$ and (b) the relationship between E_{diss} and $\log_{10}(FI)$.



(a) (b)

(c) (d)

Figure 4.8 The $\hat{f}_{FI|T_{flux}}(y \mid x)$ estimated at (a) $X = 0.2$, (b) $X = 0.4$, (c) $X = 0.6$ and (d) $X = 0.8$, using the proposed method (solid curve) and double-kernel method (dashed curve).

Two methods are also applied to predict the conditional density function $f_{Y|X}(y \mid x)$ based on $X = E_{diss}$ and did the similar experiment. The average *ISE*s of the proposed method and the double-kernel method are -1.85 and -1.86, respectively. This result also shows the performance of the proposed method is comparable to that of the double-kernel method and confirmed the result of previous experiment. In Figure 4.9, the conditional density functions estimated by two methods at different values of E_{diss} are plotted, and it can be observed that two curves are closer to each other with larger E_{diss}.

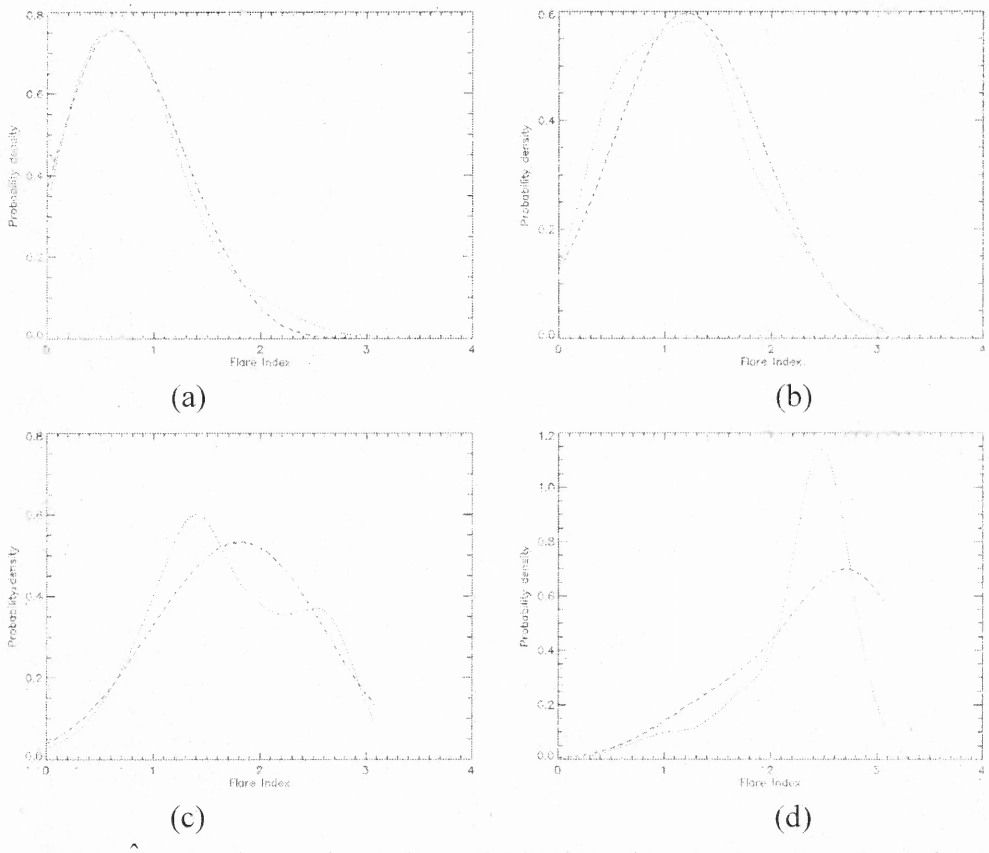


Figure 4.9 The $\hat{f}_{FI|E_{diss}}(y \mid x)$ estimated at (a) $X = 0.2$, (b) $X = 0.4$, (c) $X = 0.6$ and (d) $X = 0.8$, using the proposed method (solid curve) and double-kernel method (dashed curve).

We also predict the density function of flare index based on both T_{flux} and E_{diss}, and the *ISE* is -1.83, which does not show superiority to single magnetic parameter. The

reason could be that T_{flux} and E_{diss} are correlated to each other closely and the combination of them does not provide much more information compared with single one of them.

4.7 Discussion and Conclusions

In this chapter, a new method is presented to predict the conditional density function by combining kernel-based nonlinear regression and moment-based density function reconstruction techniques. This method is applied onto the solar flare perdition problem to predict the conditional density function of flare peak intensity or solar flare based on the magnetic parameters.

The proposed method is compared with the double-kernel density estimator through one simulation model and the solar flare prediction problem. In the simulation model, the proposed method outperformed the double-kernel density estimator, while in solar flare prediction problem, the performance of the proposed method is slightly worse than that of the double-kernel density estimator. So the performance of the proposed method is comparable with that of the double-kernel density estimator in regular case, and since it can be easily applied to high dimensional data set, which makes it attractive in solar flare prediction problem, because in the future more and more parameters would be included, and in this case the double-kernel density estimator becomes ineffective.

In summary, a mathematical tool, which can be used for conditional density estimation, is developed and it is capable of handling high dimensional data sets effectively.

CHAPTER 5

SUMMARY AND CONCLUSIONS

In this dissertation, the automatic methods for prominence eruptions detection, emerging flux regions detection and solar flare prediction are presented. These automated methods can be used for processing large scale solar images as well as real-time solar activity observation and prediction, which is important for space weather. Advanced image processing and machine learning techniques are combined in these methods to take advantage of the progress in information technologies during the last decades.

In prominence eruption detection, image segmentation, morphological operations and support vector machine are applied. The classification rates of prominence eruptions and non-prominence eruptions are 93.3% and 93.6%, respectively. However, since the number of non-prominence eruptions detected by the method is much higher than that of prominence eruptions, most limb events classified by the method are not prominence eruptions, which should be improved in the future. The new properties and segmentation technique will be considered to improve the method. Also, it would be interesting to combine the detection of filament and prominence eruptions to make the method more robust and sensitive to the true prominence eruptions.

In emerging flux region detection, multi-scale harmonic filter is applied to detect dipoles of various sizes on solar surface by using MDI magnetograms, and the rotation invariance of the filter enables to detect the dipoles of any orientation. Kalman filter is applied to trace the candidate EFR until its disappearance. The lost-and-return phenomena was noticed in the experiments, which indicates it is necessary to improve the tracing techniques, and the alternative techniques, such as IPDA, JPDA and JIPDA,

should considered and tested in the future. The classification rates of EFRs and non-EFRs are 96.4% and 98.0%, respectively, and the false alarm rate is about 25.7% due to the large number of non-EFRs detected by the method. The joint composite filter should be considered to improve the performance of identifying new candidate EFRs, and alternative statistical features should be included to improve the classification rates. This method should be extended to detect other bipolar structure on the solar surface, such as active regions in the future.

In solar flare prediction, a novel method is developed by using kernel-based nonlinear regression and moment-based density function reconstruction techniques. This method is used to predict the probability density function of either flare peak intensity or flare index based on magnetic parameters of the associated active region. It mainly involves two steps. In the first step, the conditional moments of the response variable is predicted by using kernel-based nonlinear regression, which enables it to handle high dimensional data set efficiently. In the second step, the density function is reconstructed based on the predicted moments. This method is applied to solar fare data set collecting the information of active regions from 1996 to 2006, and compared with double-kernel density estimator. This experimental results show that the proposed method yields comparable performance with the double-kernel density estimator in the regular case. Since it can be applied to high dimensional dataset, it is essential for future research in solar flare prediction, because more and more magnetic parameters will be included. But, currently, it is limited to approximate relative simple function, such as uni model functions, due to the difficulty in estimating high order conditional moments. If the high order conditional moments could be estimated more accurately, the method would be

able to provide more subtle details of the target density function. In summary, A usable mathematical tool for conditional density estimation has been developed in presence of high dimensionality of data set, and it can be used in solar flare prediction based on multiple magnetic parameters.

APPENDIX

A SPECIAL KERNEL FUNCTION CORRESPONDING TO TAYLOR EXPANSION

According to Talyor's theorem, it is possible to use a polynomial to approximate a differentiable function. Suppose there is a complicated differentiable function $y = f(\vec{x})$, where $\vec{x} = (x_1, x_2, ..., x_d)^T$. Its Taylor's expansion at the origin using the terms of the degree up to n can be defined as follows:

$$f(\vec{x}) = \sum_{i_1+...+i_d=0}^{n} \frac{\frac{\partial^{i_1}}{\partial x_1} ... \frac{\partial^{i_d}}{\partial x_d} f(\vec{0})}{i_1 ... i_d} \left(x_1^{i_1} ... x_d^{i_d}\right) + Rm, \quad (1)$$

where Rm is the remainder. The remainder Rm depends on $\vec{x}$ and is small if $\vec{x}$ is close enough to $\vec{0}$, so it is ignored in the following discussion. The eq. (1) with Rm omitted can be re-written in vector notation as follows:

$$f(\vec{x}) = \beta^T \phi(\vec{x}), \quad (2)$$

where β is a weight vector and $\phi(\vec{x}) = \left(1, ..., (x_1^{i_1} ... x_d^{i_d}), ... x_d^n\right)^T$, which is a expansion vector and contains all the terms of the degree up to n. If the exact form of $f(\vec{x})$ is available, the parameter β can be computed from the partial derivatives of $f(\vec{x})$; otherwise it could be estimated from the available observations $(\vec{x}_1, y_1)$, ..., $(\vec{x}_n, y_n)$ based on regression analysis. So to approximate a complicated function $f(\vec{x})$, the original $\vec{x}$ is first mapped to a extended vector $\phi(\vec{x})$ and then apply regression analysis to determine the weight of each component of $\phi(\vec{x})$. But the kernel trick allows to implement the mapping function

$\phi(\vec{x})$ and regression analysis implicitly by using the kernel function, which computes the inner product of the mapped vectors.

Let $C_p(\vec{x})$ denote the vector whose entries are all possible p-th degree ordered products of the entries of $\vec{x}$, and then $C_p(\vec{x}) = \left(x_1^p, x_1^{p-1}x_2, ..., x_d^p\right)^T$. Since $\phi(\vec{x}) = \left(1, ..., (x_1^{i_1}...x_d^{i_d}), ...x_d^n\right)^T$, the corresponding kernel function can be defined as:

$$k(\vec{u},\vec{v}) = \left\langle \phi(\vec{u}), \phi(\vec{v}) \right\rangle = \sum_{k=0}^{d} \left\langle C_k(\vec{u}), C_k(\vec{v}) \right\rangle = \sum_{k=0}^{d} (\vec{u}^T\vec{v})^k = \frac{1-(\vec{u}^T\vec{v})^{k+1}}{1-\vec{u}^T\vec{v}}, \text{ for } \left\langle \vec{u},\vec{v} \right\rangle \neq 1 \quad (4)$$

If $\left\langle \vec{u},\vec{v} \right\rangle = 1$, $K_{i,j} = d+1$.

In the simple case that that the input vectors are one-dimensional, indicating scalars, the Taylor series of $f(x)$ can be represented as:

$$f(x) = \sum_{j=0}^{d} c_j x^j . \quad (5)$$

With the mapping function $\phi(x) = [1, x^2, ..., x^d]^T$, then, the kernel function is defined as:

$$k(u,v) = \phi(u)^T \phi(v) = \sum_{k=0}^{d} (x_i x_j)^k = \frac{1-(uv)^{d+1}}{1-uv}, \text{ for } uv \neq 1. \quad (6)$$

If $uv \neq 1$, $k(u,v) = d+1$.

In interesting observation is that if the magnitudes of $\vec{x}$ is limited to be less than 1, such that $\|\vec{x}\| < 1$, and let $d \to \infty$, the eq. (4) would be converted to:

$$k(\vec{u},\vec{v}) = \frac{1}{1-\vec{u}^T\vec{v}} . \quad (7)$$

REFERENCES

Abramenko, V. I., "Relationship between magnetic power spectrum and flare productivity in solar active regions," *Astrophysical Journal* vol. 629, no. 2, pp. 1141-1149, 2005.

Akaike, H. "A new look at the statistical model identification," *IEEE Transactions on Automatic Control* vol. 19, no. 6, pp. 716-723, 1974.

Akhiezer, N. I., "The classical moment problem and some related questions in analysis," translated by N. Kemmer, Hafner Publishing Co., New York, 1965.

Ball, C. A. and Torous, W. N., "Unit roots and the estimation of interest rate dynamics," *Journal of Empirical Finance*, vol. 3, pp. 215-238, 1996.

Bar-Shalom, Y., "Tracking methods in a multitarget environment," *IEEE Trans. Automatic Control*, vol. 23, no. 4, pp. 618- 626, 1978.

Bar-Shalom, Y. and Fortmann, T. E., *Tracking and Data Association*, San Diego, CA: Academic Press Professional, Inc., 1988.

Bashtannyk, D. M. and Hyndman, R. J., "Bandwidth selection for kernel conditional density estimation," *Computational Statistics and Data Analysis*, vol. 36, pp. 279-298, 2001.

Baudat, G. and Anouar, F., "Kernel-based methods and function approximation," *Neural Networks, 2001. Proceedings. IJCNN '01. International Joint Conference on*, vol. 2, pp. 1244-1249, Washington, DC, July 2001.

Bruzek, A., *Illustrated Glossary for Solar and Solar-Terrestrial Physics*, Dordrecht, Holland: Kulwer Academic Publishers, 1977.

Copas, J. B., "Local Likelihood Based on Kernel Censoring," *Journal of the Royal Statistical Society*, Ser. R, vol. 57, pp. 221-235, 1995.

Denker, C., Johannesson, A., Marquette, W., Goode, P. R., Wang H. and Zirin, H., "Synoptic Hα full-disk observations of the sun from Big Bear Solar Observatory – I. instrumentation, image Processing, data products, and first results," *Solar Phys.*, vol. 184, no. 1, pp. 87-102, 1999.

Fan, J. and Yao, Q., *Nonlinear Time Series: Non-parametric and Parametric Methods*, New York: Springer-Verlag, 2003.

Fan, J., Yao, Q. and Tong, H., "Estimation of conditional densities and sensitivity measures in nonlinear dynamical systems," *Biometrika*, vol. 83, pp. 189-206, 1996.

Fan, J. and Yim, T. H., "A cross-validation method for estimating conditional probability densities," *Biometrika*, vol. 91, no. 4, pp. 819-834, 2004.

Freeland, S. L. and Handy, B. N., "Data analysis with the solarsoft system," *Solar Phys.*, vol. 182, no. 2, pp. 497-500, 1998.

Garcia-Martinez, P., Garcia, J. and Ferreiara, C., ''A new criterion for determining the expansion center for circular-harmonic filters,'' *Opt. Commun.*, vol. 117, pp. 399-405, 1995.

Gao, J., Wang, H. and Zhou, M., "Development of an automatic filament disappearance detection system," *Solar Phys.*, vol. 205, pp. 93-103, Jan. 2002.

Gilbert, H. R., Holzer, T. E., Burkepile, J. T. and Hundhausen, A. J., "Active and eruptive prominences and their relationship to coronal mass ejections," *Astrophysical Journal*, vol. 537, no. 1, pp. 503-515, 2000.

Gopalswamy, N., Shimojo, M., Lu, W., Yashiro, S., Shibasaki, K. and Howard, R. A., "Prominence eruptions and CMEs: a statistical study using microwave observations," *Astrophysical Journal*, vol. 586, no. 1, pp. 562-578, 2003.

Greenberg, M., *Advanced Engineering Mathematics*, 2nd ed., Upper Saddle River, NJ: Prentice Hall, 1998.

Grewal, M. S. and Andrews, A. P., *Kalman Filtering Theory and Practice*, Upper Saddle River, NJ: Prentice Hall, 1993.

Guyon, I. and Stork, D. G., *Linear Discriminant and Support Vector Machine*, Cambridge, MA: MIT Press, 2000.

Hagenaar, H. J., "Ephemeral regions on a sequence of full-disk Michelson Doppler Imager magnetograms," *Astrophysical Journal*, vol. 555, pp. 448-461, 2001.

Hall, P., Racine, J. and Li, Q. "Cross-validation and the estimation of conditional probability densities," *Journal of the American Statistical Association*, 2004, vol. 99, no. 468, pp. 1015-1026, 2004.

Hall, P., Wolff, R. and Yao, Q., "Methods for estimating a conditional distribution function," *Journal of the American Statistical Association,* vol. 94, pp. 154-63, 1999.

Härdle, W., *Applied Non-parametric Regression*, Cambridge, MA: Cambridge University Press, 1990.

Harvey, K. L., *Magnetic Bipoles on the Sun*, Ph.D. thesis, Rijksuniv.Utrecht, 1993.

Harvey, K. L. and Martin, S. F., "Ephemeral active regions," *Solar, Phys.*, vol. 32, pp. 389-402, 1973.

Hoffbeck, J. P. and Landgrebe, D. A., "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 763-767, 1996.

Hsu, Y. N. and Arsenault, H. H., "Optical pattern recognition using circular-harmonic expansion," *Appl. Opt.* vol. 21, pp. 4016-4019, 1982.

Hyndman, R. J., Bashtannyk, and Grunwald, G. K., "Estimating and visualizing conditional densities," *Journal of Computational and Graphical Statistics*, vol. 5, pp. 315-336, 1996.

Jacome, M. A., Gijbels, I. and Cao, R., "Comparison of presmoothing methods in kernel density estimation under censoring," *Computation Statistics*.

Jing, J., Song, H., Abramenko, V., Tan, C. and Wang, H., "The statistical relationship between the photospheric magnetic parameters and the flare productivity of active regions," *Astrophysical Journal*, vol. 644, no. 2, pp. 1796-1796, 2006.

Jing, J., Yurchyshyn, V. B., Yang, G., Xu, Y., and Wang, H. "On the relation between filament eruptions, flares, and coronal mass ejections," *Astrophys. J.*, vol. 614, no. 2, pp.1054-1062, 2004.

Kalman, R. E., "A new approach to linear filtering and prediction problems," *Transactions of the ASME - Journal of Basic Engineering,* vol. 82, pp. 35-45, 1960.

Kubo, M., Shimizu, T., and Lites, B. W., "The evolution of vector magnetic fields in an emerging flux region," *Astrophysical Journal*, vol. 595, pp. 465-482, 2003.

Loader, C. R., "Bandwidth selection: classical or plug-in?" *The Annals of Statistics*, vol. 27, pp. 415-438, 1999.

Lugt, A. V., "Signal detection by complex spatial filtering," *IEEE Trans. Inf. Theory.*, vol. 10, no. 2, pp. 139-145, 1964.

Mathai, A. M. and Saxena, R. K., *The H-function with applications in statistics and other disciplines*, New York: John Wiley & Sons, 1978.

Maybeck, P. S., *Stochastic Models, Estimation, and Control*, San Diego, CA: Academic Press, Inc. vol. 1, 1979.

Mays, L. E. and Gamlin, P. D., "Neuronal circuitry controlling the near response," *Curr. Opin. Neurobiol,* vol. 5, pp. 763–768, 1995.

Mika, S., Rätsch, G., Weston, J., Schölkopf, B. and Müller, K.-R., "Fisher discriminant analysis with kernels," *Neural Networks for Signal Processing* IX, Y.-H. Hu, J. Larsen, E. Wilson, and S. Douglas, Eds. Piscataway, NJ: IEEE, pp. 41-48, 1999.

Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K. and Schölkopfan, B. "Introduction to Kernel-Based Learning Algorithms," *IEEE Trans. On Neural Networks*, vol. 12, no. 2, pp. 181-201, 2001.

Murray, N., "On the inclination of photospheric solar magnetic fields," *Astrophysical Journal*, vol. 401, pp. 386-397, 1992.

Musicki, D. and Evans, R., "Joint integrated probabilistic data association – JIPDA," *In Proceedings of the Fifth International Conference on Information Fusion*, vol. 2, pp. 1120-1125, 2002.

Musicki, D., Evans, R. and Stankovic, S. "Integrated probabilistic data association," *IEEE Transactions on Automatic Control*, vol. 39, no. 6, pp. 1237-1241, 1994.

Nadaraya, E. A., "On estimating regression," *Theory of Probability and its Applications*, vol. 10, pp. 186-190, 1964.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P., *Numerical Recipes in C*, Cambridge, MA: Cambridge University Press, 1992.

Pritsker, M. G., "Non-parametric density estimation and tests of continuous time interest rate models," *Review of Financial Studies*, vol. 11, pp. 449-487, 1998.

Provost, S. B., "Moment-based density approximants," *The Mathematica Journal*, vol. 9, pp. 727-756, 2005.

Provost, S. B. and Rudiuk, E. M. "Moments and densities of test statistics for covariance structures," *International Journal of Mathematical and Statistical Sciences*, vol. 4, no. 1, pp. 85-104, 1995.

Polonik, W. and Yao, Q., "Conditional minimum volume predictive regions for stochastic processes," *Journal of the American Statistical Association*, vol. 95, pp. 509-19, 2000.

Premont, G. and Sheng, Y., "Fast design of circular-harmonic filters using simulated annealing," *J. Opt. Soc. Am. A*, vol. 32, no. 17, pp. 3116-3121, 1993.

Pritsker, M., "Non-parametric density estimation and tests of continuous time interest rate models," *Review of Financial Studies*, vol. 11, no. 3, pp. 449-487, 1998.

Qu, M., Shih, F. Y., Jing, J., and Wang, H., "Automatic solar flare detection using MLP, RBF and SVM," *Solar Physics*, vol. 217, no. 1, pp. 157-172, 2003.

Qu, M., Shih, F. Y., Jing, J. and Wang, H., "Automatic solar flare tracking using image processing techniques," *Solar Physics*, vol. 222, no. 1, pp. 137-149, 2004.

Qu, M., Shih, F. Y., Ju, J. and Wang, H., "Automatic solar filament detection using image processing techniques," *Solar Phys.*, vol. 228, no. 1-2, pp. 119-135, 2005.

Rao, C. R., *Linear Statistical Inference and Its Applications*, New York: John Wiley & Sons, 1965.

Reid, N., "Saddlepoint methods and statistical inference," *Statistical Science*, vol. 3, no. 2, pp. 213-238, 1988.

Ren, H., Ping, Z., Bo, W., Wu, W. and Sheng, Y., "Multidistortion-invariant image recognition with radial harmonic Fourier moments", *Opt. Soc. Am. A*, vol. 20, no. 4, pp. 631-637, 2003.

Robinson, P. M., "Consistent non-parametric entropy-based testing," *Review of Economic Studies*, vol. 58, pp. 437-453, 1991.

Rosenblatt, M., "Conditional probability density and regression estimates," *Multivariate Analysis*, vol. 2, pp. 25-31, New York: Academic Press, 1969.

Scherrer, P. H., Bogart, R. S., Bush, R. I., Hoeksema, J. T., Kosovichev, A. G., Schou, J., Rosenberg, W., Springer, L., Tarbell, T. D., Title, A., Wolfson, C. J., Zayer, I. and the MDI Engineering Team, "The Solar oscillations investigation – Michelson Doppler Imager," *Solar Phys.*, vol. 162, pp. 101-128, 1995.

Schölkopf, B., Smola, A. J. and Müller, K.-R., "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computation*, vol. 10, pp. 1299-1319, 1998.

Shih, F. Y. and Kowalski, A. J., "Automatic extraction of filaments in Hα solar images," *Solar Phys.*, vol. 218, no. 1-2, pp. 99-122, 2003.

Shih, F. Y. and Mitchell, O. R., "Threshold decomposition of grayscale morphology into binary morphology," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 11, no. 1, pp. 31-42, 1989.

Shimojo, M., Yokoyama, T., Asai, A., Nakajima, H. and Shibasaki, K., "One solar-cycle observations of prominence activities using the nobeyama radioheliograph 1992-2004," *Publications of the Astromical Society of Japan*, vol. 58, no. 1, pp. 85-92, 2006.

Shohat, J. A., Tamarkin, J. D., "The problem of moments," *American Mathematical Society*, New York, 1943.

Solomon, H. and Stephens, M., "Approximations to density functions using Pearson curves," *Journal of the American Statistical Association*, vol. 73, no. 361, pp. 153-160, 1978.

Song, W. and Wang, J., "The differential rotation and longitudinal distribution of solar magnetic flux," *Astrophysical Journal*, vol. 624, pp. 137-140, 2005.

Sorenson, H. W., "Least-squares estimation: from Gauss to Kalman," *IEEE Spectrum*, vol. 7, pp. 63-68, 1970.

Steinegger, M., Denker, C., Goode, P. R., Marquettem, W. H., Varisk, J., Wang H., Otruba, W., Freishlich, H., Hanslmeier, A., Luo, G., Chen, D. and Zhang, W., "An overview of the new global high-resolution H-alpha network," Hvar Observatory Bulletin, vol. 24, no. 1, pp.179-184, 2000.

Strous, L. H., Scharmer, G., Tarbell, T. D., Title, A. M., and Zwaan, C., "Phenomena in an emerging active region. I. horizontal dynamics," *Astronomy and Astrophysics*, vol. 306, pp. 947, 1996.

Strous, L. H. and Zwaan, C., "Phenomena in an emerging active region. II. properties of the dynamic small-scale structure," *Astrophysical Journal*, vol. 527, pp. 435-444, 1999.

Tibshirani, R. and Hastie, T., "Local likelihood estimation," *Journal of the American Statistical Association*, vol. 82, pp. 559-567, 1987.

Tjostheim, D., "Non-linear time series: a selective review," *Scand. J. Statist.*, vol. 21, pp. 97-130, 1994.

Vapnik, V. N., *The Nature of Statistical Learning Theory*. New York:Springer-Verlag, 1995.

Vapnik, N. V., *Statistical Learning Theory*, Hoboken, NJ: John Wiley & Sons, Inc., 1998.

Wahba, G., Gu, C., Wang, Y. and Chappell, R., "Soft Classification, a.k.a. Risk Estimation, via Penalized Log Likelihood and Smoothing Spline Analysis of Variance," in D.H. Wolpert, editor, *The Mathematics of Generalization, Santa Fe Institute Studies in the Science of Complexity*, Addison-Wesley Publisher, 1995.

Wang, H., "Properties of remote flare ribbons associated with coronal mass ejections," *Astrophysical Journal*, vol. 618, pp.1012-1019, 2005.

Wang, H. and Goode, P. R., "Synoptic observing programs at Big Bear Solar Observatory," *ASP Conference Series*, ed. by K. S. Balasubramaniam; Jack Harvey; and D. Rabin, vol. 140, pp. 497-509, Sep. 1998.

Wang, H. and Zirin, H., "Flows around sunspots and pores," *Solar Phys.*, vol. 140, pp. 41-54, 1992.

Watson, G. S., "Smooth regression analysis," *Shankya Series A*, vol. 26, pp. 359-372, 1964.

Webb, D. F., Lepping, R. P., Burlaga, L. F., DeForest, C. E., Larson, D. E., Martin, S. F., Plunkett, S. P., and Rust, D. M., "The origin and development of the May 1997 magnetic cloud," *Journal of Geophysical Research*, vol. 105, no. A12, pp. 27251-27260, 2000.

Welch, G. and Bishop, G., "An introduction to the Kalman filter," *ACM SIGGRAPH 2001 Course*, Los Angels, CA, Aug. 2001.

Wilson, A "Solar variability: from core to outer frontiers," *The 10th European Solar Physics Meeting*, Prague, Czech Republic, vol. 1, no. 2, pp. 85, 2002.

Yu, K. and Jones, M. C., "Local Linear Quantile Regression," *Journal of the American Statistical Association*, vol. 93, pp. 228-237, 1998.