

Copulas for Image Processing

By

Xuexing Zeng

**In the fulfilment of the requirement
for the degree of Doctor of Philosophy**

**Centre for excellence in Signal & Image Processing
Department of Electronic & Electrical Engineering
University of Strathclyde**

**Supervised by
Professor Tariq S Durrani**

© September 2010

Copyright

The Copyright to this thesis belongs to the author under the terms of the United Kingdom Copyright Acts as qualified by University of Strathclyde Regulation 3.51. Due acknowledgment must always be made of the use of any material contained in, or derived from, this thesis.

Declaration

I declare that this thesis embodies my own research work and that it was composed by myself. Where appropriate, I have made acknowledgements to the work of others.

Xuexing Zeng

Acknowledgements

First of all, I would like to express my sincere gratitude towards my supervisor Professor Tariq S Durrani, for his guidance, encouragement and continuous support in the whole period of my PhD study. Professor Durrani was always there to listen and give me advice. He showed me different ways and perspectives to deal with a research problem, which benefited me a lot. He also encouraged me to improve my English and guided me on writing academic papers. Moreover, he involved me in the Bridging the Gap (BTG) project, which gave me great financial support. In the year of 2009, under his guidance, we successfully won the third place of the Thales Scottish Technology Prize.

Besides Professor Durrani, I would like to thank Dr. Stephan Weiss for his support during the process of my study. Thank him for taking some time out of his tight schedule to company me to visit Thales.

My sincere thanks also go to Ms. Sheila Forbes, who arranged supervision meetings for me and gave me lots of help.

I also would like to thanks Thales Group for providing image data to me.

A special thanks goes to my friends, Mr. Ijteba-ul-hasnain Shah, Peiran Shi, Xiao Yang, Adel Daas, Samir Bendoukha, Shahzad Ahmed Bhatti, Andrew Millar, Paul Murray, Waleed Al-Hanafy and Sherif Abd el halim Elgamel for their helps and all the fun we have had in the last three years. Furthermore, I would like to thank all the members of Centre for Excellence in Signal & Image Processing (CeSIP).

Last but not the least, I would like to thank my loving family, my grandparents, Jufang Zhang, Guangming Zeng, my parents, Fenglan Fan and Guohe Zeng, my wife, Yi Du, my brother, Xuewen Zeng and sister-in-law, Jinyun Huang for their love and supports.

Abstract

This thesis exploits copula function theory to address critical problems in image processing, ranging from image registration, image fusion, band selection for hyperspectral images to change detection. Copula functions offer insight into the dependence structure between two entities, when their individual properties are known. Mathematically, for two random variables with known marginal distributions, the copula function acts as a bridge between the marginal distributions and their joint distribution. The copula functions also extend the concept of correlations to a wider class of dependence structures that may be non-linear.

This thesis first presents the background on copula function theory, including copula density functions, conditional copulas, and addresses aspects of simulation, and modelling of copulas, as well as the estimation of copula parameters and optimal copula selection for later applications.

It is well known that the efficacy of registration methods relies on similarity measure such as mutual information and the associated divergence fields. Thus the area of mutual information is next explored and the divergence based information between distributions is studied. These concepts are then embedded in the framework of copula functions. A large class of divergences, such as Kullback-Liebler and the more generalised divergences such as Csiszar and Renyi-like divergences, the Bregman divergences are formulated within the structure of the attendant copula functions. New definitions are offered for the modified Bregman and Burbea-Rao divergences. It is shown that for all these divergence measures the associated copula functions definitions offer a much more effective approach to computing the complicated divergences.

Computing complicated calculation of these divergence based information by using copula density is proposed. The modified Bregman divergence is validated by comparing with conventional Bregman divergence by using smallest enclosing curve and K-means classification.

As a follow on, copula functions are used for image registration. Here copula functions are exploited for maximizing the divergence based information. Experiments are conducted to show that the copula based measures improve the accuracy of registration results; and comparisons are made with conventional methods such as the joint histogram based mutual information, and those based on Gaussian assumptions.

Another application of copulas explored in this thesis is the evaluation of the performance of multi-sensor image fusion methods based on criteria such as Mutual Information, Tsallis and Renyi divergence based information. Conventional fusion techniques include simple average, Principal Component Analysis (PCA), Gradient Pyramid (GP), Laplacian Pyramid (LP), Ratio Pyramid (RP) and Discrete Wavelet Transform (DWT). Their performance is compared by copula based methods. Using the criterion of fusion performance - Fusion Symmetry (FS)- it is shown that copula based measures provides effective evaluation of fusion performance, and that by tuning its key parameter, it is shown that the Tsallis divergence based information offers improved ability of discrimination.

There has been recent interest in analysing hyperspectral images, i.e. images collected with systems that cover a fine resolution over a large number of spectral bands. In this thesis methods are developed for the selection of bands in hyperspectral images that exploit copula based divergence based information. The band images which have higher mutual information with a reference image are selected. When the reference image is not available a spectral library can be used to generate the reference image. A rejection bandwidth measure is applied to reduce the redundancy information between neighbouring bands, which may synchronously have the higher mutual information with the reference image.

Finally, conditional copula techniques are developed for change detection that deal with the complicated situation where the objects in two images respectively are very similar or even same but their statistical distribution of pixel intensities vary remarkably due to the external factors such as climate changing or the use of different sensors. It leads to wrong change indicator. Conditional copulas solve this problem by training on the ‘no change’ areas between two images. It is shown that the results of copulas based measures are better than the ‘difference’ and ‘statistical similarity’ measures. The work is validated by plotting the Receiver Operating Characteristic (ROC) curve for the different methods.

The approach taken in this thesis is to seek new techniques to solve problems, seek to develop theoretical frameworks, and then test the concepts on simulation and wherever possible on real data.

The work has been presented at several international conferences, and a number of journal papers are currently being prepared.

Table of Contents

Copyright i

Declaration ii

Acknowledgements iii

Abstract iv

Table of Contents vi

1 Introduction 1

1.1 Introduction to copulas in image processing..... 2

1.1.1 Copulas for Divergence based measures..... 3

1.1.2 Relevance to Image Registration Techniques 4

1.1.3 Copulas for Performance Evaluation of Image Fusions..... 5

1.1.4 Support for Hyperspectral Imaging..... 5

1.1.5 Conditional Copulas for Image Change Detection 6

1.2 Organisation of this thesis 7

1.3 Original Contribution 8

1.4 List of publications 9

2 Introduction to Copulas..... 11

2.1 Copula Definition 11

2.2 Properties..... 12

2.3 Copula example..... 13

2.4 Pearson correlation..... 15

2.5 Typical copulas 17

2.5.1 Clayton copula..... 18

2.5.2 Frank copula..... 18

2.5.3 Bivariate Gaussian copula 19

2.6 Copula parameter estimation..... 22

2.6.1 Kendall’s tau and Spearman correlation based method 22

2.6.2 Maximum Likelihood Estimation (MLE) method 24

2.6.3 Comparing the methods of copula parameter estimation..... 26

2.7 Copula distribution function computation..... 27

2.7.1 Calculation of Clayton, Frank and Gaussian Copula 27

2.7.2 Calculation of joint cumulative distribution using Copulas..... 28

2.8	Random variables generation for Copulas	32
2.9	Copula density function	35
2.10	Copula selection	42
2.11	Conclusion.....	44
3	Divergence based information using copulas	46
	Summary	46
3.1	Introduction	46
3.2	Problem description.....	48
	3.2.1 Use of copulas	49
3.3	Generalised divergence based information and their copula representation.....	50
	3.3.1 Csiszar divergence.....	50
	3.3.2 Renyi-like divergence	54
	3.3.3 Bregman divergence.....	57
	3.3.4 Modified Bregman divergence.....	59
	3.3.5 Smallest enclosing ball (curve)	63
	3.3.6 K-means classification by modified Bregman divergence.....	64
	3.3.7 Burbea-Rao divergence	66
	3.3.8 Modified Burbea-Rao divergence	66
	3.3.9 Comparison of divergences	68
3.4	Calculation of divergence based information using copulas.....	69
3.5	Conclusion.....	74
4	Copulas for image registration and evaluation of image fusion algorithms	76
	Summary	76
4.1	Introduction	76
4.2	Current techniques and problem description.....	78
4.3	Copulas based image registration.....	79
	4.3.1 Image registration criterion	79
	4.3.2 Optimization method.....	80
	4.3.3 Image registration techniques.....	81
	4.3.4 Synthetic medical image registration	83
	4.3.5 Medical image (CT-MRI) registration	87
	4.3.6 Visible light and thermal image registration	89
4.4	Copulas based performance evaluation of image fusion.....	91
	4.4.1 Fusion Factor.....	91
	4.4.2 Fusion Symmetry	95
	4.4.3 Divergence based information using copulas for Gaussian distributed data	95
	4.4.4 Performance evaluation of image fusion using copulas.....	96
4.5	Conclusion	99

5	Band selection of Hyperspectral images.....	101
	Summary	101
	5.1 Introduction	101
	5.2 Current techniques and problem description.....	103
	5.3 Application to hyperspectral images	104
	5.4 Conclusion.....	108
6	Conditional copula for image change detection	109
	Summary	109
	6.1 Introduction	109
	6.2 Image change detection techniques review	110
	6.3 Comparison of Pearson system, Gram-Charlie series, Edgeworth series and Kernel smoothing techniques	112
	6.4 Problems description	113
	6.5 Applications	114
	6.5.1 Change detection for remotely sensed images	116
	6.5.2 Change detection for medical images	118
	6.6 Conclusion.....	120
7	Conclusion and future work.....	121
	7.1 Introduction	121
	7.2 Further research.....	127
	Appendix	129
	Appendix 1: Computation and Simulation of Exponential and Rayleigh Copula	129
	Appendix 2: Simulation of Bivariate Gaussian distribution	134
	Appendix 3: Algorithm of rejection bandwidth and setting the complementary threshold for band selection	134
	Appendix 4: Pearson System distribution	135
	Appendix 5: Gram-Charlie and Edgeworth Series.....	136
	Appendix 6: Kernel smoothing technique.....	138
	Bibliography	140

List of Table

2.1	Copula parameter expressed in terms of the Kendall’s tau and Spearman rank correlation.....	25
2.2	Clayton, Frank and Gaussian copula density functions	38
3.1	Definition of Csiszar, Renyi-like, Bregman and Burbea-Rao divergence	48
3.2	Csiszar divergence.....	51
3.3	Csiszar divergence based information expressed in terms of copula density functions .	54
3.4	Renyi-like divergence	55
3.5	Renyi-like divergence based information expressed in terms of copula density functions	57
3.6	Bregman divergence.....	59
3.7	Modified Bregman divergence based information expressed in terms of copula density functions	62
3.8	Burbea-Rao divergence	67
3.9	Modified Bregman divergence based information expressed in terms of copula density functions	70
4.1	Result of synthetic MRI image registration without noise.....	85
4.2	Results of synthetic MRI image registration with 10% salts and pepper noise	86
4.3	Results of synthetic MRI image registration with 20% salts and pepper noise	86
4.4	Results of synthetic MRI image registration with 30% salts and pepper noise	87
4.5	Resultst of CT-MRI image registration.....	89
4.6	Results of visible light-thermal image registration for Figure 4.6	90
4.6	Results of visible light-thermal image registration for Figure 4.6	90
4.7	Results of visible light-thermal image registration for Figure 4.7	90
4.8	Results of visible light-thermal image registration for Figure 4.8	90
4.9	Performance evaluation of image fusion.....	99
5.1	15 selected bands.....	108

List of Figures

2.1	Copula distribution function for Eq. (2.3).....	14
2.2	Contour of copula distribution function for Eq. (2.3)	14
2.3	Clayton copula with parameter = 2	29
2.4	Frank copula with parameter = 0.5.....	29
2.5	Gaussian copula with parameter =0.5	29
2.6	Gaussian and Student t marginal probability density functions	30
2.7	Gaussian copula distribution	31
2.8	Contour of Gaussian copula distribution.....	31
2.9	Joint cumulative distribution	32
2.10	Contour plot of Joint cumulative distribution	32
2.11	Simulation of Clayton copula.....	36
2.12	Simulation of Frank copula	36
2.13	Simulation of Gaussian copula.....	36
2.14	Gaussian copula density function.....	38
2.15	Joint probability density function.....	38
2.16	Joint probability density function of bivariate Gaussian distribution	40
2.17	Contour of joint probability density function of bivariate Gaussian distribution	40
2.18	Joint probability density function using Gaussian copula.....	41
2.19	Contour of joint probability density function using Gaussian copula.....	41
2.20	Comparison of joint probability density function obtained by bivariate Gaussian distribution and Gaussian copula respectively	42
3.1	Comparison of Csiszar divergence.....	51
3.2	Comparison of Renyi-like divergence.....	56
3.3	Bregman divergence.....	58
3.4	Comparison of Bregman divergence	59
3.5	Comparison of modified Bregman divergence	62
3.6	Smallest enclosing Itakura-Saito ball and smallest enclosing modified Itakura-Saito curve	65
3.7	K-means classification by using squared Euclidean, Itakura-Saito and modified Itakura-Saito divergence as distance	65
3.8	Burbea-Rao divergence	67
3.9	Comparison of Burbea-Rao divergence	68
3.10	Comparison of modified Burbea-Rao divergence.....	70
3.11	Inverse standard Gaussian cumulative distribution.....	73
3.12	Bivariate Gaussian distribution and Gaussian copula based mutual information.....	74
4.1	Square neighbourhood.....	82
4.2	Circle neighbourhood.....	82
4.3	Flow chart of image registration	84

4.4	Image Registration of synthetic MRI image (a) Reference image (b) Float image (c) Float image with 10% salts and pepper noise (d) Float image with 20% salts and pepper noise (e) Float image with 30% salts and pepper noise (f) Registered image	85
4.5	(a) Referenec image (MRI) (b) Float image (CT) (c) Registred CT overlaid by semi-transparent (d) Sum of each corresponding pixel intensity of registered CT and MRI	88
4.6	The first image registration of visilbe light-thermal image (a)Visible light image (b) thermal image (c) Registered thermal image using copula method (d) Registered thermal image overlaid on the semi-transparent visible light image (e) Registered thermal image using joint histogram method (f) Registered thermal image using joint histogram method overlaid on semi-transparent visible light image	92
4.7	The second image registration of visilbe light-thermal image (a)Visible light image (b) thermal image (c) Registered thermal image using copula method (d) Registered thermal image overlaid on the semi-transparent visible light image (e) Registered thermal image using joint histogram method (f) Registered thermal image using joint histogram method overlaid on semi-transparent visible light image	93
4.8	The third image registration of visilbe light-thermal image (a)Visible light image (b) thermal image (c) Registered thermal image using copula method (d) Registered thermal image overlaid on the semi-transparent visible light image (e) Registered thermal image using joint histogram method (f) Registered thermal image using joint histogram method overlaid on semi-transparent visible light image	94
4.9	Mutual information and Tsallis, Renyi divergence based mutual information using Gaussian copula.....	96
4.10	Fused images (a) Infrared image (b) Visible image (c) Fused image by PCA (d) Fused image by averaging (e) Fused image by gradient pyramid (f): Fused image by Laplace pyramid (g): Fused image by Ratio pyramid (h): Fused image by discrete wavelet transform.	98
5.1	Hyperspectral images (a) Band 5 (b) Band 50 (c) Band 125 (d) Band 200 (e) Band 218 (f) Estimated Reference image obtained by averaging images from band 170 to 210	106
5.2	Mutual information between each band image and reference image.....	107
5.3	Selected band number using rejection bandwidth and complementary measure.....	107
6.1	Probability density function estimation.....	113
6.2	Flow chart of change detection using conditional copula	115
6.3	Change detection of SAR images of Earthquake using conditional copula (a) Image taken on 14 May 2006 (around two years before earthquake) (b) SAR image after flooding on 14 May 2008 (two days after earthquake) (c) Change detection results by pixel intensity difference based change detector (d) Change detection result by statistical similarity based change detector (e) Change detection results using Clayton copula based change detector	116

6.4	ROC plot for of performance of image change detection	118
6.5	Change detection of CT images (a) Before treatment (b) After treatment (c) Change detection result using difference measure (d) Change detection using Frank copula measure.....	119

List of Acronyms

AVIRIS Airborne Visible/Infrared Imaging Spectrometer

BSS Blind Source Separation

CDF Cumulative Distribution Function

CML Canonical Maximum Likelihood

CT Computed Tomography

DD Difference Detector

DWT Discrete Wavelet Transform

FF Fusion Factor

FGM Farlie Gumbel Morgenstern

FS Fusion Symmetry

GIS Geographic Information Systems

GP Gradient Pyramid

ICA Independent Component Analysis

IFM Inference for Margins

LP Laplacian Pyramid

MI Mutual information

MIMO Multi-Input Multi-Output

MLE Maximum Likelihood Estimation

MRD Mean Ratio Dectector

MRI Magnetic Resonance Imaging

PCA Principal Component Analysis

PET Positron Emission Tomography

PDF Probability Density Function

RP Ratio Pyramid

ROC Receiver Operating Characteristic

SAR Synthetic Aperture Radar

SKLD Symmetric Kullback-Leibler Distance

Chapter 1

Introduction

Digital processing of images has gained in popularity with the ready availability of conversion devices, and the attendant ability to perform computing operations or algorithms for a variety of purposes, in order to improve the images or to extract relevant information. Functionally an image may be considered as a two-dimensional signal and can be described by the function $z=f(x, y)$ where x and y are spatial coordinates and the amplitude z is the corresponding intensity or gray level of the image. The images are called digital images when x , y and z are finite. A digital image is composed of a finite number of element values at each location; the element is called pixel (picture element). With the availability and use of fast computers, digital image processing has become increasingly practical and popular [Gonzalez & Woods, 2008].

While there exists a large range of operations; typical digital image processing operations may be classified into a generic set of actions, including [Gonzalez & Woods, 2008]:

- Image transformations: Such as image enlargement, reduction, translating and rotation.
- Colour corrections: Such as brightness and contrast adjustments which make the images clearer to see.
- Image enhancement: Operations to improve the subjective appearance of an image, such as noise reduction, image de-blurring, i.e. bringing into focus;
- Image registration: Alignment of two or more images.
- Image change detection: Estimate change areas between two or more images.

Chapter 1: Introduction

- Image segmentation: Partitioning image into multiple segments which is more intuitive and easier to analyse.
- Image recognition: Determining whether the image data contains some specific object, feature or activity.
- Image fusion: Combining a series of relevant two or more images to one single image for obtaining more information in this single image.
- Image restoration: Making the image more closely resemble the original image which is noised or blurred.
- Image coding or compression: To reduce the footprint of an image without necessarily reducing its information content.

A large compendium of algorithms exists to perform the above operations [Gonzalez & Woods, 2008]. Almost all rely on apriori information or prior knowledge to improve an image. Several algorithms are based on implicit assumptions regarding the nature of the data or the image, such as Gaussianity or linearity, where specific evidence or information is not available. The image registration, performance evaluation of image fusion, band selection for hyperspectral images and change detection will be addressed in this thesis.

1.1 Introduction to copulas in image processing

As a preamble, in this Section we give a general background to the approach proposed in this thesis for the use of copulas in image processing. Detailed definitions and mathematical descriptions are given in later chapters.

Many digital image processing techniques such as image registration, change detection, image fusion and image classification require an estimate of dependence and the associated joint probability density function between image pixel intensities. In the familiar multivariate Gaussian case, the dependence is described by the Pearson linear correlation, and the associated joint probability density function is easy to

Chapter 1: Introduction

estimate based on a multivariate Gaussian distribution model of image pixel intensities.

However, it has been proven that non-linear dependence in stochastic processes renders the use of Pearson linear correlations impractical and mutual information is more robust than conventional Pearson correlation for intensity based image registration [Viola & Wells, 1997].

Furthermore, the Pearson linear correlation is only effective for elliptical (joint) distributions [Fang et al., 1987]. The elliptical distribution represent joint distribution that have elliptical contours, the definition is given in Chapter 2. The well-known elliptical distributions are multivariate Gaussian, Student t , Logistic and Laplace which means that their corresponding marginal distributions are Gaussian, Student t , Logistic and Laplace distributions respectively [Landsman & Valdez, 1999].

However, non-elliptical distributions always exist in the real world. For examples, radar images, it is well known that it's distribution of pixel intensity are considered as gamma distributions, therefore multi-Gamma distribution were proposed to deal with Radar image registration and change detection [Chatelain et al., 2007]. Besides, Magnetic Resonance images are considered as Rician distribution [Sijbers et al., 1998]. Specifically in the real world, the distribution of pixel intensity of images may be arbitrary, therefore their types of marginal distributions may be not consistent, and there is no multivariate distribution available to deal with these non-elliptical cases.

In order to solve these problems, a new statistical tool called the copula function will be investigated in this thesis. The copula function is a joint distribution function with uniform marginal distributions [Nelsen, 1999]. It represents a dependency structure which may be non-linear of multidimensional variables and can be estimated in terms of only related marginal distributions. Another significant advantage of copula function is that it is able to estimate the joint distribution for arbitrary marginal distributions.

1.1.1 Copulas for divergence based measures

The divergence-based information (see Section 3.1 of Chapter 3 for the definition)

Chapter 1: Introduction

such as mutual information gives a measure of the distance between two distributions, and it is actually the Kullback-Leibler based information [Thomas & Joy, 1991]. This measure represents an eminent tool to determine the quantitative dependency between two or more variables. There are four categories of conventional and generalized divergences that include the Csiszar, Renyi-like, Bregman and Burbea-Rao divergences [Pardo & Vajda, 2003], [Martin, 2006] are applied widely in the image processing field, in applications such as image registration, image classification [Banerjee, et al., 2005], [Ozturk & Abut, 1990], [Pluim, et al., 2004], [He, et al., 2003], [Maes et al., 1997]. In this thesis the corresponding divergences based information for these four categories of measures will be investigated in some detail, and copula function based frameworks will be developed to show the applicability of the copula functions to evaluate these measures.

It is well-known that the joint probability density function can be expressed as the product of copula density function and marginal probability density functions [Durrani & Zeng, 2007]. Copulas thus offer a natural way to estimate divergence-based information, and more details are given in Chapter 3.

In Chapter 3, we show that both of the Csiszar and Renyi-like divergences based information can be expressed directly in terms of copula density function only, and a modified version of the Bregman divergence based information can be expressed in terms of copula density only as well. By transforming the formula of Burbea-Rao divergence, it will be shown that Burbea-Rao divergence is a special case of Bregman divergence, so Burbea-Rao divergence based information can also be expressed by a copula density function. This means that only copula parameters are needed to estimate, the joint probability density function, and marginal density functions are not required to estimate for the divergence based information.

1.1.2 Relevance to image registration techniques

The use of divergences based information has been widely accepted as one of the most accurate and robust image registration techniques [He, et al., 2003], [Pluim, et al., 2004]. Furthermore, the four categories of divergences mentioned above, are more generalized than mutual information. And in later chapters it is shown that they provide improved ability to control the measurement sensitivity and hence better

Chapter 1: Introduction

accuracy for image processing.

The key to the estimation of divergence based information is the evaluation of the joint probability density function. It has already been stated earlier, that it is extremely difficult to estimate the joint probability density function without apriori information about the underlying models such as multi-Gaussian, multi-Gamma. However, actually these models may not be suitable since the marginal distributions may not be always Gaussian or Gamma and may be arbitrary. The comments hold for different type of marginal distributions. Copulas offer a more general model and are able to deal with cases where the marginal distributions are not consistent.

The pixel intensity based image registration process can be described as: given the first image as reference image and the second image as float image, by rotating, translating and rescaling the float image to find the optimal space transformation for the float image until the divergence based information between the overlapping parts of reference and transformed float images reaches to the maximum. Four categories and eleven types of divergences based information based on copulas will be applied to achieve image registration. The results will be compared with conventional techniques including the Gaussian assumption based mutual information and the non-parametric method which uses the joint histogram of pixel intensity to estimate the joint probability density function [Maes, et al., 1997].

1.1.3 Copulas for performance evaluation of image fusions

Conventional techniques to evaluate the effectiveness of image fusion techniques include the following: Simple average, Principal Component Analysis (PCA) [Jia, 1998], Gradient Pyramid (GP) [Burt & Kolczynski, 1993], Laplacian Pyramid (LP) [Burt & Adelson, 1983], Ratio Pyramid (RP) [Toet, 1989] and Discrete Wavelet Transform (DWT) methods [Mallat, 1989]. In this thesis, copula methods are developed as an alternative approach to assess the effectiveness of image fusion techniques for multi-sensor images without ground truth.

1.1.4 Support for hyperspectral imaging

Chapter 1: Introduction

Another application of copula functions will be explored in the context of band selection for hyper-spectral image processing. Hyperspectral imaging can cover the entire spectral band range, with around 10 nm spectral bandwidth. Hyper-spectral images represent high-dimensional data with relatively much more information than visible light images. However such large amounts of data is not convenient for the further image processing such as image classification, transmission and real time application. Hence band selection for the high-dimensional image is necessary.

Given a reference image, copulas based mutual information may be calculated and band images chosen that have the higher mutual information with the reference image. To reduce the reliance of reference image, spectral signatures may be used, and other thresholding techniques may be exploited to improve the results of band selection by removing the redundant information between neighbouring bands [Guo, et al., 2006].

1.1.5 Conditional copulas for image change detection

The requirement of automated techniques for detecting the changes between images of the same scene, arise in several image processing applications, e.g. in remote sensing [Bruzzone & Prieto, 2002], where changes in scenes offer relevant information; or in patient monitoring using biomedical imaging [Wakuya et al., 2007], where the efficacy of treatment is recognized by changes in images taken at different times, and in many more applications. In this thesis copula based techniques are developed for image change detection. Since the copula function reveals the dependence between the random variables, once the dependency structure is estimated; this characteristic can be used for complicated image change detection. For example, some observations in two images may be very similar or even same and the corresponding pixel intensity distributions should be similar or same as well. However if the two images were obtained by different sensors, or under different weather conditions or due to other external conditions, the pixel intensity distribution may be affected remarkably, and it is this difference in distributions that is readily identified by copula based techniques, that offer the prospect of new approaches to change detection, whereas statistical similarity based change detection techniques can often lead to erroneous change detection [Inglada, 2003], [Inglada & Mercier, 2007].

Chapter 1: Introduction

1.2 Organisation of this thesis

The introduction in Chapter 1 has set out some of issues concerned with Image Processing, such as the Pearson correlation being effective only when dealing with elliptical distributions, and the difficulties in the estimation of the divergence based information in image processing, and more. It is conjectured (and proven in subsequent chapters), that copula functions may resolve these problems.

Chapter 2 introduces the copula functions and their properties, description is provided of three typical copulas such as Clayton, Frank, Gaussian copula, and two new copulas called Exponential and Rayleigh copula are introduced. Details of the definition of copula density functions and associated computation of copula distribution function are considered, and aspects of the generation of random variables for copulas, copula parameter estimation and the optimal copula selection techniques are presented. This chapter lays the foundation for the use of copulas later on.

Chapter 3 is concerned with information theoretic concepts related to measure such as distances between distributions, given in terms of divergences. Four types of divergences are studied here: Csiszar, Renyi-like, modified Bregman and Burbea-Rao divergence based information, and these are embedded with the framework of associated copula density functions only. Illustrative examples are included. The modified Bregman divergence is evaluated and compared with the conventional Bregman divergence by using the smallest enclosing curve and K-means classification.

Chapter 4 proposes an approach for image registration based on maximizing divergence-based information using copulas. Theoretical concepts are first developed and then the work is illustrated first on synthetic images with added noise to verify the registration techniques. The work is then extended to test the registration algorithms firstly CT (Computerized Tomography) and MRI (Magnetic Resonance Imaging) images, then data available from Thales, visible light images and thermal images are registered. To assess the performance, copula methods are compared with the classic methods such as Gaussian assumption based mutual information and joint histogram based mutual information.

Besides, Tsallis and Renyi divergence based information is applied to evaluate the performance of image fusion, higher information between fused image and input

Chapter 1: Introduction

images are considered as better performance [Cvejic et al., 2006]. The fusion algorithms such as simple average, Principal Component Analysis (PCA), Gradient Pyramid (GP), Laplacian Pyramid (LP), Ratio Pyramid (RP) and Discrete Wavelet Transform (DWT) methods are assessed for multi-sensor image data follow the criterion of Fusion Symmetry [Stathaki, 2008]. The Tsallis divergence based information has been found that it can provide better ability of discrimination by adjusting its parameter than classic mutual information.

In Chapter 5 a new method is introduced, which uses copulas-based ‘mutual information’ to solve the classic problem of band selection for hyperspectral images. Band images which have higher mutual information with reference image are retained. The copula functions are used to evaluate the mutual information. Results of experiment are provided that compare the proposed approach with conventional methods such as Gaussian assumption and joint histogram based mutual information.

Chapter 6 addresses the problems in image change detection for remote sensing, for images taken by the different sensors, or in the different weather condition. Here the issue that the pixel intensity distributions are much different is exploited to analyse the data, while the observations may be very similar or even the same. Conditional copulas are used to solve this problem. For two registered images, information for neighbouring pixels is exploited in addition to the value of each pixel to calculate the symmetric Kullback-Leibler divergence (SKLD) between neighbours as the change indicator. The results are compared with conventional methods such as statistical similarity based method and pixel based method, and illustrated on real data.

Finally, Chapter 7 provides some discussion and conclusions on the work reported in this thesis, and further research to extend the work is also proposed.

1.3 Original contribution

The major contribution of this thesis is the invention of copulas for divergence-based information, and the divergence framework including the frequently-used Csiszar, Renyi-like, modified Bregman and Burbea-Rao divergences. The copula-based information offers a robust method to estimate all kinds of divergence-based

Chapter 1: Introduction

information even when the marginal distributions are not consistent. The approach thus handles the cases where the marginal distributions can be arbitrary. Copulas are applied for image registration by searching the optimal spatial transformation which makes the divergence-based information maximal between the overlapping parts of reference image and float image transformed through operations such as rotation, translation and rescaling.

Furthermore, copulas based mutual information is proposed for the evaluation of the performance of image fusion where the ground truth is not available. Higher information between fused image and input images is considered and shown to yield better performance [Cevjic et al., 2006].

In addition, while hyper-spectral images offer more information than visible light images, they are too large to conduct the further processing. Thus a new method named copulas-based mutual information is proposed for band selection by choosing the band images which have higher mutual information with reference image [Guo et al., 2006].

Finally, the pixel intensity distributions are affected when the images taken by the different sensors, as in the different weather condition. This can make the pixel intensity distributions be much different although the observations in a real situation may be very similar or even are the same. Since copulas can measure the complex dependency relationship which allows evaluation of this dependence, which may not be linear between the random variables; this property is exploited for change detection [Inglada, 2003], [Mercier et al., 2008].

1.4 List of Publications

During the duration of the PhD degree studies, the following research papers have been published:

- X. Zeng & T.S Durrani, “Image change detection using copulas, 9th IEEE International Conference on Signal Processing”, pp. 909-913, Oct. 26-29, Beijing, China, 2008.

Chapter 1: Introduction

- T. S Durrani & X. Zeng, “Copula based Divergence Measures and their use in Image Registration”, 17th European Signal Processing Conference, pp. 1309 -1313, Aug. 24-28, Glasgow, U.K, 2009.
- X. Zeng & T. S Durrani, “Band Selection for Hyperspectral Images using Copulas-based Mutual Information”, IEEE Workshop on Statistical Signal Processing, pp. 341-344, 31 Aug – 3 Sep, Cardiff, U.K, 2009.
- X Zeng & T.S Durrani, “Performance Evaluation of Image Fusion using Copulas”, 10th IEEE International Conference on Signal Processing, Oct. 24-28, Beijing, China, 2010. Accepted on 15 July 2010.

Earlier publication includes:

- T.S Durrani and X. Zeng, “Copulas for bivariate probability distributions”, IET electronic letters, Vol. 43 , Issue 4, pp.248-249, 2007.

Chapter 2

Introduction to Copulas

The copula function describes the ‘coupling’ between a multivariate function and its marginals. The concept was first proposed in a famous theorem by Abe Sklar in 1959 [Sklar, 1959]. Traditionally, when a multivariate joint distribution function is given, the marginal distribution functions can be found through straight forward integration. However, conversely, if the marginal distributions are known, a unique joint distribution cannot be found. With the help of copulas, the joint distribution functions can be estimated for the arbitrary marginal distribution functions since a copula can link the joint distribution to its marginal distributions, it has become a new popular statistical tool to model the dependency structure among variables. One of the most interesting new statistical ideas to emerge from financial options theory in recent years is the copula [Murphy et al., 2006] and it has been recently applied in many fields such as earth science [Li et al., 2008] and risk analysis [Perrone et al., 2006]. However, as yet copulas have not found a role in image processing, and are an almost unexplored area. Nevertheless, considering the robustness, convenience, accuracy and effectiveness of copulas, their use in the area of image processing offers a worthwhile opportunity for further exploration.

2.1 Copula definition

Sklar’s theorem [Sklar, 1959] lays the foundation for the definition of the copula functions which can be described as:

Chapter 2: Introduction to Copulas

Let $H(x, y)$ be a joint cumulative distribution function with two marginal cumulative distribution functions $F(x)$ and $G(y)$ for two random variables x and y respectively, then there exists a copula function C for all x and y such that

$$H(x, y) = C(F(x), G(y)) \quad (2.1)$$

If $F(x)$ and $G(y)$ are continuous, then the copula function C is unique. According to Sklar's theorem, the two-dimensional copula function can be defined as:

A two-dimensional copula is a bivariate cumulative distribution function defined on the unit cube with uniform marginal distributions on the interval $[0, 1]$. As the extension, the copula function $C(u, v)$ has been defined as [Nelsen, 1999]:

$$C(u, v) = H(F^{-1}(u), G^{-1}(v)) \quad (2.2)$$

where $u = F(x)$ and $v = G(y)$.

In addition, for any $u \in [0, 1]$ and $v \in [0, 1]$, copula function should satisfy the following conditions [Nelsen, 1999]:

- $C(u, 1) = u$; $C(1, v) = v$ and $C(u, 0) = 0$; $C(0, v) = 0$ (2.3)
- If $u_1 \leq u_2$; $v_1 \leq v_2$ then, $C(u_2, v_2) - C(u_2, v_1) - C(u_1, v_2) + C(u_1, v_1) \geq 0$ (2.4)

2.2 Properties

There is an important property which is called Fréchet-Hoeffding bounding for copula function and can be described as [Nelsen, 1999]:

$$W(u, v) = \max(u + v - 1, 0) \leq C(u, v) \leq \min(u, v) = M(u, v) \quad (2.5)$$

Here, $M(u, v)$ is usually called Fréchet-Hoeffding upper bound. It means that for any given copula function C , we can get $C(u, v) \leq M(u, v)$ for arbitrary $u, v \in [0, 1]$. Similarly, the Fréchet-Hoeffding lower bound is defined by $W(u, v) = \max(u + v - 1, 0)$. It has been proven that both of the Fréchet-Hoeffding upper and lower bounds are copulas themselves.

2.3 Copula example

After introduction of the definition of copula function, a copula example with one parameter will be derived from the joint cumulative distribution function $H_\theta(x, y)$.

$$H_\theta(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y+\theta xy)} & x \geq 0 \text{ and } y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

where θ is a parameter in the range of $[0, 1]$. The corresponding inverse marginal cumulative distribution function of u and v in $[0, 1]$ are:

$$F^{-1}(u) = -\ln(1-u); G^{-1}(v) = -\ln(1-v)$$

According to the copula definition in Eq. (2.2), the corresponding copula function can be found as:

$$C_\theta(u, v) = u + v - 1 + (1-u)(1-v)e^{-\theta \ln(1-u)\ln(1-v)}$$

Specifically, let $\theta = 0.5$, then the corresponding copula function becomes:

$$C(u, v) = u + v - 1 + (1-u)(1-v)e^{\frac{\ln(1-u)\ln(1-v)}{2}} \quad (2.6)$$

The diagram of this copula distribution function is shown in Figure 2.1. To obtain the better visual effects, the contour figure for displaying the isoline of copula distribution function has been drawn and shown in Figure 2.2.

This example shows how does the copula distribution function can be generated when the joint probability distribution is known. It can be achieved by the following two steps:

Firstly, estimate the marginal distribution function and inverse marginal distribution function from the joint distribution function.

Secondly, calculate the copula distribution function by using the Eq. (2.2), this shows that the copula distribution function is just the joint cumulative distribution of two inverse marginal cumulative distribution functions.

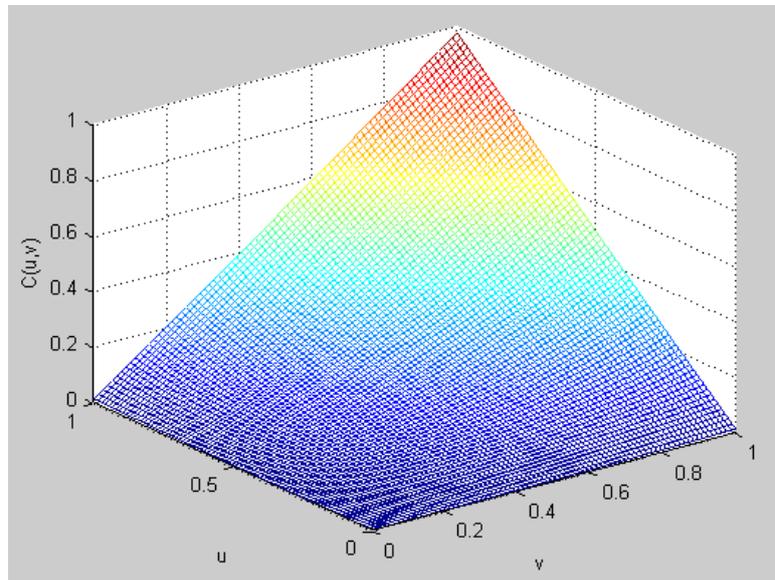


Figure 2.1: Copula distribution function for Eq. (2.6)

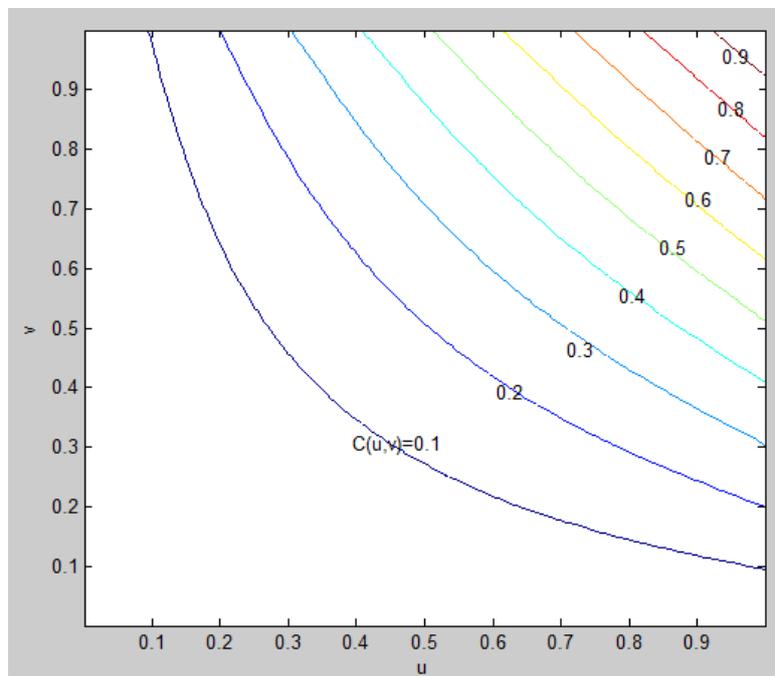


Figure 2.2: Contour of copula distribution function for Eq. (2.6)

2.4 Pearson correlations

Before the research of copula function, we review the Pearson correlation which is an important correlation measure in statistics. The Pearson correlation shows whether a pairs of variables are related to each other and how strongly they are related. The Pearson correlation coefficients describe the strength of the relationship between two variables, and the range of correlation coefficient is from -1.0 to +1.0. The Pearson correlation coefficient of 1.0 indicates a perfect positive relationship since high values in one of the variables are related to high values in another variable. The correlation coefficient 0 means no relationship between two variables. Moreover, the correlation coefficient -1 means a perfect negative relationship since variables with high values are related perfectly to low values in another variable, and conversely, low values in one of variable are perfectly related to high values in another variable.

Pearson correlation coefficient is also called linear correlation and is defined as follows [Gibbons, 1985]:

$$\rho_{XY} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y} \quad (2.7)$$

where ρ_{XY} is the Pearson correlation between two random variables X and Y , μ_X and μ_Y and are expected values, σ_X and σ_Y are standard deviations of X and Y respectively. Here E is the expectation operation and cov is covariance.

For $\mu_X = E(X)$; $\sigma_X^2 = E(X^2) - E^2(X)$ and $\mu_Y = E(Y)$; $\sigma_Y^2 = E(Y^2) - E^2(Y)$. The Pearson correlation becomes:

$$\rho_{XY} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad (2.8)$$

The correlation exists only when two standard deviations are finite and both of them are nonzero. If the variables are independent then the correlation is 0, however, the converse is not true, that is to say, the variables are not possibly independent when the correlation is zero because correlation coefficient can be used to describe linear

Chapter 2: Introduction to Copulas

dependencies between two variables. For example: assume a random variable X is uniformly distributed on the interval from -1 to 1 , and $Y = X^2$. Then Y is completely determined by X , so that X and Y are obviously dependent, but their correlation is zero.

Pearson correlation is a very important correlation and is applied in several areas. However, it is not the perfect measure of dependence since correlation is limited to elliptical distributions [Kumar & Shoukri, 2007] which can be defined as following [Fang et al., 1987]:

An elliptical distribution [Fang et al., 1987] is the generalization of the multivariate normal distribution. Let ψ_n be a class of functions $\psi(t)$ where $t \in [0, +\infty)$ such that function $\psi(\sum_{i=1}^n t_i^2)$ is an n -dimensional characteristic function. It is clear that

$$\psi_n \subset \psi_{n-1} \dots \subset \psi_1$$

Consider a n -dimensional random vector $X = (X_1, X_2, \dots, X_n)^T$, the random vector X has a multivariate elliptical distribution, denoted by $X \sim E_n(\mu, \Sigma, \psi)$, if its characteristic function can be expressed as:

$$\varphi_X(t) = e^{it^T \mu} \varphi\left(\frac{1}{2} t^T \Sigma t\right) \quad (2.9)$$

for some column vector μ , n by n positive definite matrix Σ and for some function $\psi(t) \in \psi_n$ which is called the characteristic generator. If the joint density function exists, it has the following form [Fang et al., 1987]:

$$f_X(x) = \frac{c_n}{\sqrt{|\Sigma|}} g_n\left[\frac{1}{2}(x - \mu)^T \Sigma^{-1}(x - \mu)\right] \quad (2.10)$$

where function g_n is called the density generator and

$$c_n = \frac{\Gamma(n/2)}{(2\pi)^{n/2}} \left[\int_0^\infty x^{\frac{n}{2}-1} g_n(x) dx \right]^{-1} \quad (2.11)$$

where function Γ is called Gamma function [Abramowitz & Stegun, 1965] which is defined as:

$$\Gamma(z) = \int_0^{\infty} t^{z-1} e^{-t} dt \quad (2.12)$$

when z is a positive integer, then $\Gamma(z) = (z-1)!$.

For example, the multivariate normal distribution has the density generator function:

$g(u) = e^{-u}$, then it is easy to get the joint density function:

$$f_X(x) = \frac{c_n}{\sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (2.13)$$

and $c_n = (2\pi)^{-\frac{n}{2}}$, so the joint multivariate normal distribution function can be written as:

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{|\Sigma|}} e^{-\frac{1}{2}(x-\mu)^T \Sigma^{-1}(x-\mu)} \quad (2.14)$$

Moreover, an important property of elliptical distribution is that any linear combination of elliptically distributed variables is elliptical. The common elliptical distributions are multivariate Gaussian, Student t distributions [Landsman & Valdez, 2002].

Pearson correlation is effective for elliptical distribution but may not be a suitable way of dealing with joint multivariate distribution without the assumption of elliptical distribution and it is also useless when the types of marginal distribution are not consistent.

2.5 Typical copulas

A number of copula functions have been defined in [Nelsen, 1999]. This thesis will consider some typical copulas. Firstly, the Archimedean copulas are widely applied in several fields such as finance by the reasons of simplicity and efficiency [Cherubini et al., 2006]. A bivariate Archimedean copula function $C(u, v)$ can be expressed as

Chapter 2: Introduction to Copulas

[Nelsen, 1999]:

$$C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v)) \quad (2.15)$$

It should satisfy the following conditions:

- a) $0 \leq u \leq 1, 0 \leq v \leq 1$. Function φ is strictly decreasing and $\varphi(1) = 0$.
- b) φ is convex function, that is to say, the second order of derivative of φ , $\varphi'' \geq 0$.

The function φ is called generator function. Two important Archimedean copulas which are used widely named Clayton and Frank copula will be introduced as follows.

2.5.1 Clayton copula:

The Clayton copula was firstly mentioned by [Clayton, 1978], it can be written by

$$C(u, v) = \max([u^{-\theta} + v^{-\theta} - 1]^{-1/\theta}, 0) \quad (2.16)$$

where $\theta \in [-1, \infty)$ and $\theta \neq 0$ and when $\theta \rightarrow 0$, the variables are independent. The generator function is:

$$\varphi(t) = (t^{-\theta} - 1)$$

2.5.2 Frank copula

The Frank copula was firstly introduced by [Frank, 1979] and it can be written by

$$C(u, v) = -\frac{1}{\theta} \ln \left[1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right] \quad (2.17)$$

where $\theta \in (-\infty, +\infty)$ and $\theta \neq 0$, when $\theta \rightarrow 0$, the variables are independent. The generator function is

Chapter 2: Introduction to Copulas

$$\varphi(t) = -\ln \frac{e^{-\theta t} - 1}{e^{-\theta} - 1}$$

Note that Archimedean copula function cannot independently exist without generator function, so the generator function must be specified.

2.5.3 Bivariate Gaussian copula

The Gaussian copula [Nelsen, 1999] is also called Normal copula. The bivariate Gaussian joint probability density function can be written as [Hamedani & Tata, 1975]:

$$f_{XY}(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_x}{\sigma_x}\right)^2 - \frac{2\rho xy}{\sigma_x\sigma_y} + \left(\frac{y-\mu_y}{\sigma_y}\right)^2\right]} \quad (2.18)$$

where μ_x , μ_y are the mean values, σ_x , σ_y are the standard deviations and ρ is the Pearson correlation of two random variables x and y respectively. The standard bivariate Gaussian joint probability density function can be obtained by letting $\mu_x = \mu_y = 0$ and $\sigma_x = \sigma_y = 1$ as:

$$f_{XY}(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} \quad (2.19)$$

The standard joint Gaussian cumulative distribution function $H(x, y)$ can be obtained by the following double integrals as:

$$H(x, y) = \iint_{x,y} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} dx dy \quad (2.20)$$

So, the Gaussian copula distribution function can be written as:

$$C(u, v) = \int_{-\infty}^{\varphi^{-1}(u)} \int_{-\infty}^{\varphi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{x^2-2\rho xy+y^2}{2(1-\rho^2)}} dx dy \quad (2.21)$$

where function φ is the standard Gaussian cumulative distribution function. The

Chapter 2: Introduction to Copulas

variables are independency when $\rho=0$. The closer the Pearson correlation coefficient is to 1 and -1, the stronger the correlation and negative correlation respectively between the variables.

More copula functions can be found in [Nelsen, 1999], besides, except in [Nelsen, 1999], two new copulas called Exponential and Rayleigh copula have been developed in [Durrani & Zeng, 2007]. Similar to Gaussian copula which extracts the dependency structure between two standard Gaussian distributions; these two new copulas extract the dependency structures between two exponential distributions and two Rayleigh distributions respectively. Considering two exponential marginal probability density functions $f_X(x)$ and $f_Y(y)$ are respectively as:

$$f_X(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} ; f_Y(y) = \begin{cases} \mu e^{-\mu y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

where λ, μ are the variances of variables X and Y respectively.

Then the two cumulative distribution functions: $F_X(x)$ and $F_Y(y)$ can be calculated as:

$$F_X(x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases} ; F_Y(y) = \begin{cases} 1 - e^{-\mu y} & y > 0 \\ 0 & y \leq 0 \end{cases}$$

And the two inverse marginal cumulative distribution functions of u and v are:

$$x = F_X^{-1}(u) = -\frac{1}{\lambda} \ln(1-u)$$

$$y = F_Y^{-1}(v) = -\frac{1}{\mu} \ln(1-v)$$

The joint exponential probability density function can be written as [Downton, 1970]:

$$f_{XY}(x, y) = \frac{\lambda\mu}{1-\rho} \exp\left(-\frac{\lambda x}{1-\rho} - \frac{\mu y}{1-\rho}\right) I_0\left(\frac{2}{1-\rho} \sqrt{\rho\lambda\mu xy}\right) \quad (2.22)$$

Chapter 2: Introduction to Copulas

where ρ is the associated Pearson linear correlation and I_0 is the first kind Modified Bessel function [Carrier et al., 1983] with order 0. So, the exponential copula function can be defined as:

$$C(u, v) = \int_0^{\frac{\log(1-u)}{\lambda}} \int_0^{\frac{\log(1-v)}{\mu}} \frac{\lambda\mu}{1-\rho} \exp\left(-\frac{\lambda x}{1-\rho} - \frac{\mu y}{1-\rho}\right) I_0\left(\frac{2}{1-\rho} \sqrt{\rho\lambda\mu xy}\right) dx dy \quad (2.23)$$

Similarly, considering two Rayleigh marginal probability density functions $f_X(x)$ and $f_Y(y)$ of two variables X and Y as:

$$f_X(x) = \frac{x}{\sigma_X^2} \exp\left(-\frac{x^2}{2\sigma_X^2}\right); \quad f_Y(y) = \frac{y}{\sigma_Y^2} \exp\left(-\frac{y^2}{2\sigma_Y^2}\right)$$

The joint Rayleigh probability density function can be written as: [Abu-Dayya & Beaulieu, 1994]

$$f_{XY}(x, y) = \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho^2)} \exp\left(-\frac{1}{(1-\rho^2)} \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2}\right)\right) I_0\left(\frac{xy\rho}{(1-\rho^2)\sigma_X\sigma_Y}\right) \quad (2.24)$$

The two marginal cumulative distribution functions are:

$$F_X(x) = 1 - \exp\left(-\frac{x^2}{2\sigma_X^2}\right); \quad F_Y(y) = 1 - \exp\left(-\frac{y^2}{2\sigma_Y^2}\right)$$

The two inverse marginal cumulative distributions can be calculated as:

$$x = F_X^{-1}(u) = \sqrt{-2\sigma_X^2 \ln(1-u)}$$

$$y = F_Y^{-1}(v) = \sqrt{-2\sigma_Y^2 \ln(1-v)}$$

The corresponding copula function is called Rayleigh copula and it can be defined as:

$$C(u, v) = \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho^2)} \exp\left(-\frac{1}{(1-\rho^2)} \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2}\right)\right) I_0\left(\frac{xy\rho}{(1-\rho^2)\sigma_X\sigma_Y}\right) dx dy \quad (2.25)$$

where $a_1 = \sqrt{-2\sigma_X^2 \ln(1-u)}$; $a_2 = \sqrt{-2\sigma_Y^2 \ln(1-v)}$,

Chapter 2: Introduction to Copulas

σ_X and σ_Y are standard deviations of variables X and Y respectively, and ρ is the associated linear correlation and I_0 is the first kind Modified Bessel function with order 0.

Considering Clayton, Frank and Gaussian are applied frequently, so these three copula functions will be computed for modeling and simulated by generation of random variables in the following sections. As for the two new copulas: Exponential and Rayleigh copula will be investigated in Appendix 1.

2.6 Copula parameter estimation

There are two approaches for estimation of copula parameter. The first method is parametric method which relies on the special relationships between the copula parameter and Kendall's tau or Spearman rank correlations [Nelsen, 1999]. The second method is non-parametric method which is called maximum likelihood method [Cherubini et al., 2004].

2.6.1 Kendall's tau and Spearman correlation based method

Kendall's tau correlation is a non-parametric statistic tool which is used to measure the association or statistical dependence between random variables.

Let $\{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\}$ denote a random sample of observations from a vector (X, Y) of continuous random variables, there are $\binom{n}{2}$ different pairs (x_i, y_i) and (x_j, y_j) of observations in the sample and each pair is either concordant or discordant.

If $x_i > x_j$ and $y_i > y_j$ or $x_i < x_j$ and $y_i < y_j$, then we say (x_i, y_i) and (x_j, y_j) are concordant pairs. Conversely, if $x_i > x_j$ and $y_i < y_j$ or $x_i < x_j$ and $y_i > y_j$, we say

Chapter 2: Introduction to Copulas

(x_i, y_i) and (x_j, y_j) are discordant pairs. More specifically, the Kendall's tau correlation for the sample can be defined as [Kruskal, 1958]:

$$\tau = \frac{c-d}{c+d} = \frac{c-d}{\binom{n}{2}} = \frac{(c-d)}{\frac{1}{2}n(n-1)} \quad (2.26)$$

where c and d are the number of concordant pairs and discordant pairs respectively for samples. The denominator in the definition of τ can be interpreted as the total number of concordant and discordant pairs. So, a high value in the numerator means that most pairs are concordant. The arrangement of Kendall's tau correlation is between $[-1, 1]$. If the ranking correlation of two random variables is in perfect agreement then the Kendall's tau correlation is 1. If the disagreement between the two rankings is perfect, the coefficient has the value -1 . If the rankings are completely independent, the coefficient has value 0. Another version of Kendall's tau correlation is defined as the probability of concordance minus the probability of discordance and can be written as [Schweizer & Wolff, 1981], [Nelsen, 1999]:

$$\tau = \Pr[(X_1 - X_2)(Y_1 - Y_2) > 0] - \Pr[(X_1 - X_2)(Y_1 - Y_2) < 0] \quad (2.27)$$

where (X_1, Y_1) and (X_2, Y_2) are independent vectors of continuous random variables with same joint distribution function and common marginal distributions. Kendall's tau correlation can also be expressed simply in term of the copula functions as [Schweizer & Wolff, 1981], [Nelsen, 1999]:

$$\tau = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1 \quad (2.28)$$

This formula shows the relationship between the Kendall's tau correlation and the copula parameter, some typical copula parameter estimations by Kendall's tau correlation have been listed in the Table 2.1 [Huard et al., 2006].

Another important rank correlation is called Spearman correlation which is also a non-parametric rank correlation used to measure the strength of the association between two variables. Given n raw samples x_i and y_i and convert x_i and y_i to the

Chapter 2: Introduction to Copulas

rank r_{xi} and r_{yi} . The difference between the ranks is $d_i = r_{xi} - r_{yi}$. The Spearman rank correlation coefficient can be defined as [Maritz, 1981]:

$$\rho_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (2.29)$$

If there is a perfect agreement between two sets of ranks, then $\rho_s = 1$; If there is a complete disagreement between two sets of ranks, then, $\rho_s = -1$; If the variables are independent, then, $\rho_s = 0$. Spearman correlation also can be expressed in term of the copula distribution function as [Nelsen, 1999]:

$$\rho_s = 12 \iint_{[0,1]^2} uv dC(u,v) - 3 \quad (2.30)$$

The corresponding formulas between copula parameter and Spearman rank correlation are also been listed in the Table 2.1 [Huard et al., 2006].

2.6.2 Maximum Likelihood Estimation (MLE) method

Maximum likelihood estimation is a popular statistical method for parameter estimation. Given a fixed dataset, MLE chooses the parameter value which makes the dataset more likely to happen than other parameter values [Stuart & Ord, 1987].

Let $X = \{(X_1^t, X_2^t)\}_{t=1}^T$ denotes a sample where X_1^t and X_2^t are two vectors with the length T , the expression of log-likelihood of joint probability density function can be written in terms of the copula density and marginal probability density functions as [Cherubini et al., 2004]:

$$\ell(\alpha) = \sum_{t=1}^T \ln c(F_1(x_1^t, \theta_1), F_2(x_2^t, \theta_2)) + \sum_{t=1}^T \sum_{n=1}^2 \ln f_n(x_n^t; \theta_n) \quad (2.31)$$

Copula	Kendall's τ and Range	Spearman ρ_s and Range
Clayton	$\tau = \frac{\theta}{\theta + 2}$ $\tau \in (0,1)$	<p>No Simple Form</p> $0 < \rho_s < 1$
Frank	$\tau = 1 - \frac{4}{\theta} [1 - D_1(\theta)]$ $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$ $\tau \in [-1,1]$	$\rho_s = 1 - \frac{12}{\theta} [D_1(\theta) - D_2(\theta)]$ $D_k(x) = \frac{k}{x^k} \int_0^x \frac{t^k}{e^t - 1} dt$ $-1 \leq \rho_s \leq 1$
Gaussian	$\tau = \frac{2}{\pi} \arcsin \theta$ $\tau \in [-1,1]$	$\rho_s = \frac{6}{\pi} \arcsin\left(\frac{\theta}{2}\right)$ $-1 \leq \rho_s \leq 1$

Table 2.1: Copula parameter expressed in terms of the Kendall's tau and Spearman rank correlation [Huard et al., 2006].

where c is the copula density function, F_1 and F_2 are two marginal distributions with parameters θ_1 and θ_2 respectively and function f_n are marginal probability density functions. There are two popular MLE methods for this parameter estimations, the first is called Inference for Margins (IFM) which estimates the copula parameter by two steps [Cherubini et al., 2004].

Step 1: Estimate the parameter θ of marginal firstly.

$$\theta'_n = \arg \max \sum_{t=1}^T \sum_{n=1}^2 \ln f_n(x'_n; \theta_n) \quad (2.32)$$

Chapter 2: Introduction to Copulas

Step 2: Estimate the copula parameter α by using marginal parameter θ which has been found above.

$$\alpha' = \arg \max \ell(\alpha) = \arg \max \sum_{t=1}^T \ln c(F_1(x_1^t; \theta_1), F_2(x_n^t; \theta_2), \alpha)$$

The second MLE method is called Canonical Maximum Likelihood method (CML) [Cherubini et al., 2004], [Durrleman et al., 2000]. Here, the data x_1^t and x_2^t are first transformed into the uniform variants u_1^t and u_2^t by using empirical distributions, then use the following MLE formula to estimate the parameters of the copula.

$$\alpha' = \arg \max \sum_{t=1}^2 \ln c(u_1, u_2; \alpha)$$

The empirical distribution has been defined as [Shorack & Wellner, 1986]:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(X_i \leq x) \quad (2.33)$$

where function $I_A(x)$ is indicator function given by:

$$I_A(x) = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{if } x \notin A \end{cases}$$

2.6.3 Comparing the measures of copula parameter estimation

The first method of copula parameter estimation which is based on Kendall's tau and Spearman correlation are the methods of precise formula derivation. However, the significant disadvantage is the time consuming computing effort which increases with the increasing size of data. Specifically, the direct computation of Kendall's tau correlation between n pairs of samples is $O(n^2)$ in complexity and becomes very slow on large dataset. Fast algorithm with expected calculating complexity of $O(n \log n)$ have been proposed in [Dwork et al., 2001], [Christensen, 2005]. Besides, not all copulas have the simple expression linking the copula parameter and Kendall's tau or

Chapter 2: Introduction to Copulas

Spearman correlation, some of them may be very complicated and someone cannot be found.

The MLE method is an alternative method which is feasible for data with a large size, and deals with the case of high dimension data. Furthermore, CML method considers the attractive advantage of copula that copula separates the dependence structure and marginal distributions.

Consequently, the measure based on the Kendall's tau and Spearman correlation is used when the size of the dataset is small and a simple expression can be found between copula parameter and Kendall's tau or Spearman correlation. The MLE measure is suitable for the data with large size or high dimensional data.

2.7 Copula distribution function computation

In this part, Clayton, Frank and Gaussian copula distribution functions will be computed for modelling copula distributions.

2.7.1 Clayton, Frank and Gaussian Copula

After the estimation of copula parameters, the copula can be modeled by computing copula distribution function with uniform marginal distributions. The expression of Clayton and Frank copulas are quite simple, and are easy to calculate. As for the Gaussian copula, it is not feasible to process the double integrals directly. The following transforming which reduces the double integrals to one integral can be used to calculate the Gaussian copula distribution function [Genz, 2004]:

$$\begin{aligned} & \int_{-\infty}^h \int_{-\infty}^k \frac{1}{2\pi(1-\rho)^2} \exp\left(-\frac{x^2 - 2\rho xy + y^2}{2(1-\rho^2)}\right) dy dx \\ &= \varphi(-h)\varphi(-k) + \frac{1}{2\pi} \int_0^{\sin^{-1}\rho} \exp\left(-\frac{h^2 + k^2 - 2hk \sin \theta}{2(\cos \theta)^2}\right) d\theta \end{aligned} \quad (2.34)$$

Chapter 2: Introduction to Copulas

where φ and ρ are the standard Gaussian cumulative distribution function and Pearson correlation of variables x and y respectively. The modelling results of Clayton copula with parameter = 2, Frank copula with parameter =0.5 and Gaussian copula with parameter = 0.5 are shown in Figure 2.3, 2.4 and 2.5 respectively. It may be seen that these three copulas always have the similar pyramid shapes but different values for the same variables u and v which were demonstrated clearly in the contour plots.

2.7.2 Calculation joint distribution by using copulas

Once the copula cumulative distribution function is estimated, the joint cumulative distribution can be calculated by these copula distribution functions and the inverse marginal distribution functions determined. For example, given two vector samples with size 2000 respectively, the first vector is a Gaussian distributed data with the parameters: mean value $\mu= 2.0441$ and standard deviation $\sigma= 2.0220$. The Gaussian probability density function can be written as:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.35)$$

The second vector is a Student t distributed data with the parameter freedom degree $\nu = 5.3654$ and the Student t probability density function can be written as:

$$f(x) = \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \frac{1}{\sqrt{\nu\pi}} \frac{1}{(1 + \frac{x^2}{\nu})^{\frac{\nu+1}{2}}} \quad (2.36)$$

Here, Γ is called Gamma function which has been defined in Eq. (2.12). These two marginal probability density functions are shown in Figure 2.6.

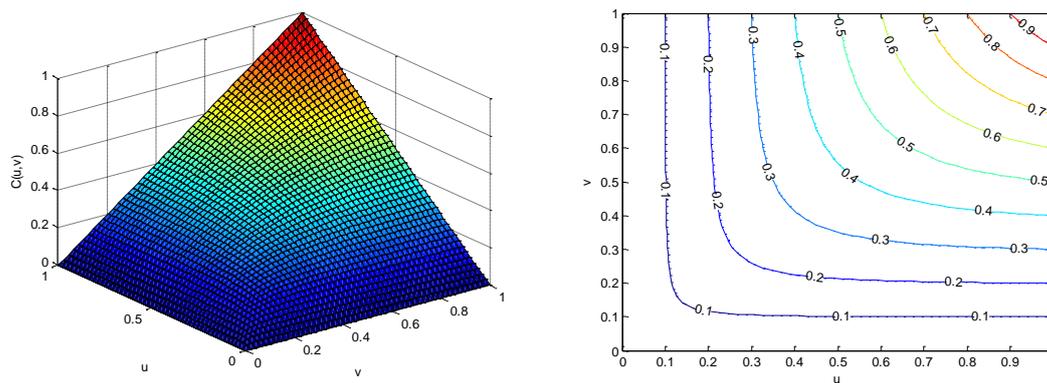


Figure 2.3: Clayton copula with parameter = 2

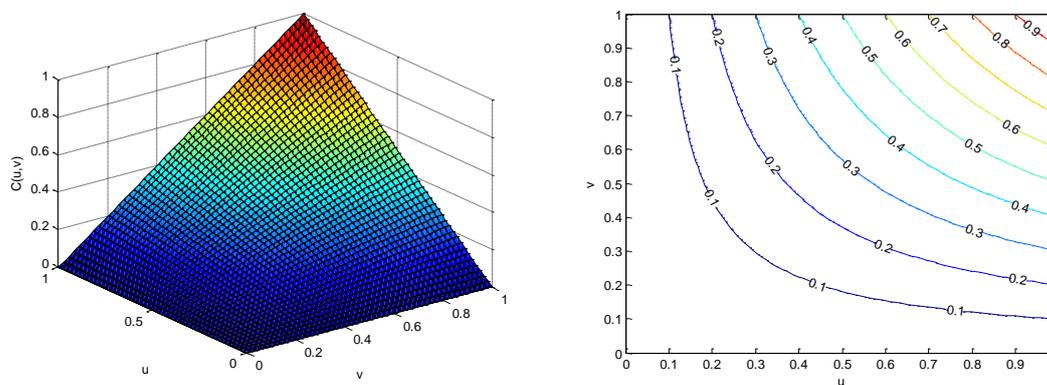


Figure 2.4: Frank copula with parameter = 0.5

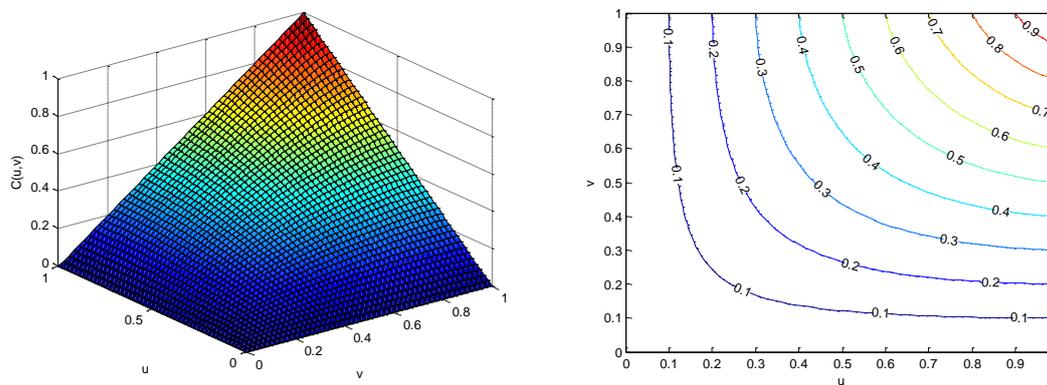


Figure 2.5: Gaussian copula with parameter = 0.5

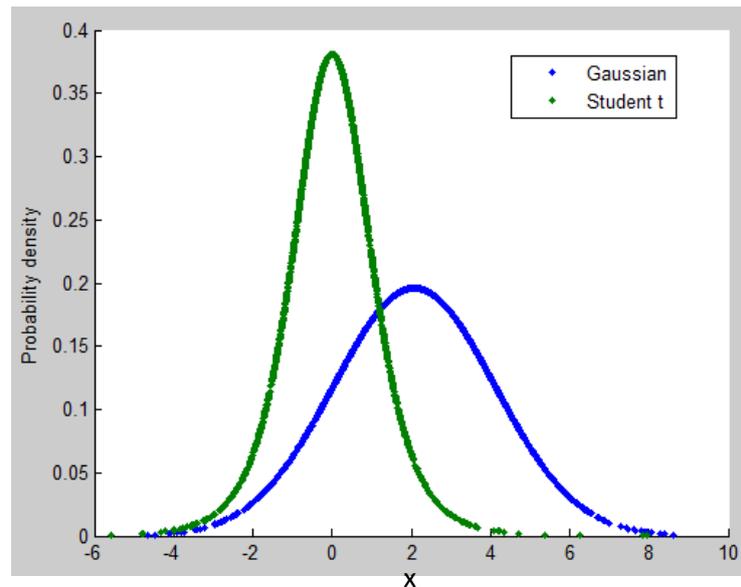


Figure 2.6: Gaussian and Student t marginal probability density functions

To calculate the joint cumulative distribution, the first step is to estimate the copula parameter. Here CML method which has been introduced in Eq. (2.21) is adopted to estimate the Gaussian copula parameter and the result is:
$$\begin{bmatrix} 1 & 0.6992 \\ 0.6992 & 1 \end{bmatrix}.$$

After the estimation of Gaussian copula parameter, the next step is to calculate the corresponding Gaussian copula cumulative distribution function by the method which has been described in Eq. (2.23). The result of Gaussian copula cumulative distribution and the corresponding contour plot are given in Figure 2.7 and 2.8 respectively.

Finally, compute the inverse Gaussian distribution and inverse Student t distribution for the two marginals respectively to obtain the joint cumulative distribution function. The figure of the joint cumulative distribution function and its corresponding contour figure are given in Figure 2.9 and Figure 2.10 respectively and x_1 and x_2 represents the Gaussian and Student t distributed data respectively.

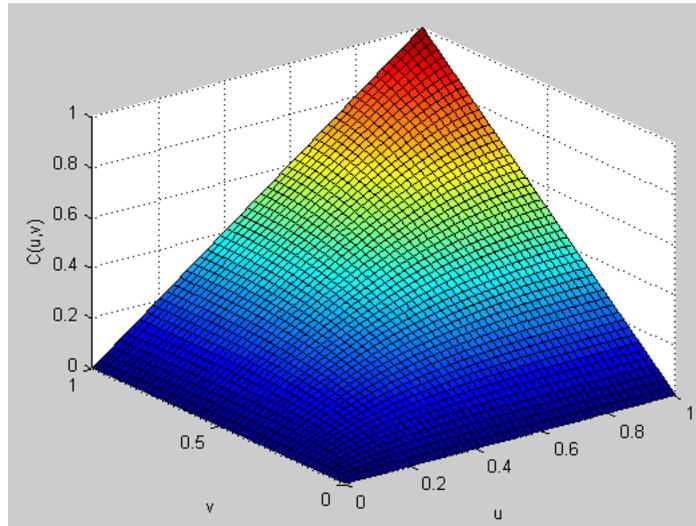


Figure 2.7: Gaussian copula distribution

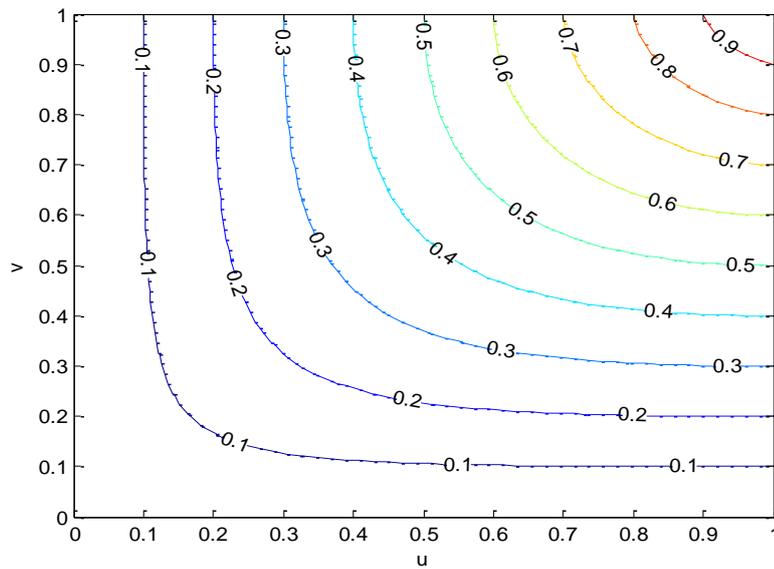


Figure 2.8: Contour of Gaussian copula distribution

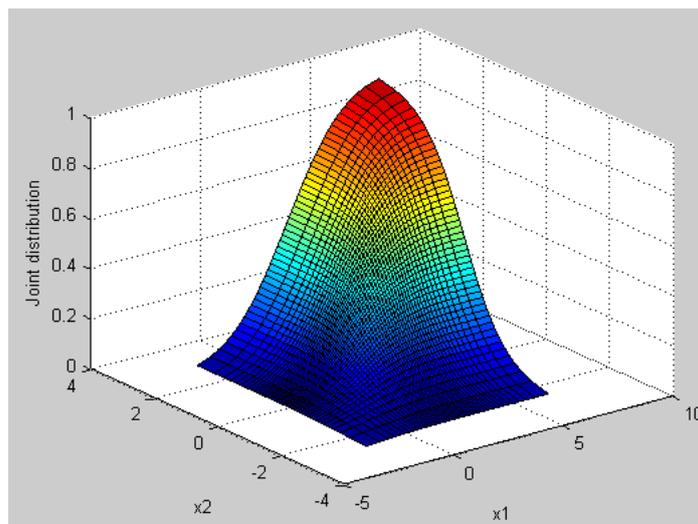


Figure 2.9: The joint cumulative distribution

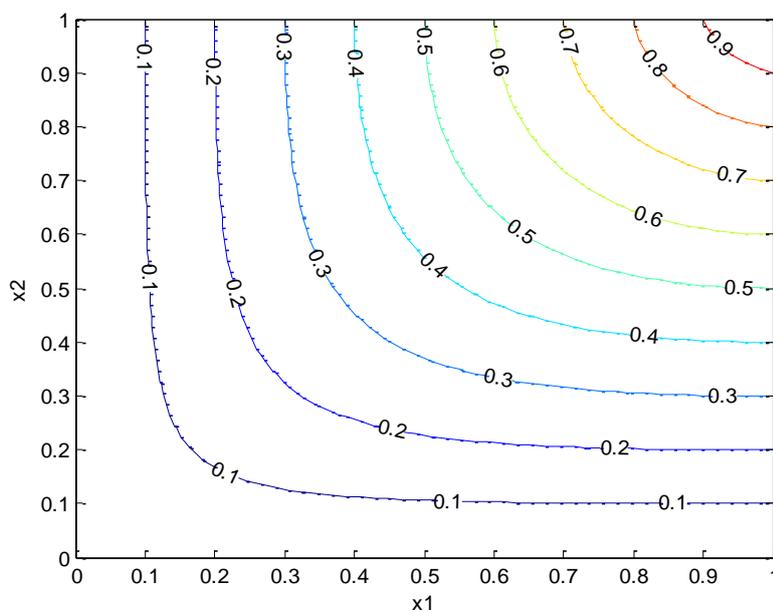


Figure 2.10: Contour plot of joint cumulative distribution

2.8 Random variables generation for copulas

In this section, the generation of random variables by using conditional copulas will be investigated. This is also called copula simulation. The process can be described as: considering a copula function $C(u, v)$, if the first variable u and the copula parameter

Chapter 2: Introduction to Copulas

which indicates the dependence between random variables are known, then how can the second copula variable v be estimated?

Similar to the definition of joint conditional probability, the bivariate copula case can be expressed by [Trivedi & Zimmer, 2007]

$$C_u(v) = \Pr[V \leq v | U = u] = \lim_{\Delta u \rightarrow 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} = \frac{\partial C(u, v)}{\partial(u)} \quad (2.37)$$

For example, given a joint cumulative distribution function $H(x, y)$ as:

$$H(x, y) = \begin{cases} \frac{(x+1)(e^y - 1)}{x + 2e^y - 1} & x \in [-1, 1]; y \in [0, \infty] \\ 1 - e^{-y} & x \in (1, \infty]; y \in [0, \infty] \\ 0 & \text{elsewhere} \end{cases}$$

with two marginal cumulative distribution functions: $F(x)$ and $G(y)$ as:

$$F(x) = \begin{cases} 0 & x < -1 \\ \frac{x+1}{2} & x \in [-1, 1] \\ 1 & x > 1 \end{cases} \quad ; \quad G(y) = \begin{cases} 0 & y < 0 \\ 1 - e^{-y} & y \geq 0 \end{cases}$$

Then, $F^{-1}(u) = 2u - 1$ and $G^{-1}(v) = -\ln(1 - v)$ for $0 \leq u \leq 1$ and $0 \leq v \leq 1$, then, substitute the two inverse marginal cumulative distributions into $H(x, y)$, let $x = F^{-1}(u) = 2u - 1$, and $y = G^{-1}(v) = -\ln(1 - v)$, then the copula function can be expressed as:

$$C(u, v) = \frac{uv}{u + v - uv} \quad (2.38)$$

The corresponding conditional copula can be calculated by using the partial differential operation described in Eq. (2.37) and the following result is obtained:

$$C_u(v) = \frac{\partial C(u, v)}{\partial(u)} = \left(\frac{v}{u + v - uv} \right)^2 \quad (2.39)$$

The inverse conditional copula is also can be obtained as:

$$C_u^{-1}(t) = \frac{u\sqrt{t}}{1 - (1-u)\sqrt{t}} \quad (2.40)$$

The method of generation of random variables (u, v) for copula function $C(u, v)$ can be described as following steps:

Firstly, generate two independent uniform variables u and t with the range in $[0, 1]$.

Secondly, let

$$v = \frac{u\sqrt{t}}{1 - (1-u)\sqrt{t}} \quad (2.41)$$

Furthermore, the corresponding random variables (x, y) are also can be achieved by setting:

$$\begin{cases} x = 2u - 1 \\ y = -\ln(1 - v) \end{cases}$$

Similar to the example above, the formulas for generation of the random variables (u, v) for Clayton and Frank copula function $C(u, v)$ can be found as:

For Clayton copula:

$$v = (u^{-\theta} (t^{-\frac{\theta}{\theta+1}} - 1) + 1)^{-\frac{1}{\theta}} \quad (2.42)$$

For Frank Copula:

$$v = -\frac{1}{\theta} \ln \left(1 - \frac{t(1 - e^{-\theta})}{t(1 - e^{-\theta u}) + e^{-\theta u}} \right) \quad (2.43)$$

As for Gaussian copula simulation, firstly, we can review the generation of random variables $[X, Y]$ for joint bivariate Gaussian distribution.

Chapter 2: Introduction to Copulas

Similar to the simulation of joint standard Gaussian distribution which is shown in Appendix 2, the Gaussian copula simulation can be achieved using the following steps:

Firstly, using Cholesky decomposition [David & Lloyd, 1997] to produce an upper triangular A so that $A^T A = R$ where R is the strictly positive defined covariance matrix.

This can be achieved by: $\begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$, where ρ is the parameter of Gaussian Copula,

then the lower triangular can be calculated as:

$$A = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$$

The next step is to generate two independent standard normal random variables Z_1 and

Z_2 , and $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$; Let $X = A * Z = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix}$

The last step is to calculate the cumulative distribution function by using standard Gaussian distribution for X .

$$u = \Phi(X_1) \text{ and } v = \Phi(X_2)$$

where Φ is univariate standard Gaussian distribution function.

The simulation of Clayton, Frank and Gaussian copula function with different parameters from low to high dependence are given in Figure 2.11, 2.12 and 2.13 respectively. It may be found the variables of copulas concentrates around the diagonal when they have high dependence and disperse when they have low dependence.

2.9 Copula density function

The joint probability density function (pdf) can be derived by the second order partial differential operation on the joint cumulative distribution function as:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y}$$

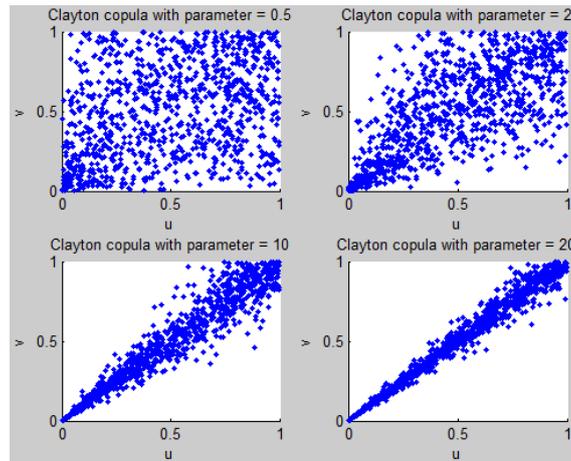


Figure 2.11: Simulation of Clayton copula

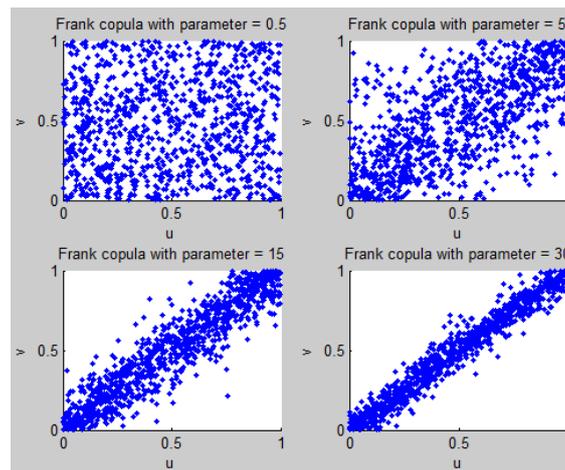


Figure 2.12: Simulation of Frank copula

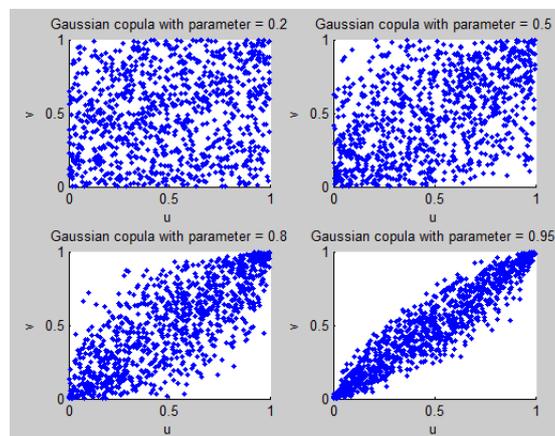


Figure 2.13: Simulation of Gaussian copula

Chapter 2: Introduction to Copulas

where $f_{XY}(x, y)$ and $F_{XY}(x, y)$ are joint probability density function and joint cumulative distribution function of random variables x and y respectively. Note from the Eq. (2.1), the following equation can be obtained as [Durrani & Zeng, 2007]:

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} = \frac{\partial^2 C(F_X(x), F_Y(y))}{\partial x \partial y}$$

Let $c(u, v) = \frac{\partial^2 C(F_X(x), F_Y(y))}{\partial x \partial y}$ be the copula density function, then

$$f_{XY}(x, y) = \frac{\partial^2 F_{XY}(x, y)}{\partial x \partial y} = \frac{\partial^2 C[F_X(x), F_Y(y)]}{\partial x \partial y} = \frac{\partial^2 C(u, v)}{\partial u \partial v} \frac{\partial F_X(x)}{\partial x} \frac{\partial F_Y(y)}{\partial y} = c(u, v) f_X(x) f_Y(y)$$

(2.44)

So, the copula density function can be expressed as:

$$c(u, v) = \frac{f_{XY}(x, y)}{f_X(x) f_Y(y)}$$

(2.45)

and $f_{XY}(x, y) = c(u, v) f_X(x) f_Y(y)$, the copula density function can be considered as the dependency structure part of the two marginal distributions and the product of the two marginal probability density functions can be regarded as the independent part of the marginal distributions. The copula density functions for Clayton, Frank and Gaussian copulas are listed in Table 2.2 [Huard et al., 2006].

The next section gives an approach for estimation of the copula probability density and joint probability density functions. The same dataset which was used in Section 2.72 and Gaussian copula is applied. Firstly, estimate the Gaussian copula parameter, and then calculate the joint probability density function. The result is shown in Figure 2.14. According to the Eq. (2.35), the joint probability density function can be calculated by the product of copula density function and the two marginal probability density functions. The plot of joint probability density function is shown in Figure 2.15.

Copula	Copula density function $c(u, v)$
Clayton	$(1 + \theta)u^{-1-\theta}(-1 + u^{-\theta} + v^{-\theta})^{-2-1/\theta}$
Frank	$\frac{\theta e^{\theta(1+u+v)}(-1+e^\theta)}{[e^\theta - e^{\theta(1+u)} - e^{\theta(1+v)} + e^{\theta(u+v)}]^2}$
Gaussian	$\frac{1}{\sqrt{1-\rho^2}} e^{-\frac{[\varphi^{-1}(u)]^2 + [\varphi^{-1}(v)]^2 - 2\rho\varphi^{-1}(u)\varphi^{-1}(v)}{2(1-\rho^2)}}$

Table 2.2: Clayton, Frank and Gaussian Copula density

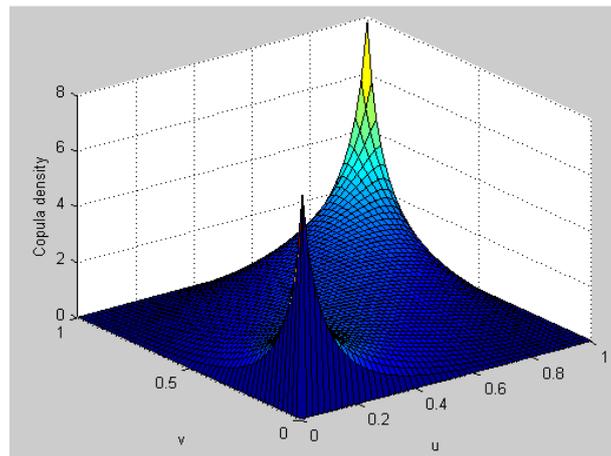


Figure 2.14: Gaussian copula density function

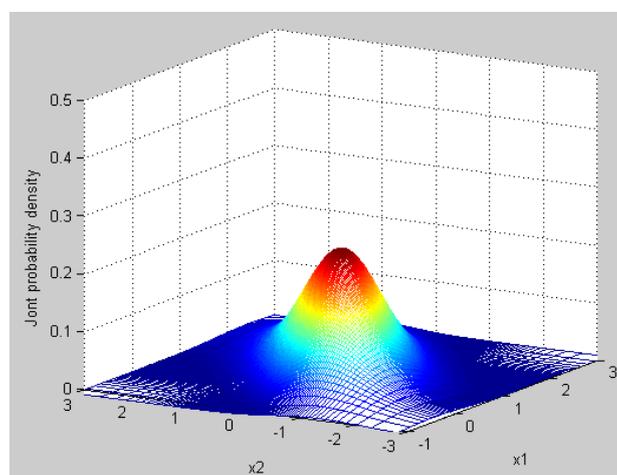


Figure 2.15: Joint probability density function

Another experiment was carried out to validate the effectiveness of copula density function in representing the joint probability density function. Firstly, generate a two-dimensional dataset which has the bivariate Gaussian distribution with the parameters: mean values μ and covariance cov as:

$$\mu = \begin{bmatrix} \mu_X \\ \mu_Y \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix}$$

$$cov = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix} = \begin{bmatrix} 0.9 & 0.4 \\ 0.4 & 0.9 \end{bmatrix}$$

Step 1: use the joint Gaussian assumption probability density function in Eq. (2.9) to calculate the joint Gaussian probability density function. The joint density function and its contour plot results are shown in Figure 2.16 and 2.17 respectively.

Step 2: Estimate the Gaussian copula parameter for this two-dimensional dataset and the result is:

$$\rho_{Gau} = \begin{bmatrix} 1 & 0.7743 \\ 0.7743 & 1 \end{bmatrix}$$

Step 3: Estimate the copula density function and then the joint probability density function can be calculated as the product of copula density function and two Gaussian marginal probability density functions. The joint density function and its contour plots of copulas based joint probability density function are given in Figure 2.18. and 2.19 respectively.

Step 4: In order to obtain better visual effects, we overlap the two contour plots to observe the difference between the joint Gaussian probability density function and copulas based joint probability density function.

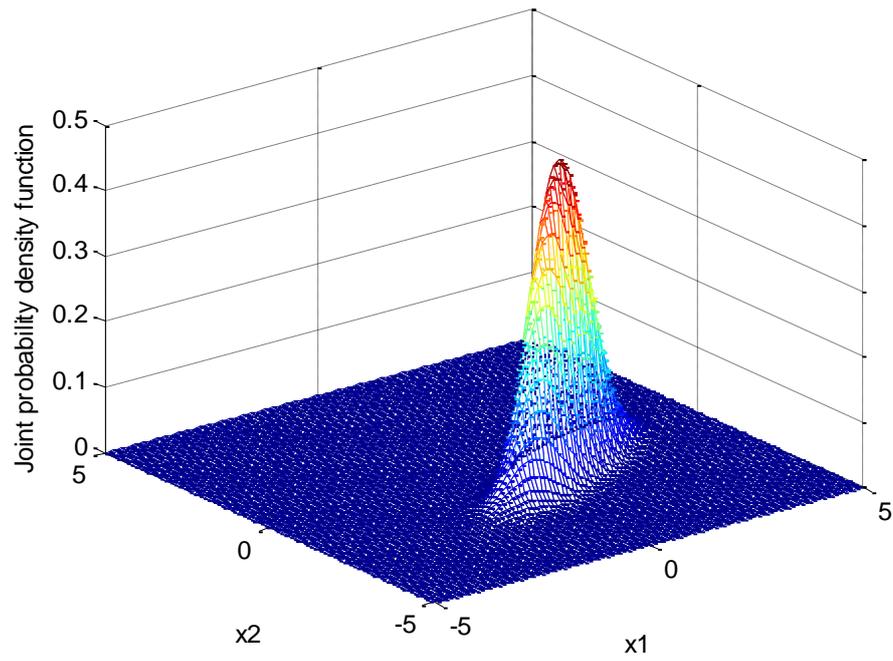


Figure 2.16: Joint probability density function of bivariate Gaussian distribution

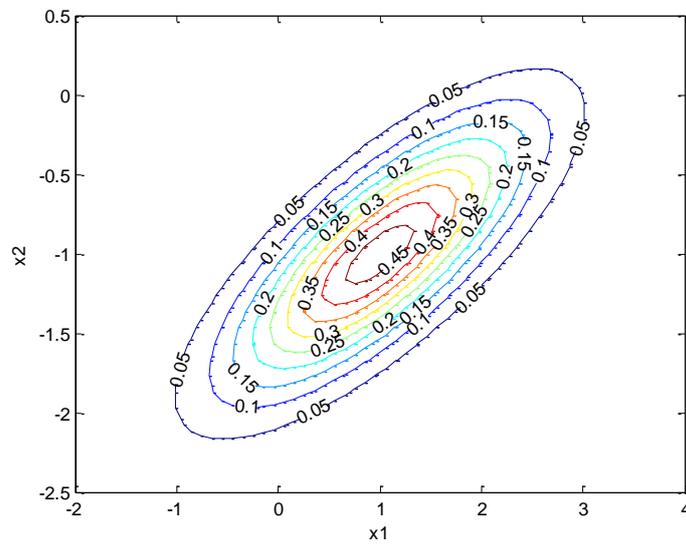


Figure 2.17: Contour of joint probability density function of bivariate Gaussian distribution

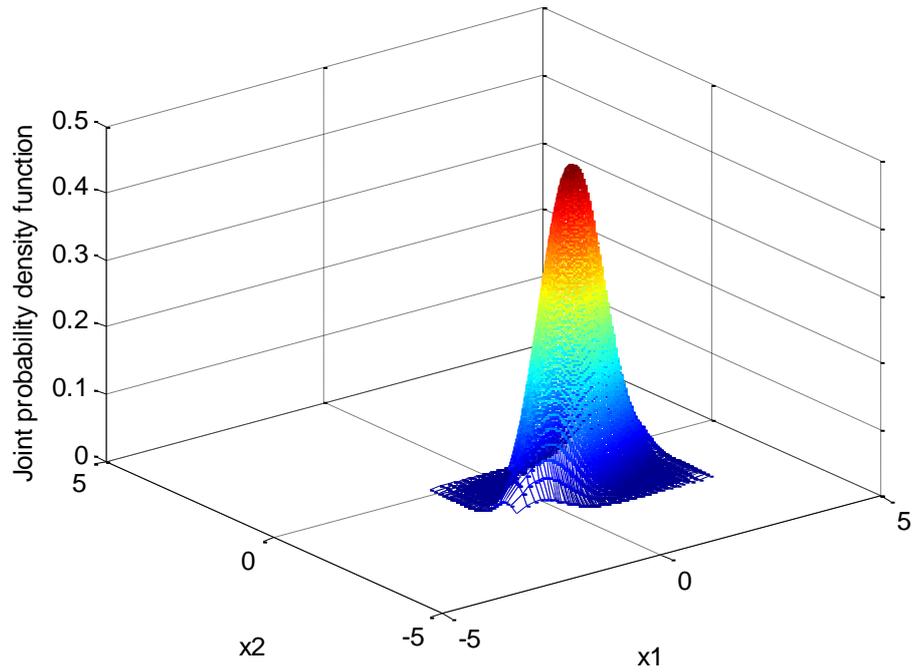


Figure 2.18: Joint probability density function using Gaussian copula

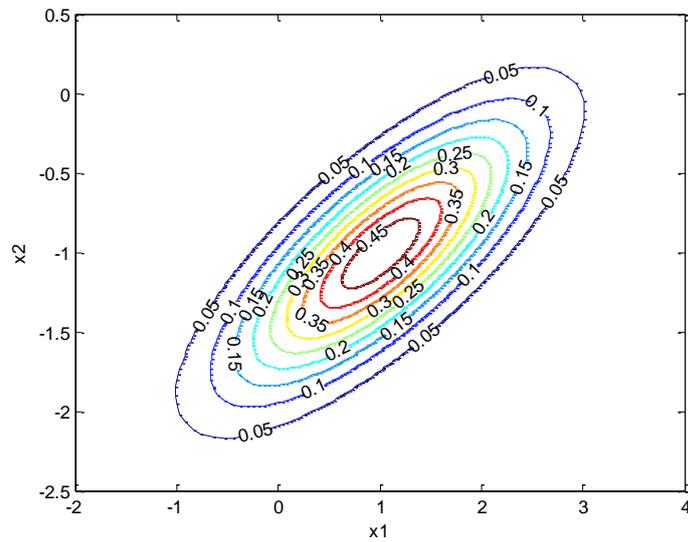


Figure 2.19: Contour of joint probability density function using Gaussian copula

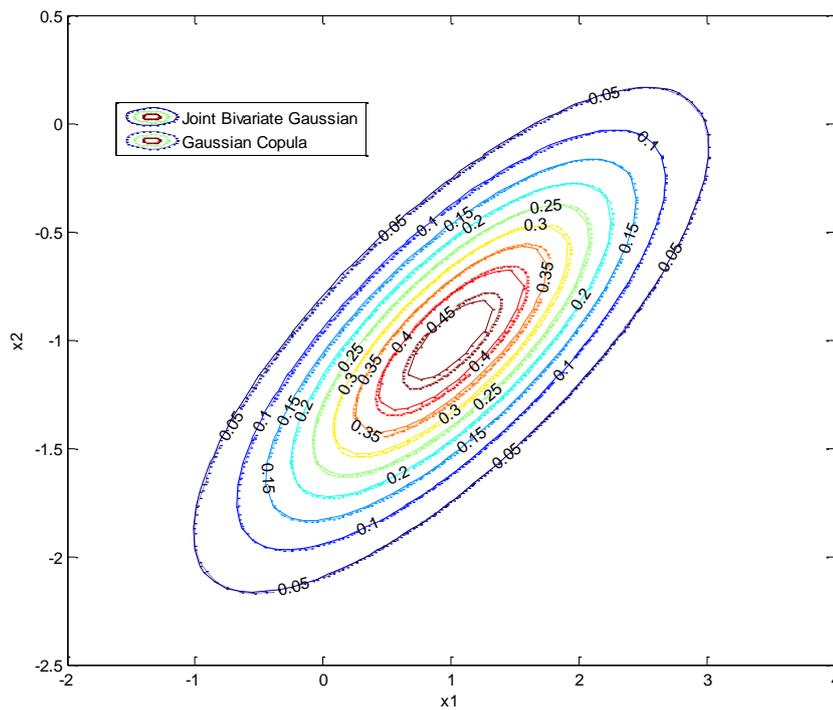


Figure 2.20: Comparison of joint probability density function obtained by bivariate Gaussian distribution and Gaussian copula respectively

As can be found in Figure 2.20, the difference between Bivariate Gaussian based joint probability density shown in solid lines and Gaussian copula based joint probability density function shown in dashed lines are very similar, and even almost same in the areas with low joint density values. It shows copula method is competent for estimation of joint probability density function between two random variables.

2.10 Copula selection

As there are several copula functions that have been defined, and subtle differences exist among these copulas. Hence, it is necessary to choose the ‘best’ copula which is optimal to describe the dependence between variables.

Chapter 2: Introduction to Copulas

Firstly, it should be noted that some copulas have a narrow range to describe the dependency between random variables. For example, Farlie-Gumbel-Morgenstern (FGM) copula is defined as [Nelsen, 1999]:

$$C(u, v) = uv + \theta uv(1-u)(1-v) \quad (2.46)$$

where $-1 \leq \theta \leq 1$, with the corresponding copula density function is:

$$c(u, v) = 1 + \theta(1-2u)(1-2v) \quad (2.47)$$

For FGM copula, there are simple expressions for both copula distribution and copula density function. However, note the corresponding formula between copula parameter and Kendall's tau correlation [Nelsen, 1999]:

$$\tau = 4 \iint_{[0,1]^2} C(u, v) dC(u, v) - 1 = 4\left(\frac{1}{4} + \frac{\theta}{18}\right) - 1 = \frac{2}{9}\theta \quad (2.48)$$

Hence, the Kendall's tau correlation range is limited to the narrow bound as:

$$-\frac{2}{9} \leq \tau \leq \frac{2}{9}$$

Due to the narrow range of correlation, some copulas are not robust to process the different dataset.

Secondly, considering some copulas have very complicated expressions and are not convenient for applications. However, among the known copulas, seems Clayton, Frank and Gaussian copula are reasonable choices.

More specifically, Empirical copula based method for the selection of the optimal copula has been proposed [Durrleman et al., 2000].

Given a two vector data sample a and b with the vector length d , Euclidean distance: can be calculated as:

$$\overline{d}_2(a, b) = \sqrt{\sum_{i=1}^d (a_i - b_i)^2} \quad (2.49)$$

where $a = [a_1, a_2, \dots, a_d]$ and $b = [b_1, b_2, \dots, b_d]$.

Chapter 2: Introduction to Copulas

Similar to the definition of the empirical distribution in Eq. (2.33), the empirical copula distribution can be defined as follows:

Let $\{(x_k, y_k)\}_{k=1}^n$ denotes a sample of size n from a continuous bivariate distribution, and then the empirical copula C_n can be defined by [Nelsen, 1999]:

$$C_n\left(\frac{i}{n}, \frac{j}{n}\right) = \frac{\text{(number of pairs}(x, y) \text{ in the sample with } x \leq x_{(i)}, y \leq y_{(j)})}{n} \quad (2.50)$$

where $x \leq x_{(i)}, y \leq y_{(j)}, 1 \leq i, j \leq n$ denotes order statistics from the sample.

Considering a finite subset of copulas, there should be a Minimal Euclidean distance between the optimal copula and the empirical copula. This distance can be written as:

$$\overline{d}_2(C_e, C) = \|C_e - C\|_{L^2} = \left\{ \sum_{t_1=1}^T \cdots \sum_{t_n=1}^T \cdots \sum_{t_N=1}^T [C_e\left(\frac{t_1}{T}, \dots, \frac{t_n}{T}, \dots, \frac{t_N}{T}\right) - C\left(\frac{t_1}{T}, \dots, \frac{t_n}{T}, \dots, \frac{t_N}{T}\right)]^2 \right\}^{\frac{1}{2}}$$

For the bivariate case, the distance becomes:

$$\overline{d}_2(C_e, C) = \|C_e - C\|_{L^2} = \left\{ \sum_{t_1=1}^T \sum_{t_2=1}^T C_e\left(\frac{t_1}{T}, \frac{t_2}{T}\right) - C\left(\frac{t_1}{T}, \frac{t_2}{T}\right) \right\}^{\frac{1}{2}} \quad (2.51)$$

where C_e is empirical copula and C is the candidate copula. It means that both the cumulative distribution function (CDF) of candidate copula and the empirical copula are required to be calculated respectively.

2.11 Conclusion

In this chapter, the copula function theory has been introduced which include copula function definition and its properties, copula density function, conditional copula, estimation of copula parameters, modelling of copula distribution function and

Chapter 2: Introduction to Copulas

generation of random variables for typical copulas. It also analyzed the benefits of copula function, and compared with the conventional dependency tools such as Pearson correlation which is only effective for the elliptical distributions such as multivariate Gaussian, Student t distributions that require the marginal distributions are Gaussian or Student t distributions respectively and must be consistent. Copulas extend the concept of dependence which may not be linear and are able to deal with arbitrary marginal distributions. Furthermore, the estimations of the joint cumulative distribution function and density function for arbitrary marginal distributions using copulas are proposed. Finally, dependency range of copula and empirical copula were explored for the selection of optimal copula.

Chapter 3

Divergence based Information using Copulas

Summary

In this chapter, four classes of generalized divergences: Csiszar, Renyi-like, modified Bregman and Burbea-Rao divergence based information which are generally defined in terms of joint and marginal density functions are explored and frameworks are developed to express these divergence in terms of copula density functions only. Algorithms for the calculation of such divergence based information using copula density functions are proposed. Algorithm of computation of these divergence based information by using Gaussian copula density is proposed. Copula based method is compared with the Gaussian assumption and joint histogram based method, and the modified Bregman divergence is validated using smallest enclosing curve and K-means classification.

3.1 Introduction

Based on the works of Csiszar [Csiszar, 1967], Ali and Silvery [Ali & Silvey, 1966], the definition of divergence can be summarized as:

Consider $p=p(x)$ and $q=q(x)$ are two probability density functions of two probability distribution P and Q respectively in the same space X , where $p(x) \geq 0$ and $q(x) \geq 0$ for each $x \in X$ and $\sum_{x \in X} p(x) = 1$; $\sum_{x \in X} q(x) = 1$.

In the discrete case, the divergence can be defined as:

$$D(P, Q) = \sum_{x \in X} \phi(p, q)$$

In the continue case, it can be written as:

$$D(P, Q) = \int \phi(p, q) dx$$

where the function $\phi(p, q)$ is called kernel function. For instance, if the kernel function $\phi(p, q) = p \log(\frac{p}{q})$, then the corresponding divergence is Kullback-Leibler divergence. Obviously, divergence is determinate by kernel functions, and the divergences vary with different kernel functions. Moreover, divergence D must satisfy the following conditions: First, divergence D must be defined for all the samples. Secondly, with the moving apart of P and Q , the divergence D should increase, and it reaches the maximal value when P is orthogonal to Q . It has the only minimal value 0 when $P = Q$. Finally, considering another distribution S , if S is closer to Q than P , then, $D(P, S) \geq D(Q, S)$ [Osterreicher, 2002].

The divergence based information can be defined as: The divergence between the joint probability density function and the product of the marginal density functions. For example, mutual information is Kullback-Leibler based information and Kullback-Leibler divergence is the special case of Csiszar divergence. It has been recognised that generalized divergence based information can provide the ability to control the measurement sensitivity and hence better accuracy and is more efficient than the classic Kullback-Leibler divergence based information. [He, et al., 2003], [Pluim, et al., 2004]. So, the generalized divergences based information are worth exploring.

In this chapter, four classes of generalized divergences [Pardo & Vajda, 2003] & [Martin, 2006]: Csiszar, Renyi-like, Bregman and Burbea-Rao divergence are explored. These four classes of divergence have been summarized in Table 3.1 and the function $f()$ is convex function and called kernel function. For example, if the kernel function $f(t) = t \log(t)$ for the Csiszar divergence, then Csiszar divergence becomes:

$$D_{CS}(p, q) = qf\left(\frac{p}{q}\right) = q\left(\frac{p}{q}\right)\log\left(\frac{p}{q}\right) = p\log\left(\frac{p}{q}\right)$$

Let $p = p_{XY}(x, y)$ be the joint probability density function, $q = q_X(x)q_Y(y)$ be the

Csiszar Divergence
$D_{CS}(p, q) = qf\left(\frac{p}{q}\right)$
Renyi-like Divergence
$D_{Renyi}(p, q) = \log\left[f^{-1}\left(pf\left(\frac{p}{q}\right)\right)\right]$
Bregman Divergence
$D_{Bre}(p, q) = f(p) - f(q) - (p - q)f'(q)$
Burbea-Rao Divergence
$D_{BR}(p, q) = \frac{f(p) + f(q)}{2} - f\left(\frac{p + q}{2}\right)$
$p=p(x)$ and $q=q(x)$ are probability density functions

Table 3.1: Definition of Csiszar, Renyi-like, Bregman and Burbea-Rao Divergence

product of two marginal probability density function: $q_X(x)$ and $q_Y(y)$ of random variables X and Y respectively, then this Csiszar divergence based information becomes the familiar mutual information:

$$D(x, y) = \iint_{XY} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{q_X(x)q_Y(y)} dx dy$$

3.2 Problem description

The key to divergence based information method is the calculation of joint probability density functions. However, in the non-elliptical distribution case, it is difficult to

estimate the joint probability density function precisely. The measures of computing the joint probability density function in image processing can be classified into two categories: Non-parametric and parametric methods. The typical Non-parametric methods compute the normalization of the joint histograms of pixel intensity of the overlapping parts of two images. This is called joint histogram method which is a simple approach to estimate the joint probability density function [Maes et al., 1997]. This method usually requires a large amount of data for reliable results. However, some data within a small size such as that of neighbours for certain pixel may be required to process. Moreover, the pixel intensity distributions usually offer more stable information than pixel intensities themselves, while the joint histogram method just counts the number of occurrences of pixel intensity pairs.

For the parametric methods, although some multivariate models such as multivariate Gaussian, Gamma distribution have been used [Chatelain et al., 2007], the distributions of the image pixel intensity in the real world usually do not obey the Gaussian or other certain probability distributions. Furthermore, the multivariate distributions require that the types of marginal distributions are consistent. However, if the margins do not have the same type of distributions, for example, one image is Gaussian distributed, and another one is Gamma distributed, then there is no obviously known multivariate distribution model that can deal with this situation.

3.2.1 Use of copulas

Copulas [Nelsen, 1999] offer an alternative robust parametric based technique for the modeling of the dependency structure between random variables. The copula function extends the correlation concept to a wider dependence class, which may be not linear and it separates the marginal distributions and dependency structure.

Due to the special relationship between copula density functions, joint probability density function and marginal density functions shown in Eq. (2.45), copula functions offer a natural and robust way for the estimation of the divergence based information from observed data. It is shown later that that the Csiszar, Renyi-like, modified Bregman and Burbea-Rao divergence based information can be expressed only by copula density function.

3.3 Generalized divergence based information and their copula representation

In this section, the definitions of Csiszar, Renyi-like, modified Bregman and Burbea-Rao divergences based information in terms of copula density functions are introduced. Some limitations of Bregman and Burbea-Rao divergences have been found. To improve them, modified definitions are proposed and the test of smallest enclosing curve for Itakura-Saito and modified Itakura-Saito are carried out. Another test is K-means classification by using Itakura-Saito [Banerjee et al., 2005] and the modified Itakura-Saito as the distance have been applied to check the efficiency of the modified Bregman divergence.

3.3.1 Csiszar divergence

Csiszar divergence is also called f divergence which was introduced and studied independently by Csiszar [Csiszar, 1967], Ali and Silvery [Ali & Silvey, 1966]. Considering two probability density functions $p(x)$ and $q(x)$ for two probability distributions P and Q . As mentioned earlier, the Csiszar divergence measures the difference between P and Q and it can be written as:

$$D_{CS}(P, Q) = \int q(x) f\left(\frac{p(x)}{q(x)}\right) dx \quad (3.1)$$

where kernel function $f(\cdot)$ is a convex function on $[0, +\infty]$. With the different kernel functions, several Csiszar divergences such as Kullback-Leibler, Tsallis, I_α , Chi-square and Matusita divergences have been defined, and are shown in Table 3.2. Considering P and Q as two symmetrical Bernoulli distributions [Stuart & Ord, 1987] such as $P = \{t, 1-t\}$ and $Q = \{1-t, t\}$ with $0 \leq t \leq 1$. The Csiszar divergence for P and Q

Chapter 3: Divergence based Information using Copulas

Csiszar Divergence	Kernel function $f(t)$	Csiszar Divergence $D_{Cs}(p, q)$	Range
Kullback –Leibler	$t \log t$	$\int_x p(x) \log \left[\frac{p(x)}{q(x)} \right] dx$	
Tsallis	$\frac{t^\alpha - \alpha t + \alpha - 1}{(\alpha - 1)}$	$\frac{1}{\alpha} \left[\int_x \frac{p^\alpha(x)}{q^{\alpha-1}(x)} dx - 1 \right]$	$\alpha \neq 0$
I_α	$\frac{t^\alpha - \alpha t + \alpha - 1}{\alpha(\alpha - 1)}$	$\frac{1}{\alpha(\alpha - 1)} \left[\int_x \frac{p^\alpha(x)}{q^{\alpha-1}(x)} dx - 1 \right]$	$\alpha \neq 0$ $\alpha \neq 1$
χ^2	$(t - 1)^2$	$\int_x \frac{p^2(x)}{q(x)} dx - 1$	
Matusita	$ t^\alpha - 1 ^\frac{1}{\alpha}$	$\int_x p^\alpha(x) - q^\alpha(x) ^\frac{1}{\alpha}$	$0 < \alpha \leq 1$

Table 3.2: Csiszar Divergence

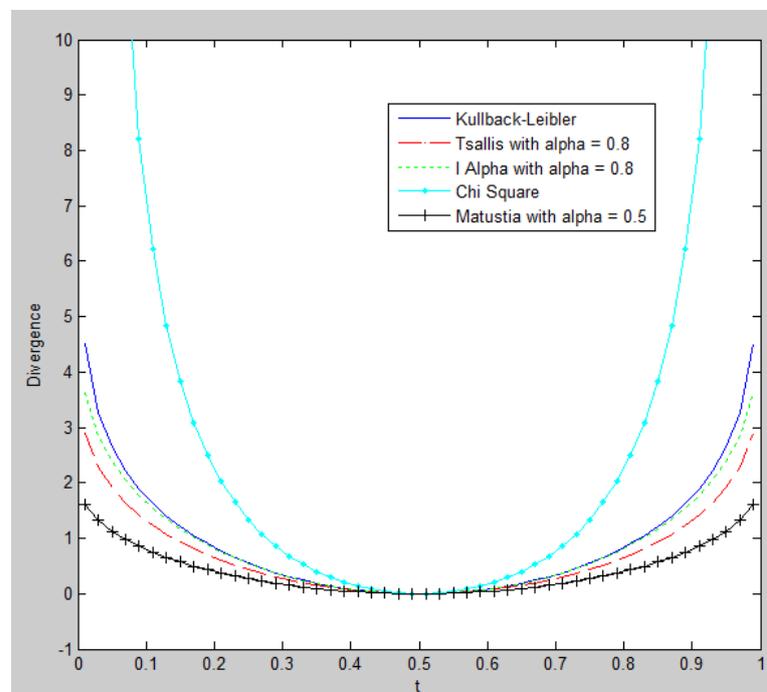


Figure 3.1: Comparison of Csiszar divergence

Chapter 3: Divergence based Information using Copulas

is shown in Figure 3.1. It can be found that all the divergence reach the minimal value 0 when $t=0.5$ ($P=Q$). All the divergences become greater when t moves away from 0.5 (P moves away from Q).

It has been pointed out in Section 3.2 that it is difficult to estimate the joint density probability function for non-elliptical case. This difficulty can be solved by exploiting copula density function. For example, considering the Kullback-Leibler divergence with the kernel function: $f(t) = t \log t$.

$$D(P, Q) = \int_x q(x) f\left[\frac{p(x)}{q(x)}\right] dx = \int_x q(x) \frac{p(x)}{q(x)} \log\left[\frac{p(x)}{q(x)}\right] dx = \int_x p(x) \log\left[\frac{p(x)}{q(x)}\right] dx$$

Then, the corresponding Kullback-Leibler divergence based information (Mutual information) becomes:

$$D_{KL}(x, y) = \iint_{XY} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{q_X(x)q_Y(y)} dx dy \quad (3.2)$$

The joint probability density function has been found that it is the product of copula density and marginal densities and shown in the Eq. (2.45):

$$p_{XY}(x, y) = c(u, v)q_X(x)q_Y(y)$$

The Kullback-Leibler divergence based information can be obtained as:

$$D_{KL}(x, y) = \iint_{X,Y} p_{XY}(x, y) \log \frac{p_{XY}(x, y)}{q_X(x)q_Y(y)} dx dy = \iint_{X,Y} c(u, v)q_X(x)q_Y(y) \log c(u, v) dx dy$$

where u and v are marginal distributions with the marginal density function $q_X(x)$ and $q_Y(y)$ respectively, and note following equations,

$$\begin{cases} du = q_X(x)dx \\ dv = q_Y(y)dy \end{cases}$$

Chapter 3: Divergence based Information using Copulas

So, the final expression of Kullback-Leibler divergence based information by using copulas can be found as:

$$D_{KL}(x, y) = \iint_{[0,1]^2} c(u, v) \log c(u, v) dudv \quad (3.3)$$

Precisely, the more generalized formula for Csiszar divergence can be expressed as:

$$D_{CS}(x, y) = \iint_{x,y} q_X(x)q_Y(y)f(c)dxdy = \iint_{[0,1]^2} f[c(u, v)]dudv \quad (3.4)$$

According to the Eq. (3.4), all the Csiszar divergence based information can be expressed by using copula density function only and are listed as given in Table 3.3.

Csiszar divergence has the non-negative and convexity properties. It has the minimal value 0 only when two distributions are equivalent. Another important property is the linear property [Osterreicher, 2002], considering two kernel functions of two distributions P and Q with marginal probability density function $p=p(x)$ and $q=q(x)$ respectively, the first kernel function is $f_1()$ and the corresponding Csiszar divergence D_1 can be calculated as:

$$D_1(P, Q) = \iint_{x,y} qf_1\left(\frac{p}{q}\right)dxdy = \iint_{[0,1]^2} f_1[c(u, v)]dudv$$

The second kernel function is $f_2(x)$ and the corresponding Csiszar divergence D_2 is calculated as:

$$D_2(P, Q) = \iint_{x,y} qf_2\left(\frac{p}{q}\right)dxdy = \iint_{[0,1]^2} f_2[c(u, v)]dudv$$

Let the third kernel function be $f_{12}() = f_1() + f_2()$, then the corresponding Csiszar divergence is:

$$D_{12}(P, Q) = \iint_{x,y} qf_{12}\left(\frac{p}{q}\right)dxdy$$

Csiszar Divergence	Csiszar Divergence based information	Using copulas
Kullback-Leibler	$\iint_{x,y} p_{XY}(x,y) \log \left[\frac{q_X(x)}{q_Y(y)} \right] dx dy$	$\iint_{[0,1]^2} c(u,v) \log c(u,v) du dv$
Tsallis	$\frac{1}{\alpha} \left[\iint_{x,y} \frac{p_{XY}^\alpha(x,y)}{q_X^{\alpha-1}(x) q_Y^{\alpha-1}(y)} dx dy - 1 \right]$	$\frac{1}{(\alpha-1)} \iint_{[0,1]^2} c^\alpha(u,v) du dv - 1$
I_α	$\frac{1}{\alpha(\alpha-1)} \left[\iint_{x,y} \frac{p_{XY}^\alpha(x,y)}{q_X^{\alpha-1}(x) q_Y^{\alpha-1}(y)} dx dy - 1 \right]$	$\frac{1}{\alpha(\alpha-1)} \iint_{[0,1]^2} c^\alpha(u,v) du dv - 1$
χ^2	$\iint_{x,y} \frac{p_{XY}^2(x,y)}{q_X(x) q_Y(y)} dx dy - 1$	$\iint_{[0,1]^2} [c(u,v) - 1]^2 du dv$
Matusita	$\iint_{x,y} p_{XY}^\alpha(x,y) - q_X^\alpha(x) q_Y^\alpha(y) ^{\frac{1}{\alpha}}$	$\iint_{[0,1]^2} [c(u,v)]^\alpha - 1 du dv$

Table 3.3: Csiszar divergence based information expressed in terms of copula density functions

$$D_1(P, Q) + D_2(P, Q) = q \left[f_1\left(\frac{P}{q}\right) + f_2\left(\frac{P}{q}\right) \right]$$

Notice $f_{12}(0) = f_1(0) + f_2(0)$ then, the above formula becomes:

$$D_1(P, Q) + D_2(P, Q) = q f_{12}\left(\frac{P}{q}\right) = D_{12}(P, Q) \quad (3.5)$$

The Csiszar divergence based information has been successfully applied for medical image registration in [Pluim et al., 2004].

3.3.2 Renyi-like divergence

Chapter 3: Divergence based Information using Copulas

Considering two probability density functions $p=p(x)$ and $q=q(x)$ for two probability distributions P and Q , Renyi-like divergence can be defined as [Martin & Durrani, 2007], [Martin, 2006]:

$$D_{\text{Renyi}}(P, Q) = \log[f^{-1}(pf(\frac{P}{q}))] \quad (3.6)$$

where the kernel function $f(\cdot)$ is a monotonously continuous function on $[0, +\infty)$. The Renyi divergence is defined as the special case when the kernel function $f(t) = t^{r-1}$.

The inverse kernel function is $f^{-1}(x) = x^{\frac{1}{r-1}}$, so the Renyi divergence can be calculated as:

$$D_{\text{Ren}}(P, Q) = \log[f^{-1}(pf(\frac{P}{q}))] = \frac{1}{r-1} \log(p^r q^{1-r}) \quad (3.7)$$

It has been found that Renyi-like divergence is the generalized expression of the Kullback-Leibler divergence by L'Hopital Rule [He et al., 2003]:

$$\lim_{r \rightarrow 1} D_{\text{Ren}}(P, Q) = D_{\text{KL}}(P, Q)$$

The Renyi-like divergences have been summarized in the Table 3.4.

Considering P and Q as two symmetrical Bernoulli distributions such as $P = \{t, 1-t\}$ and $Q = \{1-t, t\}$ with $0 \leq t \leq 1$. The Renyi-like divergence as a function of t is shown in Figure 3.2. Note that when $t=0.5$, the Renyi divergence reaches a minimal value 0 for all the Renyi-like divergence and becomes greater when t moves away from 0.5.

Similar to the case of the Csiszar divergence based information, copula density function can be used to calculate the Renyi divergence based information as:

$$D_{\text{Ren}} = \frac{1}{(r-1)} \log \iint_{X,Y} \frac{p_{XY}^r(x,y)}{q_X^{r-1}(x)q_Y^{r-1}(y)} dx dy = \frac{1}{(r-1)} \log \iint_{[0,1]^2} c^r(u,v) du dv \quad (3.8)$$

Renyi-like Divergence	Kernel function $f(t)$	Renyi-like Divergence D_{Renyi}	Range
Renyi	t^{r-1}	$\frac{1}{r-1} \int_x \log \left[\frac{p^r(x)}{q^{1-r}(x)} \right] dx$	$r > 0, r \neq 1$
Bhattacharyya	$t^{-\frac{1}{2}}$	$-2 \log \int_x \sqrt{p(x)q(x)} dx$	

Table 3.4 Renyi-like divergence

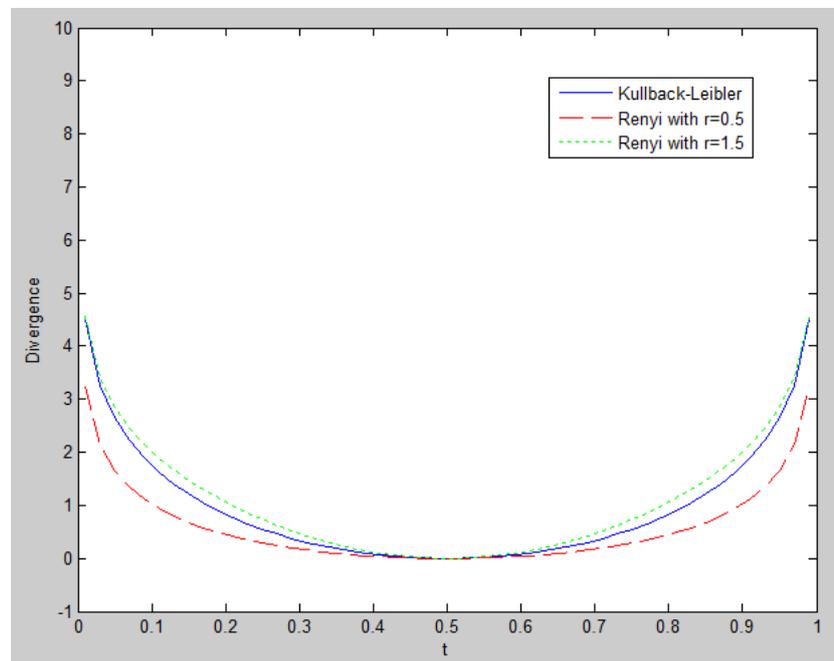


Figure 3.2: Comparison of Renyi-like divergence

Some of Renyi-like divergence based information are listed in the Table 3.5. It has been found that the Renyi-like divergence based information offers a more robust method than mutual information, achieved by adjusting its parameters and successfully applied for image registration [He et al., 2003].

Renyi-like Divergence	Divergence based information	Using copulas
Renyi	$\frac{1}{(r-1)} \log \iint_{x,y} \frac{p_{XY}^r(x,y)}{q_X^{r-1}(x)q_Y^{r-1}(y)} dx dy$	$\frac{1}{(r-1)} \log \iint_{[0,1]^2} c^r(u,v) du dv$
Bhattacharyya	$-2 \log \iint_{x,y} \sqrt{p_{XY}(x,y)q_X(x)q_Y(y)} dx dy$	$-2 \log \iint_{[0,1]^2} \sqrt{c(u,v)} du dv$

Table 3.5: Renyi-like divergence-based information expressed in terms of copula density functions

3.3.3 Bregman divergence

Bregman divergence is derived from Taylor expansion of a function and can be defined as [Banerjee et al, 2005]: Any function $f(x)$ satisfying the differentiable conditions can be expressed as a Taylor series at the point $x=y$ as:

$$f(x) = f(y) + f'(y)(x-y) + \frac{f''(y)}{2!}(x-y)^2 + \dots + \frac{f^{(n)}(y)}{n!}(x-y)^n + \dots$$

Bregman divergence can be intuitively thought of as the difference between the value of strictly convex and differentiable function $f()$ at the point x and the value of the first order Taylor expansion of $f()$ around point y evaluated at point x in Figure 3.3.

Specifically, Bregman divergence has been defined as [Bregman, 1967]:

$$D_{Bre}(x, y) = f(x) - f(y) - (x-y)f'(y) \quad (3.9)$$

Considering two probability distribution $p(x)$ and $q(x)$ for random variables X , some Bregman divergences can be obtained by different kernel functions. For example, the

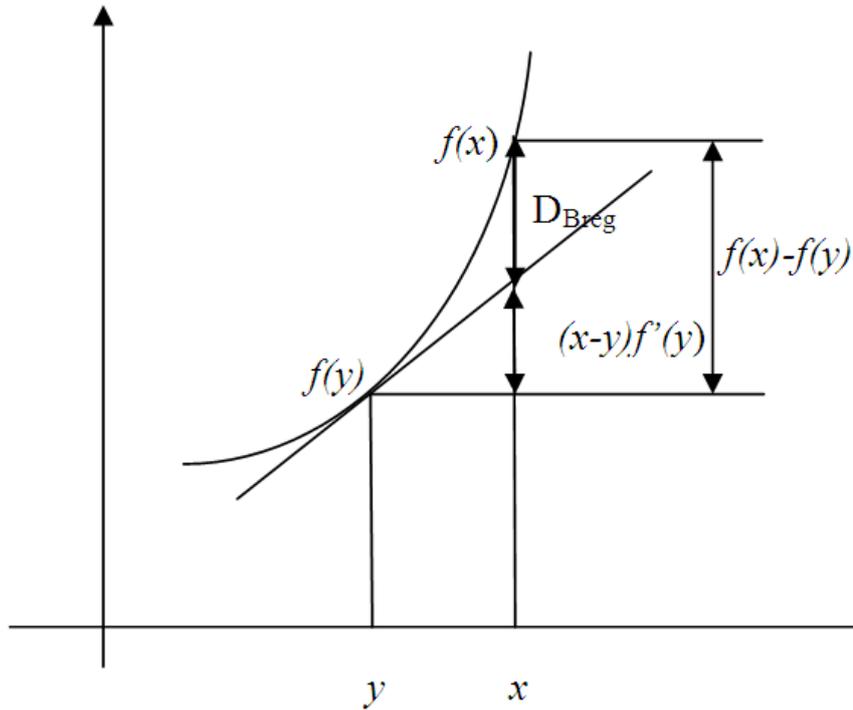


Figure 3.3: Bregman divergence

square loss divergence is defined by the kernel function $f(t) = t^2$, then the corresponding Bregman divergence becomes $[p(x) - q(x)]^2$.

If the kernel function is $f(t) = -\log(t)$, the corresponding Bregman divergence is

$$\frac{p(x)}{q(x)} - \log \frac{p(x)}{q(x)} - 1$$

and this is called Itakura-Saito divergence.

In the special case $f(t) = t \log(t)$, the corresponding Bregman divergence becomes

$$p(x) \log \frac{p(x)}{q(x)} - [p(x) - q(x)]$$

and is the Kullback-Leibler divergence.

The Bregman divergence has the properties of non-negativity, convexity, linearity, and so on [Banerjee et al., 2005]. All the Bregman divergences have been summarized in the Table 3.6.

Assume P and Q as two symmetrical Bernoulli distributions such as $P = \{t, 1-t\}$ and $Q = \{1-t, t\}$ with $0 \leq t \leq 1$. The diagram of Bregman divergence is given in the Figure 3.4. All the divergences have a minimal value 0 when $P=Q$ and their value increases when P moves away from Q .

Bregman Divergence	Kernel function $f(t)$	Bregman divergence D_{Breg}
Squared loss	t^2	$\int_x [p(x) - q(x)]^2 dx$
Itakura-Saito	$-\log(t)$	$\int_x \left(\frac{p(x)}{q(x)} - \log \frac{p(x)}{q(x)} - 1 \right) dx$
Kullback-Leibler	$t \log(t)$	$\int_x \left\{ p(x) \log \frac{p(x)}{q(x)} - [p(x) - q(x)] \right\} dx$

Table 3.6 Bregman divergence

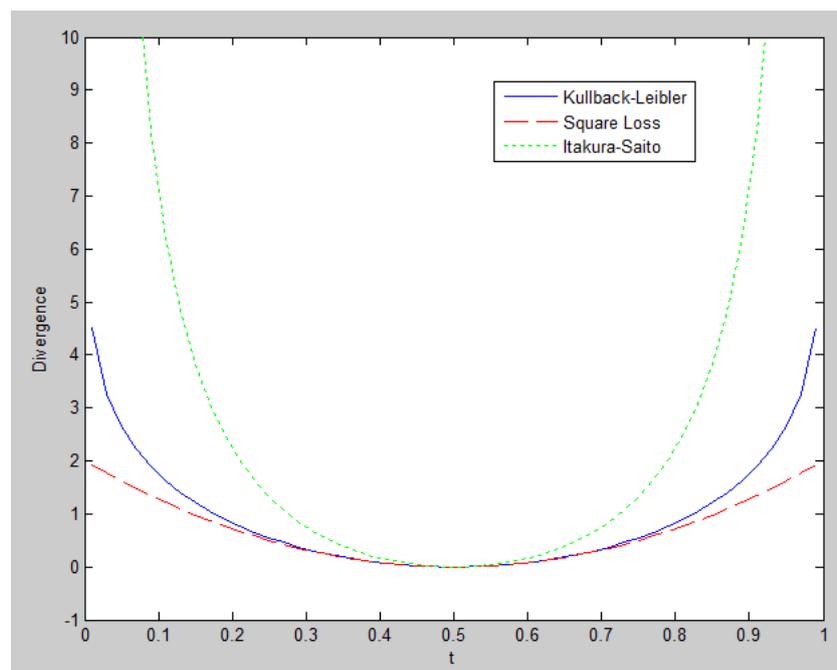


Figure 3.4: Comparison of Bregman divergence

The Bregman divergence has been applied for data clustering in [Banerjee et al., 2005], and the Itakura-Saito divergence is used to image analysis and classification [Ozturk & Abut 1990] and enhance speech [Cho et al., 2001].

3.3.4 Modified Bregman divergence

Chapter 3: Divergence based Information using Copulas

It may be noted from the square loss divergence that as long as $(x-y)$ does not change then the square loss divergence will be invariable. Similarly, as long as (x/y) does not change, the Itakura-Saito divergence will be invariable as well. It means that the square loss and Itakura-Saito divergences are only sensitive to the value of $(x-y)$ and (x/y) respectively. Note that they cannot distinguish cases that have the same value of $(x-y)$ and (x/y) respectively. Furthermore, since the Bregman divergence is derived from the Taylor expansion, and only the first order item is considered. To improve the sensitivity of the Bregman divergence, the second order of Taylor expansion may be considered. According to the Taylor expansion:

$$f(x) \approx f(y) + (x-y)f'(y) + \frac{(x-y)^2}{2} f''(y)$$

$$D_{Bre} = f(x) - f(y) - (x-y)f'(y) \approx \frac{(x-y)^2}{2} f''(y)$$

$$\frac{f(x) - f(y) - (x-y)f'(y)}{yf''(y)} \approx \frac{(x-y)^2}{2y} \quad (3.10)$$

Notice the right hand side of the equation (3.10), $\frac{(x-y)^2}{2y}$ is just the χ^2 (Chi-square) divergence [Pardo et al., 2003] which is given in Table 3.3 as belonging to the family of Csiszar divergence.

Considering the second order item of Taylor expansion as:

$$f(x) = f(y) + (x-y)f'(y) + \frac{(x-y)^2}{2} f''(y) + D_1$$

$$D_1 = f(x) - f(y) - (x-y)f'(y) - \frac{(x-y)^2}{2} f''(y)$$

$$D_{Bre} = D_1 + \frac{(x-y)^2}{2} f''(y) \quad (3.11)$$

Divide $yf''(y)$ both sides of equation (3.11),

Chapter 3: Divergence based Information using Copulas

$$\frac{D_{Bre}}{yf''(y)} = \frac{D_1}{yf''(y)} + \frac{(x-y)^2}{2y}$$

Defining $D_{Bnew} = \frac{D_{Bre}}{yf''(y)}$ and $D_{1new} = \frac{D_1}{yf''(y)}$, then

$$D_{Bnew} = D_{1new} + \frac{(x-y)^2}{2y}$$

So, the new Bregman divergence can be defined as:

$$D_{Bnew} = \frac{f(x) - f(y) - (x-y)f'(y)}{yf''(y)} = \frac{D_{Bre}}{yf''(y)} \quad (3.12)$$

The modified Bregman divergence has the non-negative property, and it has the minimum value 0 only when $x=y$. It also has the convexity property only when $y>0$.

For example, if the kernel function $f(t) = t^2$, then $D_{1new} = 0$; $D_{Bnew} = \frac{(x-y)^2}{2y}$ is just the Chi-Square divergence. From the above modified squared loss divergence based information becomes:

$$\iint_{x,y} \frac{[p_{XY}(x,y) - q_X(x)q_Y(y)]^2}{2q_X(x)q_Y(y)} dx dy = \frac{1}{2} \iint_{[0,1]^2} [c(u,v) - 1]^2 dudv$$

The modified Bregman divergence based information expressed in terms of copulas are listed in Table 3.7. It may be found that the modified Bregman divergence can be expressed by copula density function only, the joint and marginal probability density functions do not need to be considered any more.

Assume P and Q as two symmetrical Bernoulli distributions such as $P = \{t, 1-t\}$ and $Q = \{1-t, t\}$ with $0 \leq t \leq 1$. Then the modified Bregman divergence may be compared with the conventional Bregman divergence, and the results are shown in Figure 3.5. It can be found that all the divergence values are equivalent to 0 when $t=0.5$ and, the Kullback-Leibler divergence is congruent with the modified Itakura-Saito divergence in this experiment of symmetrical Bernoulli distribution.

Modified Bregman	Divergence based information	Using copulas
Squared loss	$\iint_{x,y} \frac{[p_{XY}(x,y) - q_X(x)q_Y(y)]^2}{2q_X(x)q_Y(y)} dx dy$	$\frac{1}{2} \iint_{[0,1]^2} [c(u,v) - 1]^2 du dv$
Itakura-Saito	$\iint_{x,y} \left[\frac{p_{XY}(x,y)}{q_X(x)q_Y(y)} - \log \frac{p_{XY}(x,y)}{q_X(x)q_Y(y)} - 1 \right] \times q_X(x)q_Y(y) dx dy$	$- \iint_{[0,1]^2} \log c(u,v) du dv$
Kullback-Leibler	$\iint_{x,y} [p_{XY}(x,y)] \times \left[\log \frac{p_{XY}(x,y)}{q_X(x)q_Y(y)} \right] - [p_{XY}(x,y) - q_X(x)q_Y(y)] dx dy$	$\iint_{[0,1]^2} c(u,v) [\log c(u,v)] du dv$

Table 3.7: Modified Bregman divergence based information expressed in terms of copulas

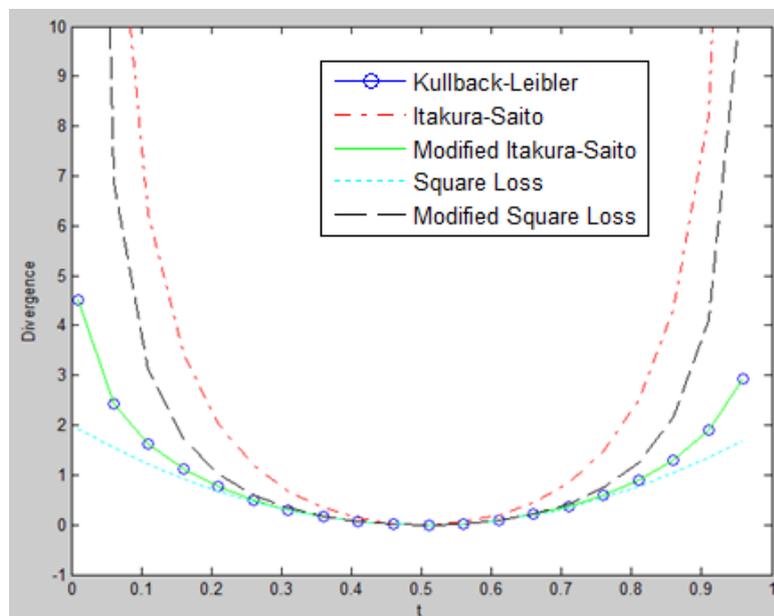


Figure 3.5: Comparison of modified Bregman divergence

3.3.5 Smallest enclosing ball (curve)

Given a set of observed data $S = \{S_1, S_2\}$ where S_1 and S_2 are vector data, the smallest enclosing Bregman ball for S is a Bregman ball $B_{c,r}$ with the minimal radius r so that $S \in B_{c,r}$ [Nock et al., 2005].

$$B_{c,r} = \{s \in S : D_{Breg}(c, s) \leq r\}$$

The smallest enclosing Bregman ball may be calculated using centre c and radius r . The radius r can be considered as the maximal Bregman divergence between the centre c and s which is an arbitrary point in data S . The algorithm for estimation of the centre has been described as [Nock et al., 2005]:

Input: Data $S = \{S_1, S_2\}$

Output: Centre c

Initialization: choose a point randomly in data S as centre c .

Choose a loop size T

Loop start

for $t=1$ to $T-1$

Find the point s_{max} in S which has the maximal Bregman divergence between c and each point in S .

$$c = f'^{-1}\left[\frac{t}{t+1} f'(c) + \frac{1}{t+1} f'(s_{max})\right]$$

where function $f()$ is kernel function

end

Loop end

Chapter 3: Divergence based Information using Copulas

Considering a point with coordinate $(x = 0.3709, y = 0.2878)$ as the centre of Itakura-Saito divergence and the radius is 0.7276, then the corresponding smallest enclosing ball can be drawn as shown in the Figure 3.6. Similarly, the smallest enclosing curve for modified Itakura-Saito divergence can be drawn by using the same algorithm as well. It should note that the shape for modified Itakura-Saito divergence is not a closed figure so that it is called smallest enclosing curve instead of ball. This is because that there are two corresponding answers when we look for the maximal Itakura-Saito divergence between c and each point in data S , however, there is only one answer for modified Itakura-Saito divergence.

3.3.6 K-means classification by modified Bregman divergence

K-means [Seber, 1984] is an unsupervised algorithm for data classification, it assigns each sample to the cluster whose centre is nearest, and the centre is the average of all the samples in the cluster. Firstly, choose the number of cluster k , and then randomly choose k samples as the cluster centre then assign each sample to the cluster centre which has the shortest distance between the centre and sample, and re-compute the new cluster centre. Repeat until the convergence criterion is met. The distance can be squared Euclidean, Bregman divergence [Banerjee et al., 2005] such as Itakura-Saito divergence [Jain, 2010], and so on.

In the following experiment, a 1000x2 dataset is randomly generated with Gaussian distribution. The squared Euclidean, Itakura-Saito and modified Itakura-Saito as the distances for K-means algorithm are tested to check their ability for data classification.

The classification results are shown in Figure 3.7. The top left corner is the original data. The top right corner, the lower left corner and the lower right corner are the K-means classification results with the distance of Itakura-Saito, squared Euclidean and modified Itakura-Saito divergence respectively. The original data have been separated to two classifications: blue points and red points with symbol and the crosses in circle are their centres.

As can be seen, the results in Figure 3.7 show that the modified Itakura-Saito divergence and squared Euclidean as the distances of K-means are more reliable than conventional Itakura-Saito divergence for this Gaussian distributed dataset since some

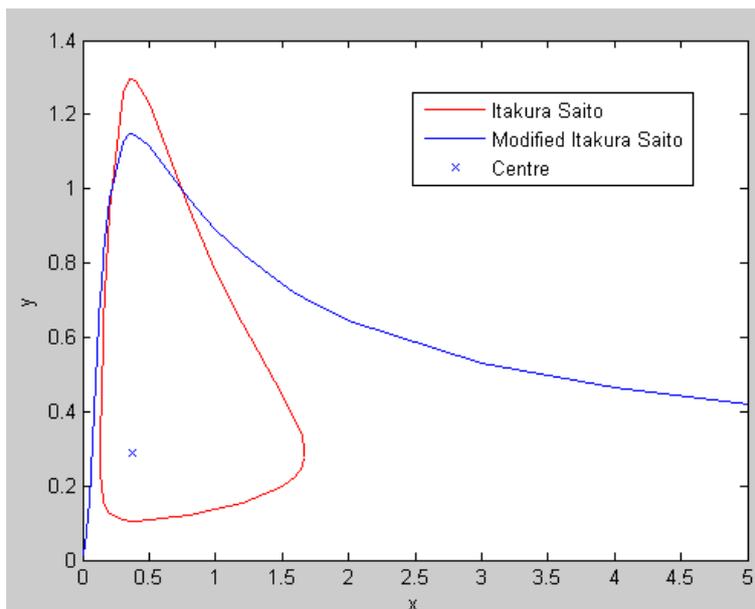


Figure 3.6: Smallest enclosing Itakura-Saito ball and smallest enclosing modified Itakura-Saito curve

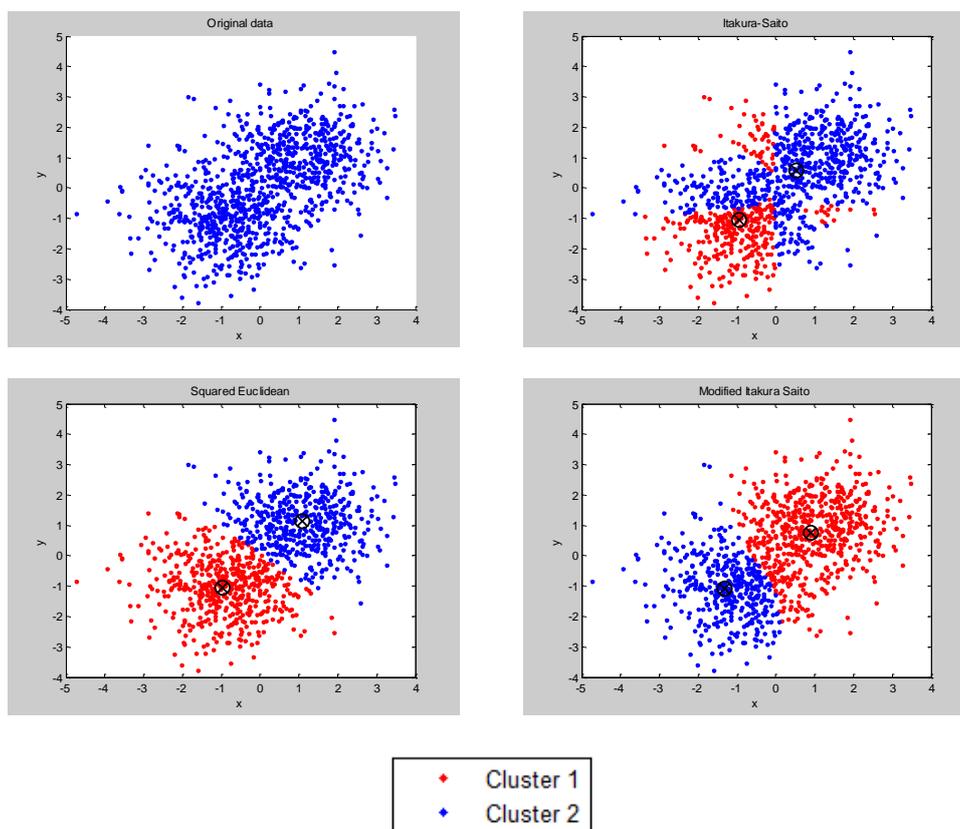


Figure 3.7: K-means classification using squared Euclidean, Itakura-Saito and modified Itakura-Saito divergence as distance

samples are classified into cluster 1 by using Itakura-Saito divergence, however, they are obviously more close to cluster 2, and more reasonable to be classified into cluster 2.

3.3.7 Burbea-Rao divergence

The definition of Burbea-Rao divergence can be found in [Burbea et al., 1982] and is defined as follow:

$$D_{BR} = \frac{f(x) + f(y)}{2} - f\left(\frac{x+y}{2}\right) \quad (3.13)$$

where the kernel function $f(\cdot)$ is convex function, the intuitive diagram of Burbea-Rao divergence can be found in Figure 3.8. Some Burbea-Rao divergences have been summarized in Table 3.8.

Assume P and Q to be two symmetrical Bernoulli distributions such as $P = \{t, 1-t\}$ and $Q = \{1-t, t\}$ with $0 \leq t \leq 1$. The diagram of Burbea-Rao divergence is shown in Figure 3.9. All the divergences reach the minimal value 0 when $P = Q$ and become greater when P moves apart from Q .

3.3.8 Modified Burbea-Rao divergence

Referring to the definition of Burbea-Rao divergence in Eq. (3.13) and considering the following equation:

$$2D_{BR} = \left[f(x) - f\left(\frac{x+y}{2}\right) \right] + \left[f(y) - f\left(\frac{x+y}{2}\right) \right]$$

Let D_{Bre1} be the Bregman divergence between x and $(x+y)/2$. D_{Bre2} be the Bregman divergence between y and $(x+y)/2$.

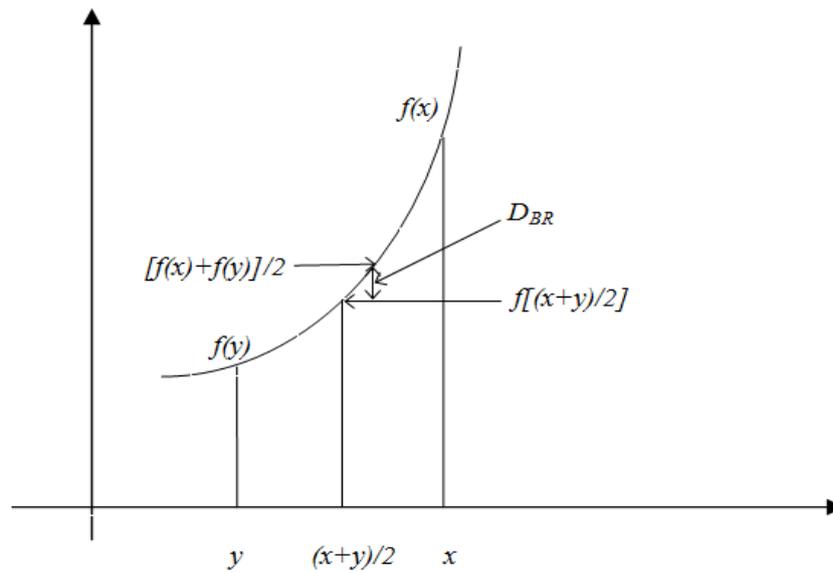


Figure 3.8: Burbea-Rao Divergence

Kernel function $f(t)$	Burbea-Rao divergence based information
t^2	$\int_x \frac{[p(x) - q(x)]^2}{4} dx$
$-\log(t)$	$\int_x \left(-\frac{\log[p(x)] + \log[q(x)]}{2} + \log \frac{p(x) + q(x)}{2} \right) dx$
$t \log(t)$	$\int_x \left\{ \frac{1}{2} \left[p(x) \log \frac{2p(x)}{p(x) + q(x)} + q(x) \log \frac{2q(x)}{p(x) + q(x)} \right] \right\} dx$

Table 3.8: Burbea-Rao divergence

$$\begin{cases} D_{Bre1} = f(x) - f\left(\frac{x+y}{2}\right) + \frac{x-y}{2} f'\left(\frac{x+y}{2}\right) \\ D_{Bre2} = f(y) - f\left(\frac{x+y}{2}\right) - \frac{x-y}{2} f'\left(\frac{x+y}{2}\right) \end{cases}$$

Then we can obtain the relationship between Bregman and Burbea-Rao divergence as:

$$D_{BR} = \frac{D_{Bre1} + D_{Bre2}}{2} \quad (3.14)$$

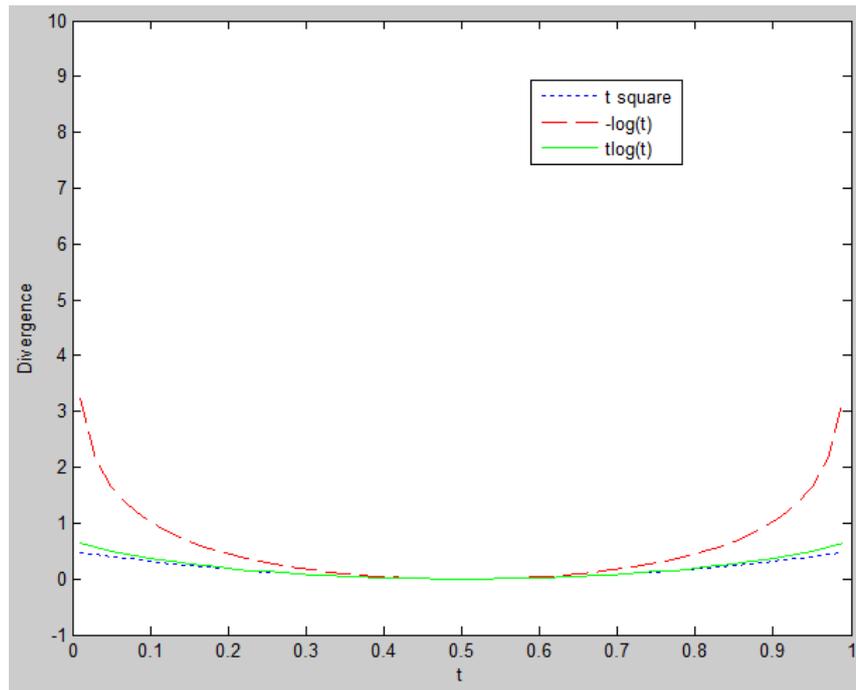


Figure 3.9: Comparison of Burbea-Rao Divergence

So, Burbea-Rao can be considered as the linear combination of two Bregman divergences. Hence, we can use the same approach as in Eq. (3.12) to modify the Burbea-Rao divergence as:

$$D_{BRNew} = \frac{D_{BR}}{yf''(y)} \quad (3.15)$$

The modified Burbea-Rao divergence based information may be expressed in terms of copula density functions as given in Table 3.9. The modified Burbea-Rao divergence and the modified Bregman divergence have similar properties, since Burbea-Rao divergence is the linear combination of Bregman divergence. Assume P and Q are two symmetrical Bernoulli distributions such as $P = \{t, 1-t\}$ and $Q = \{1-t, t\}$ with $0 \leq t \leq 1$. The diagram of modified Burbea-Rao divergence is shown in Figure 3.10. All the divergences reach the minimal value 0 when $P = Q$ and increases when P moves apart from Q .

3.3.9 Comparison of divergences

Chapter 3: Divergence based Information using Copulas

After introduction of four categories divergences: Csiszar, Renyi-like, modified Bregman and Burbea-Rao divergence and corresponding divergence based information; it may be found that all the divergences have similar performance: divergence has the minimal value when marginal distributions equal and increases when two marginal distributions move apart.

Moreover, Chi-square (Csiszar), Itakura-Saito (Bregman), modified square loss (modified Bregman) and modified square t divergence (modified Burbea-Rao divergence) can offer better ability to control the measurement sensitivity since these divergences change more obviously for the same change of marginal distributions (same change of t) than other divergences and they are different with other divergence which have the approximately value 0 when two distributions are close ($P \approx Q$ or $t \approx 0.5$).

In addition, better ability of discrimination can be acquired by adjusting the parameters of Tsallis, $I\alpha$ and Renyi divergences.

Finally, modified Itakura-Saito (Bregman) divergence offers much better classification performance of Gaussian distributed data than conventional Itakura-Saito divergence with the algorithm of K-means classification.

3.4 Calculation of divergence based information using copulas

Since all kinds of divergence based information have been defined, the next step is to compute them. Gaussian copula density function has complicated expression, so it is difficult to compute its double integral. However, it has the perfect correlation range so it is worth exploring the calculation of its double integral. The Gaussian copula density function has been defined as:

$$\begin{aligned}
 c(u, v) &= \frac{1}{\sqrt{1-\rho^2}} e^{\frac{[\varphi^{-1}(u)]^2 + [\varphi^{-1}(v)]^2}{2}} e^{\frac{-[\varphi^{-1}(u)]^2 - [\varphi^{-1}(v)]^2 + 2\rho\varphi^{-1}(u)\varphi^{-1}(v)}{2(1-\rho^2)}} \\
 &= \frac{1}{\sqrt{1-\rho^2}} \left(\frac{2\rho\varphi^{-1}(u)\varphi^{-1}(v) - \rho^2([\varphi^{-1}(u)]^2 + [\varphi^{-1}(v)]^2)}{2(1-\rho^2)} \right)
 \end{aligned} \tag{3.16}$$

Kernel function $f(t)$	Modified Burbea-Rao Divergence	Using Copulas
t^2	$\iint_{x,y} \frac{[p_{xy}(x,y) - q_x(x)q_y(y)]^2}{8q_x(x)q_y(y)}$	$\frac{1}{8} \iint_{[0,1]^2} [c(u,v) - 1]^2 dudv$
$-\log(t)$	$\iint_{x,y} q_x(x)q_y(y) \log \frac{p_{xy}(x,y) + q_x(x)q_y(y)}{2\sqrt{p_{xy}(x,y)q_x(x)q_y(y)}} dx dy$	$\iint_{[0,1]^2} \log \frac{c(u,v) + 1}{2\sqrt{c(u,v)}} dudv$
$t \log(t)$	$\frac{1}{2} \iint_{x,y} p_{xy}(x,y) \log \frac{2p_{xy}(x,y)}{p_{xy}(x,y) + q_x(x)q_y(y)} dx dy +$ $\frac{1}{2} \iint_{x,y} q_x(x)q_y(y) \log \frac{2q_x(x)q_y(y)}{p_{xy}(x,y) + q_x(x)q_y(y)} dx dy$	$\frac{1}{2} \iint_{[0,1]^2} c(u,v) \log \frac{2c(u,v)}{c(u,v) + 1} + \log \frac{2}{c(u,v) + 1} dudv$

Table 3.9: Modified Burbea-Rao divergence based information expressed in terms of copula density functions

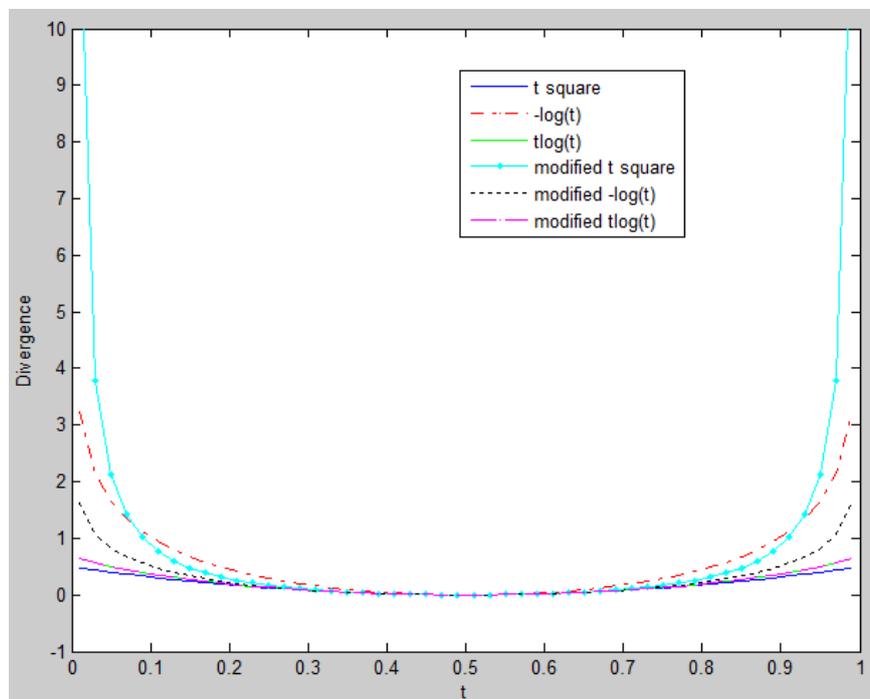


Figure 3.10: Comparison of Modified Burbea-Rao Divergence

Chapter 3: Divergence based Information using Copulas

The copula density based mutual information can be written as:

$$\iint_{[0,1]^2} [c(u,v)] \log[c(u,v)] dudv = \iint_{[0,1]^2} \frac{1}{\sqrt{1-\rho^2}} e^{\frac{2\rho\varphi^{-1}(u)\varphi^{-1}(v)-\rho^2([\varphi^{-1}(u)]^2+[\varphi^{-1}(v)]^2)}{2(1-\rho^2)}} \log\left[\frac{1}{\sqrt{1-\rho^2}} e^{\frac{2\rho\varphi^{-1}(u)\varphi^{-1}(v)-\rho^2([\varphi^{-1}(u)]^2+[\varphi^{-1}(v)]^2)}{2(1-\rho^2)}}\right] dudv \quad (3.17)$$

where function $\varphi^{-1}()$ is the inverse standard Gaussian cumulative distribution function. As can be seen, the expression of Gaussian copula based mutual information is very complicated to compute, and the inverse standard Gaussian cumulative distribution should be calculated firstly. The essence of calculation $\varphi^{-1}()$ is the process of calculation of inverse error function $erf^{-1}()$ by the following formula [Philip, 1960]:

$$\varphi^{-1}(z) = \sqrt{2}erf^{-1}(2z-1) \quad (3.18)$$

where $z \in [0,1]$. The following two ways can be used to calculate the inverse error function. The first way is based on the Taylor expansion and iteration algorithm, the inverse error function can be written as [Philip 1960]:

$$\left\{ \begin{array}{l} erf^{-1}(z) = \sum_{k=0}^{\infty} \frac{c_k}{2k+1} \left(\frac{\sqrt{\pi}}{2}z\right)^{2k+1} \\ c_k = \sum_{m=0}^{k-1} \frac{c_m c_{k-m-1}}{(m+1)(2m+1)} = \left\{1, 1, \frac{7}{6}, \frac{127}{90}, \dots\right\} \\ c_0 = c_1 = 1 \end{array} \right.$$

Then the inverse error function can be written as:

$$erf^{-1}(z) = \frac{\sqrt{\pi}}{2} \left(z + \frac{\pi}{12} z^3 + \frac{7\pi^2}{480} z^5 + \frac{127\pi^3}{40320} z^7 + \frac{4369\pi^4}{5806080} z^9 + \frac{34807\pi^5}{182476800} z^{11} + \dots \right) \quad (3.19)$$

The second way has been proposed by [Winitzki, 2008], the error function $erf(z)$ and inverse error function $erf^{-1}(z)$ can be expressed as:

$$erf(z) \approx \frac{z}{|z|} \sqrt{1 - e^{\frac{-x^2(\frac{\pi}{4} + ax^2)}{1+ax^2}}} \quad (3.20)$$

Chapter 3: Divergence based Information using Copulas

$$\operatorname{erf}^{-1}(z) \approx \frac{z}{|z|} \sqrt{\sqrt{\left(\frac{2}{\pi a} + \frac{\ln(1-z^2)}{2}\right)^2 - \frac{\ln(1-z^2)}{a}} - \left(\frac{2}{\pi a} + \frac{\ln(1-z^2)}{2}\right)} \quad (3.21)$$

where $a = \frac{8(\pi - 3)}{3\pi(4 - \pi)} \approx 0.140012$

The result of inverse Gaussian distribution by Taylor expansion shown in Eq. (3.19) and Winitzki shown in Eq. (3.21) are given in Figure 3.11.

These two methods have been validated by the rule: the double integral of copula density function equals to 1, it can be expressed by the following formula:

$$\iint_{[0,1]^2} c(u,v) du dv = \iint_{[0,1]^2} \frac{1}{\sqrt{1-\rho^2}} e^{\frac{2\rho\phi^{-1}(u)\phi^{-1}(v) - \rho^2([\phi^{-1}(u)]^2 + [\phi^{-1}(v)]^2)}{2(1-\rho^2)}} du dv = 1 \quad (3.22)$$

Finally, numerically evaluate double integral [Press et al., 1992], [Gander & Gautschi, 2000] to compute the Gaussian copula based mutual information. The results are given in Figure 3.12 and it can be found that the Gaussian copula based mutual information is very close to the bivariate Gaussian distribution based mutual information for the bivariate Gaussian distributed data.

Given two marginal Gaussian probability density functions $f_X(x)$ and $f_Y(y)$ of random variables X and Y respectively:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma_X^2}} e^{-\frac{(x-\mu_X)^2}{2\sigma_X^2}} ; f_Y(y) = \frac{1}{\sqrt{2\pi\sigma_Y^2}} e^{-\frac{(y-\mu_Y)^2}{2\sigma_Y^2}}$$

The joint density function $f_{XY}(x, y)$ has been given in Eq. (2.9):

According to the information theoretic definition of mutual information [Thomas et al., 1991]:

$$MI(X, Y) = H(Y) - H(Y | X) = H(X) + H(Y) - H(X, Y) \quad (3.23)$$

where H is Shannon entropy and the Shannon entropy of these two marginals can be

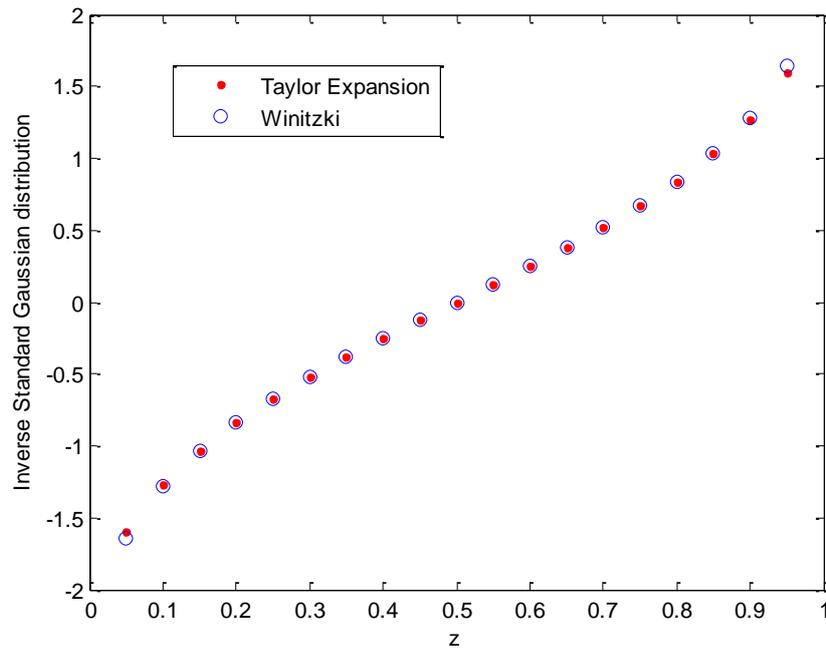


Figure 3.11: Inverse standard Gaussian cumulative distribution

calculated as:

$$H(X) = -\int_X f_X(x) \ln f_X(x) dx = \frac{1}{2} \ln(2\pi e \sigma_X^2)$$

$$H(Y) = -\int_Y f_Y(y) \ln f_Y(y) dy = \frac{1}{2} \ln(2\pi e \sigma_Y^2)$$

The entropy of joint distribution is:

$$H(X, Y) = \frac{1}{2} \ln(2\pi e)^2 |\Sigma|$$

where Σ is covariance and $\Sigma = \begin{bmatrix} \sigma_X^2 & \rho\sigma_X\sigma_Y \\ \rho\sigma_X\sigma_Y & \sigma_Y^2 \end{bmatrix}$

So, the mutual information of these two Gaussian marginal distribution is:

$$MI_{Gau} = -\frac{1}{2} \ln(1 - \rho^2) \quad (3.24)$$

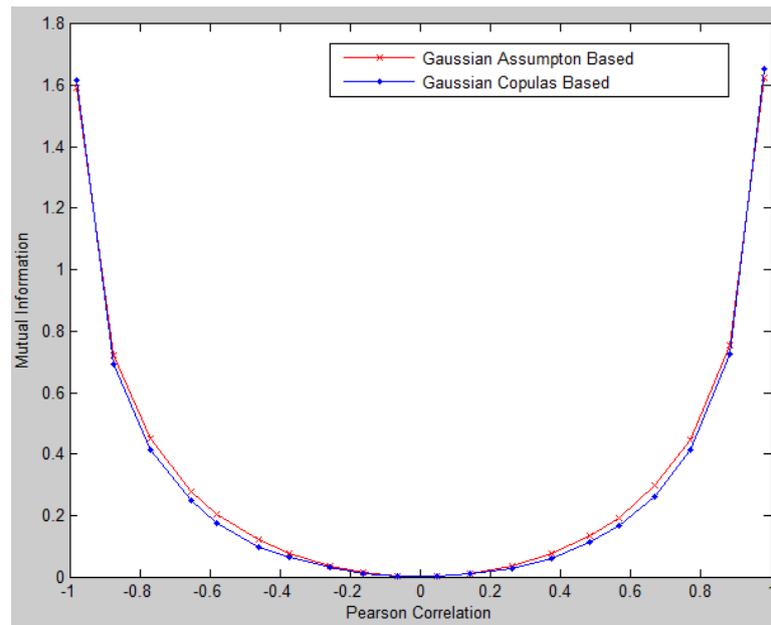


Figure 3.12: Bivariate Gaussian distribution and Gaussian copula based mutual information

It should note that if the probability distribution of data is far from Gaussian distribution then the bivariate Gaussian distribution is not appropriate any more. Moreover, if the types of marginal distributions are not consistent, there is no appropriate bivariate distribution can be used, and copula is a good choice to deal with the joint distributions with arbitrary marginal distributions.

3.5 Conclusion

In this chapter, the divergence-based information in terms of copula density function is introduced. It is shown that the Csiszar, Renyi-like divergence based information can be directly expressed using copula density function only. The limitations of Bregman divergence have been improved by proposing a modified definition. Similarly, the Burbea-Rao divergence is modified to provide a more effective definition. Copulas have been applied to estimate the mutual information, the results of mutual information by using copula is very close to the bivariate Gaussian

Chapter 3: Divergence based Information using Copulas

distribution based mutual information for Gaussian distributed data. The modified Bregman divergence has been validated by smallest enclosing curve and K-means classification.

Chapter 4

Copulas for Image Registration and Evaluation of Image Fusion Algorithms

Summary

Following the definition of four generalized divergence based information given in Chapter 3, in this chapter, image registration techniques are developed by maximizing the divergence-based information using copula density functions between the overlapping areas of the first (reference) image and the second (float) image which is transformed by rotating, translating and rescaling. Experiment results are provided on both synthetic and real data that show that the copulas-based methods offer more accurate registration results than Gaussian assumption and joint histogram based mutual information.

Algorithms such as simple average, Principal Component Analysis (PCA) [Jia, 1998], Gradient Pyramid (GP) [Burt and Kolczynski, 1993], Laplacian Pyramid (LP) [Burt and Adelson, 1983], Ratio Pyramid (RP) [Toet, 1989] and Discrete Wavelet Transform (DWT) [Mallat, 1989] methods have been successfully applied in conventional system for image fusion. A concern here is the evaluation of performance, which is difficult to determine. Mutual information (MI), Tsallis and Renyi divergence based information have been applied to assess the performance of image fusion without ground truth [Cvejic et al., 2006]. However they are difficult to estimate precisely. In this chapter, a copula based approach is proposed for evaluating the performance of image fusion algorithms. Copulas are proposed for the estimation of the MI, Tsallis and Renyi divergence based information and are used to evaluate the quality of image fusion.

4.1 Introduction

Image registration is the process of alignment geometrically two or more images taken for the same scene usually at different times, or from different viewpoints or by different sensors. It transforms the different sets of data into the same coordinate system. Image registration is a fundamental technique of image processing and has been applied widely for a range of fields such as in medical imaging for monitoring tumor growth, treatment verification; remote sensing, geographic information systems (GIS), image fusion, target localization, automatic quality control, and so on [Barbara et al., 2003].

The key for effective image registration is to find the optimal spatial relationship between two or more images. In the recent decades, image acquisition devices have developed and grown rapidly, the advanced automatic image registration techniques are required to be explored. Image registration techniques can be classified into two categories: [Barbara, et al., 2003] Feature based method and Area based methods. The feature based methods do not directly process image pixel intensity values. These techniques process features which represent higher level of information. Image features can be region features, line features, edge features and point features such as centroid. Feature based methods require that the image contain distinctive and easily detectable objects. These methods usually find application in remote sensing and compute vision. Area based methods work directly with the image pixel intensity in the region of interest of images. For example, medical images usually do not contain enough distinctive objects and thus area based methods are usually employed [Barbara et al., 2003]. For area based techniques, divergence based measure have been widely accepted as one of the most accurate and robust registration techniques [Maes et al., 1997], [Viola & Wells, 1997], [He et al., 2003]. In this chapter, divergence based information by using copulas is applied for image registration and is compared with Gaussian assumption and joint histogram based measure.

Image fusion is the process of combining relevant information from two or more images into a single image which is more informative than any of the input images. Image fusion has been applied widely in the fields of medical imaging, remote sensing image applications, and so on.

The area of information fusion has been received increasing attention in recent years due to the ready availability of multiple sensors. For multi-sensor images, algorithms such as simple average, PCA and multi-scale decomposition method which may be pyramidal transform such as Gradient Pyramid (GP) method, Laplacian Pyramid (LP) method and Ratio Pyramid (RP) method or wavelet transform such as Discrete Wavelet Transform (DWT) methods have been successfully applied for image fusion. Nevertheless a crucial issue that arises in image fusion is that the performance of image fusion with the associated algorithms is difficult to evaluate, especially when a clearly defined ground-truth image is not available.

To evaluate the effectiveness of image fusion techniques, divergence based information measures are usually used since it is feasible to assess the case without the ground truth [Cvejic et al., 2006]. As stated in the section 3.1, divergence is a measure of distance between the distributions P and Q of two random variables X and Y . One of the commonly used divergences is Kullback-Leibler divergence which has been defined in Table 3.2, and the Kullback-Leibler divergence based information: mutual information are defined in Table 3.3.

4.2 Current techniques and problem description

Considering the situation that some images maybe do not contain distinctive and easily detectable objects, so that the Feature based method are not feasible. In this thesis, we focus on the Area based registration by maximizing divergence based information. The key of divergence based information for image registration is to precisely estimate the joint probability density function. Gaussian assumption and joint histogram are two conventional approaches to estimate joint probability density function for image registration. The disadvantages of these two conventional measures have been pointed out in Section 3.2. Considering the special relationship between joint probability density function, copula density function and marginal probability density functions, and copulas can deal with any marginal distributions, copula functions offer a nature and robust way to estimate divergence based information, and so it will be applied for image registration in this chapter.

Chapter 4: Copulas for Image Registration and Evaluation of Image Fusion Algorithms

The disadvantage of mutual information based method is that it does not have parameter that could be adjusted to achieve the better results of performance evaluation of image fusion. In this section, some generalized divergence such as Tsallis and Renyi divergence which have been defined in Chapter 3 are used to evaluate the performance of image fusion since these divergence based information are considered that they can provide the better ability of discrimination by adjusting their parameters.

4.3 Copulas based image registration

In this section, an image registration criterion, optimal parameter searching method and image registration techniques are introduced, and we tested the synthetic MRI images, real medical images (CT-MRI) and visible light-thermal images. The results of image registration using copulas are compared with Gaussian assumption and joint histogram based mutual information.

4.3.1 Image registration criterion

Let I_1 and I_2 be two digital image pixel intensities. The objective of image registration is to find the optimal spatial transformation parameters which maximize the divergence-based information between the overlapping areas of reference image and transformed float image.

These transforming parameters usually include $c1$ and $c2$ which represent the geometric centre of I_1 and I_2 respectively, rescaling parameters and the rotating degree θ . For example, given the first image I_1 as reference image, we translate, rotate and rescale the second image I_2 until that the maximal divergence-based information between the overlapping area of reference image and transformed float image is found, and then the images are considered to be aligned correctly.

Two methods such as Gaussian assumption based mutual information and the joint histogram based mutual information [Maes, et al., 1997] will be computed and compared with the results obtained using copulas method. The copulas based mutual information has been defined as:

$$MI_{Cop} = \iint_{[0,1]^2} c(u,v) \log c(u,v) dudv \quad (4.1)$$

Gaussian assumption based mutual information has been defined in Eq. (3.24) as: [Thomas and Joy, 1991]

$$MI_{Gau} = -\frac{1}{2} \ln(1 - \rho^2) \quad (4.2)$$

where ρ is the Pearson correlation between variables.

The joint image intensity histogram of the overlapping volume of both images is computed by binning the image intensity pairs, the number of bins in the joint histogram is typically 256 [Maes, et al., 1997].

4.3.2 Optimization method

In image registration, estimation of maximal divergence based information is a multidimensional optimization problem, where the number of dimensions depends on how many parameters of geometrical transformation such as size rescale at x -coordinate and y -coordinate, translating at x -coordinate and y -coordinate and rotation degree are required to be estimated. The computation is complicated when several parameters are required to be estimated together or when the images have the large size.

In the case of transformations with more parameters, sophisticated optimization algorithms are necessary. Powell's multidimensional direction set method with Brent's method for one-dimensional optimization [Press, et al., 1992] has been applied to search the maximal value of divergence based information [Maes, et al., 1997]. In their work, the images are initially positioned such that their centre coincide and have the same orientation, to reduce the complexity of computing.

The Nelder-Mead Simplex method [Press, et al., 1992], [Lagarias, et al., 1998] is sometimes used as it is derivative-free, and as such is very fast.

The application of Gauss-Newton numerical minimization algorithm for minimizing the sum of squared differences is described in [Sharma, et al., 1997]; while [Viola, et al., 1997] have used the gradient descent optimization method to estimate the maximal mutual information and [Sawhney, et al., 1999] have used Levenberg-Marquardt optimization method for image registration.

It is worth noting that the initial parameters are crucial for these searching methods, since it is possible that these algorithms may yield local optimal value. To improve the accuracy and to reduce the complexity of computing, some useful techniques for image registration are introduced as follows.

4.3.3 Image registration techniques

In this section, some useful techniques of the selection of initial parameters for image registration are introduced. Firstly, the centre of gravity of image [Assen, et al., 2002] is often used to evaluate the initial position parameters for image registration. The centre of gravity normally keeps more stable position information than the geometrical centre for medical images, so the centre of gravity is a good choice for the initial searching parameter. The may be defined as:

Let $f(i, j)$ where $1 \leq i \leq N, 1 \leq j \leq N$ denotes the pixel intensity of image. The centre of gravity $C(C_x, C_y)$ can be calculated by the formula:

$$C_x = \frac{\sum_{i=1}^M \sum_{j=1}^N i \cdot f(i, j)}{\sum_{i=1}^M \sum_{j=1}^N f(i, j)} ; C_y = \frac{\sum_{i=1}^M \sum_{j=1}^N j \cdot f(i, j)}{\sum_{i=1}^M \sum_{j=1}^N f(i, j)} \quad (4.3)$$

Secondly, once an original pixel is given, there are some ways to obtain its neighbourhood. Firstly, the square neighbourhood [van den Boomgard and van Balen, 1992] are applied frequently. A square neighbourhood with parameter $r=3$ and window size 7×7 is shown in Figure 4.1.

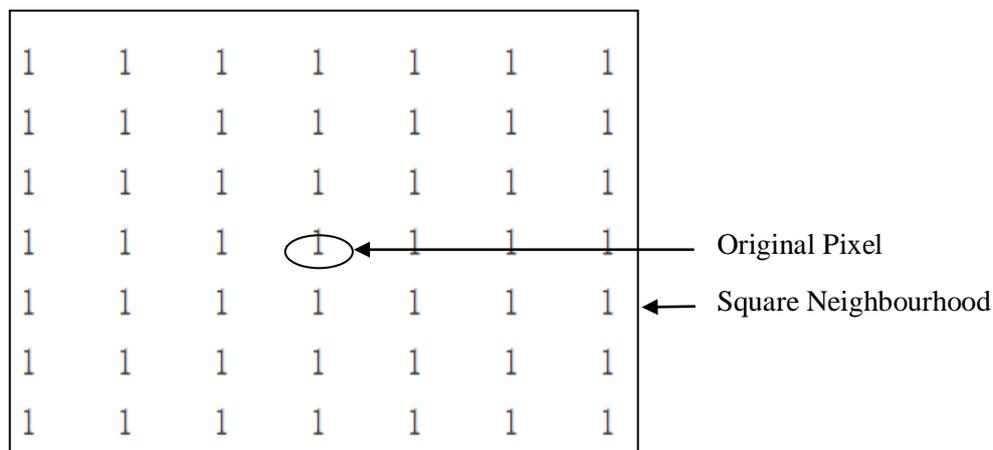


Figure 4.1: Square neighbourhood

Alternatively, the circle neighbourhood [Adams, 1993] can be used to estimate the rescale parameter since it is not affected by any rotation of image. The circle neighbourhood with radius $r=3$ is shown in Figure 4.2.

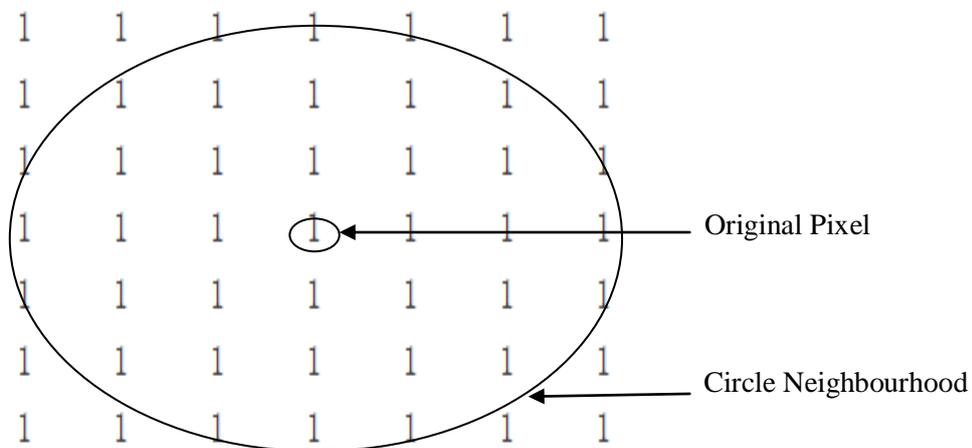


Figure 4.2: Circle neighbourhood

Another important technique for image registration is interpolation; the common methods are nearest neighbour, bilinear and bicubic interpolation.

In the nearest neighbour interpolation process [Parker, 1983], the problem involves approximating the value for a point in some space, when some values of points around that point are given. The nearest neighbour algorithm simply selects the value of the nearest point, and does not consider the values of other neighbouring points at all. The algorithm is very simple to implement however it may just make the neighbouring pixels have the same intensity values.

Bilinear interpolation [Parker, 1983] considers the closest 2x2 neighbourhood that surround the unknown pixel. A weighted average is then taken of the four pixels to arrive at its final interpolated value. The results usually have much smoother looking than nearest neighbour interpolation.

Bicubic interpolation [Keys, 1981] considers the surrounding 4x4 neighbours of unknown pixels. The closest pixels are given a higher weighting in the calculation. Bicubic interpolation is a more sophisticated and the interpolated surface is smoother than the nearest neighbourhood and bilinear interpolation methods.

4.3.4 Synthetic medical image registration

The image shown in Figure 4.4 (a) is a Magnetic Resonance Imaging (MRI) image of a human brain scan which can be found online at: (www.itk.org) and is used as the reference image with the size 249x249 pixels. The coordinate of centre of gravity of the reference image is [125.27; 124.96]. The float image shown in Figure 4.4 (b) which is obtained from the first image by rotating it by 10 counter clockwise degrees, and the algorithm of Bicubic interpolation is used, it has the same size with reference image. To improve robustness, 10%, 20% and 30% salts and pepper noise were added to the float images, as shown in Figure 4.4 (c), (d) and (e) respectively.

The objective is to find a point in float image as the centre, and then rotate the float image with respect to this centre until the maximal Gaussian assumption based mutual information, or the joint histogram based mutual information or divergence based information using copulas is found. Then to check how far these centres from the centres of gravity of float images are, and to evaluate the difference between the computed degree of rotation from the original rotation degree.

To initiate the algorithms, the centres of gravity of the float images are calculated as the initial parameter values. The copula parameters are calculated using the canonical maximal likelihood (CML) given in Equation (2.21). The Clayton copula model is selected for calculation of the divergence based information by copulas, and square neighbours are applied to choose the overlapping part of images. The Bicubic algorithm is used for image interpolations, and Nelder-Mead Simplex method is used for the searching the maximal value. The process of image registration is described in the following flow chart shown in Figure 4.3.

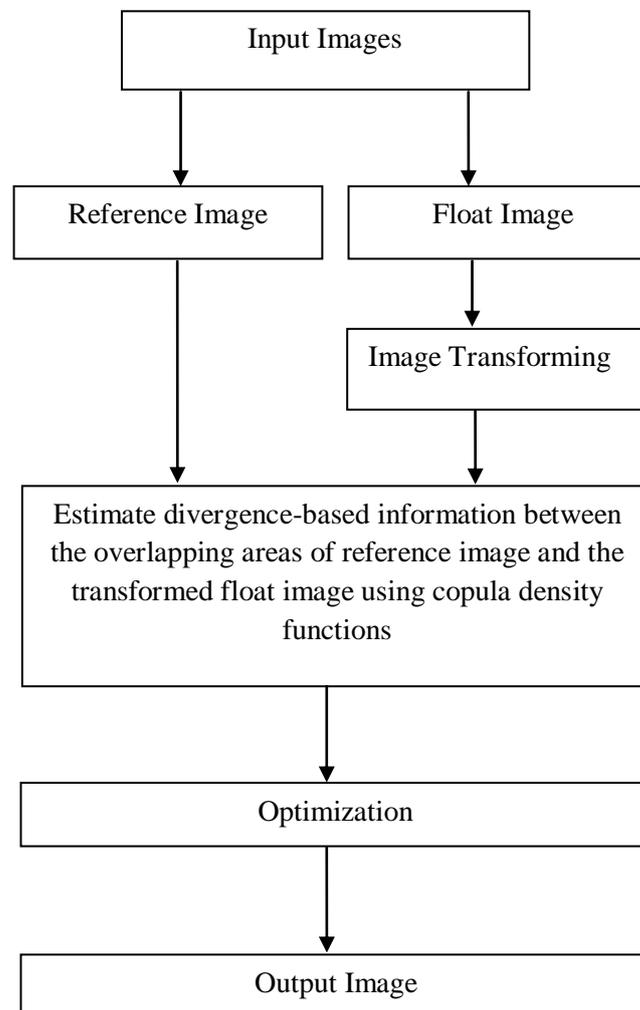
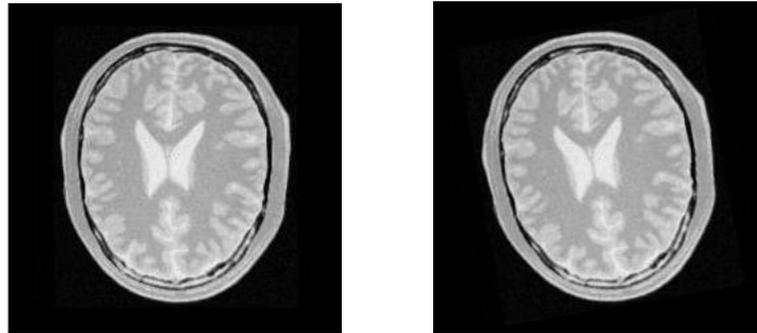


Figure 4.3: Flow Chart of Image Registration



(a) Reference image

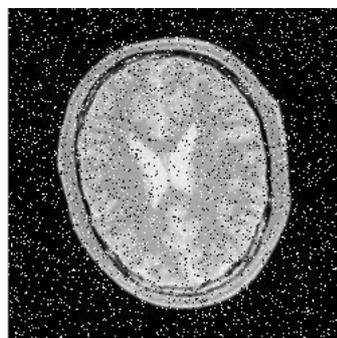
(b): Float image

Figure 4.4 Image Registration of synthetic MRI image

Method	centre on x axis	centre on y axis	Rotation degree
Gaussian	125.17	125.05	-10.00
Joint histogram	124.45	125.12	-10.08
Copulas	125.08	125.05	-9.98

Gravity centre of float image = [125.26; 124.91]

Table 4.1: Results of synthetic MRI image registration without noise

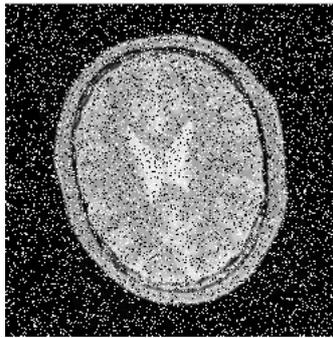


(c): Float image with 10% salt and pepper noise

Method	centre on x axis	centre on y axis	Rotation degree
Gaussian	125.23	125.19	-10.17
Joint histogram	123.97	125.06	-10.08
Copulas	124.85	125.22	-9.90

Gravity centre of float image = [124.7795; 124.8561]

Table 4.2: Results of synthetic MRI image registration with 10% salt and pepper noise

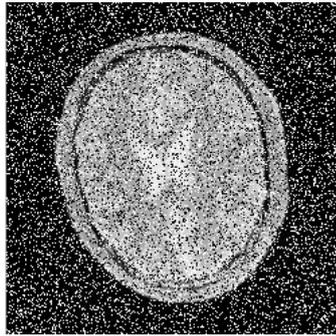


(d): Float image with 20% salt and pepper noise

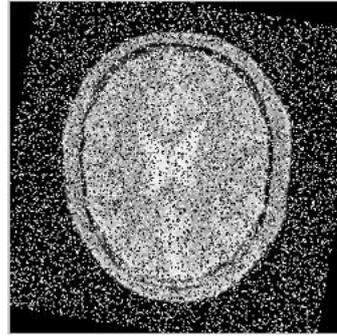
Method	centre on x axis	centre on y axis	Rotation degree
Gaussian	125.32	125.18	-10.21
Joint histogram	124.32	124.43	-10.32
Copulas	125.10	124.86	-10.05

Gravity centre of float image = [125.41; 125.00]

Table 4.3: Results of synthetic MRI image registration with 20% salt and pepper noise



(e): Float image with 30% salt and pepper noise



(f): Registered Image

Method	centre on x axis	centre on y axis	Rotation degree
Gaussian	124.81	125.07	-10.17
Joint histogram	123.75	124.65	-10.33
Copulas	124.77	125.02	-9.94

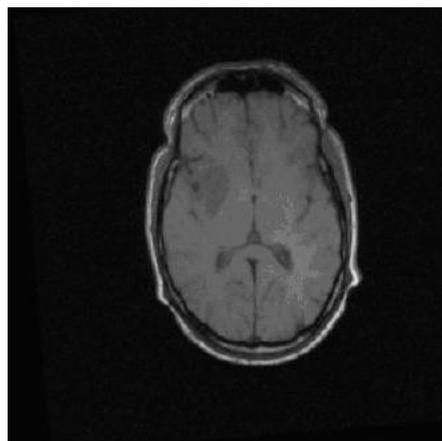
Gravity centre of float image = [125.12; 125.23]

Table 4.4: Results of image registration with 30% salt and pepper noise

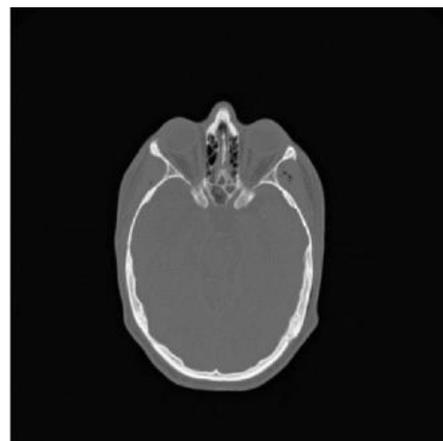
The registration results are shown in Table 4.1, Table 4.2, Table 4.3 and Table 4.4. As may be seen from these tables, the copula measure offers acceptable results for image registration since the results are close to the original values applied. Furthermore, with increasing level of the added noise, the copula based measure provides better results than Gaussian assumption based and histogram based mutual information.

4.3.5 Medical image (CT-MRI) registration

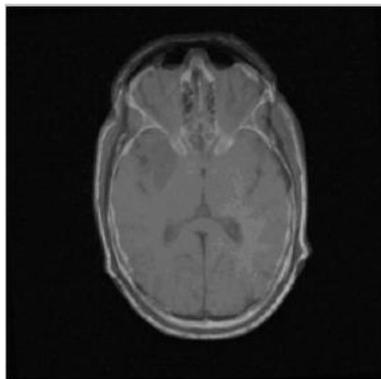
The image shown in the Figure 4.5 (a) is a brain MRI image with size 328x328 pixels and Figure 4.5 (b) is a Computed Tomography (CT) brain image with size 656x656



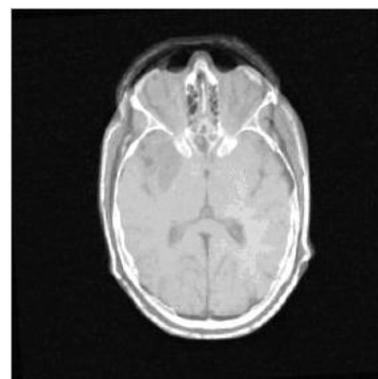
(a): Reference image (MRI)



(b): Float image (CT)



(c): Registered CT overlaid by semi-transparent MRI



(d): Sum of each corresponding pixel intensity of registered CT and MRI

Figure 4.5: Image registration of CT-MRI images

pixels for the same person and can be found online (www.itk.org). Firstly, the size of CT image is adjusted to be the same as the MRI image by subsampling. Similar to the approach taken for the synthetic image registration, the centre of gravity, circle neighbour and Bicubic interpolation are applied. The image registration result is shown in Figure 4.5 (c) where the registered CT image is overlaid on the semi-transparent MRI image. To improve the visual effects, the pixel intensity sum of the registered CT image and MRI image is illustrated in Figure 4.5 (d). The parameters results are given in Table 4.5.

Method	centre on x axis	centre on y axis	Rotation degree
Gaussian	152.36	179.92	3.13
Joint Histogram	152.74	182.84	3.07
Copulas-based	152.16	180.01	3.44

Table 4.5: Result of CT-MRI image registration

4.3.6 Visible light image and thermal image registration

In this section, the registration techniques are applied to three real datasets of visible light images and thermal images provided by Thales. In Figure 4.6, 4.7 and 4.8, (a) and (b) represent visible light images and thermal images respectively while (c) and (e) are the registered images by using joint histogram method and copula method respectively, (d) and (f) are registered thermal images overlaid on the semi-transparent visible light images. The registration technique is based on the following approach. A reference point was firstly selected in visible light image, and then searches the corresponding point in the thermal image. This corresponding point in the thermal image should be located at the same position as the reference point of visible light image. This is achieved by adjusting the rescaling parameter on x and y plane. So, the transforming parameters include the rescaling parameters on x and y plane and the coordinates of corresponding point in the thermal image after selecting the reference point in reference image. The initial parameters are selected as follows: $[x\text{-coordinate of corresponding point; } y\text{-coordinate of corresponding point; rescaling parameter on } x \text{ plane; rescaling parameter on } y \text{ plane}] = [340; 217; 1.4; 1.4]$. The coordinates of reference point of visible light image and corresponding point of thermal image and rescaling parameters on x and y are listed in Table 4.6, 4.7 and 4.8 for Figure 4.6, 4.7 and 4.8 respectively.

Method	Corresponding point	Rescaling on x	Rescaling on y
Joint histogram	339.17; 219.39	1.48	1.40
Copulas	339.60; 219.93	1.45	1.44

The coordinates of reference point of visible light image is [370; 64].

Table 4.6: Results of visible light/thermal image registration for Figure 4.6

Method	Corresponding point	Rescaling on x	Rescaling on y
Joint histogram	339.58; 221.12	1.45	1.43
Copulas	339.31; 220.89	1.44	1.43

The coordinates of reference point of visible light image is [370; 65].

Table 4.7: Results of visible light-thermal image registration for Figure 4.7

	Corresponding point	Rescaling on x	Rescaling on y
Joint histogram	341.99; 219.37	1.46	1.41
Copulas	342.03; 219.84	1.44	1.44

The coordinates of reference point of visible light image is [374; 61].

Table 4.8: Results of visible light/thermal image registration for Figure 4.8

It may be seen that the results of copula method shown in (d) of Figure 4.6, 4.7 and 4.8 are better than joint histogram method shown in (f) of Figure 4.6, 4.7 and 4.8 by observing the distinct objects such as roads and vehicles in images. It can be clearly found that the vehicles are registered in (f) by copula based measure, but they are not registered accurately by joint histogram based measure in (d).

In this experiment for visible light – thermal image registration, 20 image pairs were tested for registration, and in all cases the copulas based method always offer better result than joint histogram based method. The figures given include representative results from this test. Besides, it is worth noting that the same image registration parameters are obtained by Csiszar, Renyi-like, modified Bregman and modified Burbea-Rao divergence based information using Clayton copula, for the synthetic MRI, real image (CT-MRI) and visible light-thermal images registration.

4.4 Copulas based performance evaluation of image fusion

The objective of image fusion is to integrate the complementary information from input images, so that the fused image is more informative and suitable for visual perception. Based on the information theory, the performance evaluation is usually achieved by the estimation of the amount of information obtained from the individual input images. The common evaluation methods of information theory are Fusion Factor (FF) and Fusion Symmetry (FS) [Stathaki, 2008].

4.4.1 Fusion Factor

Fusion Factor (FF) simply sums the information between fused image and input images [Stathaki, 2008]. Considering X and Y as two input images, and F as the fused image, then the Fusion Factor has been defined as [Cvejic et al., 2006]:

$$FF = I_{FX}(F, X) + I_{FY}(F, Y) \quad (4.4)$$

where I_{FX} represents the information between fused image and image X and I can be mutual information, Tsallis divergence based information and Renyi divergence based information. Higher value of FF means better performance of image fusion. However, a significant disadvantage of FF is that, FF cannot indicate whether the input images are fused symmetrically or not. To deal with this problem, the concept of Fusion Symmetry is introduced as follow.

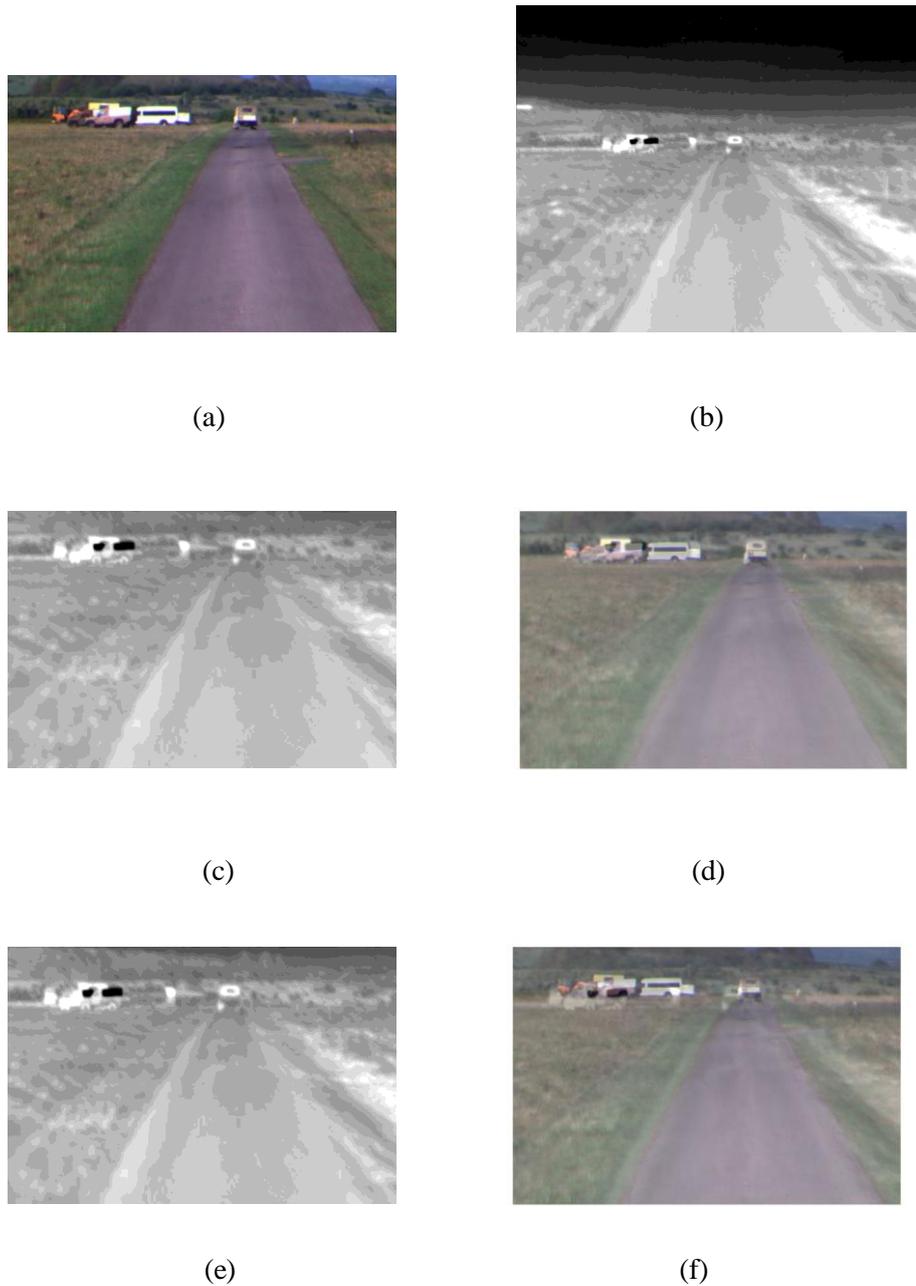


Figure 4.6: The first image registration of visible light/thermal image

(a): Visible light image, (b): Thermal image, (c): Registered thermal image using copula method, (d): Registered thermal image using copulas overlaid on the semi-transparent visible light image, (e): Registered thermal image using joint histogram method, (f): Registered thermal image using joint histogram method overlaid on semi-transparent visible light image.

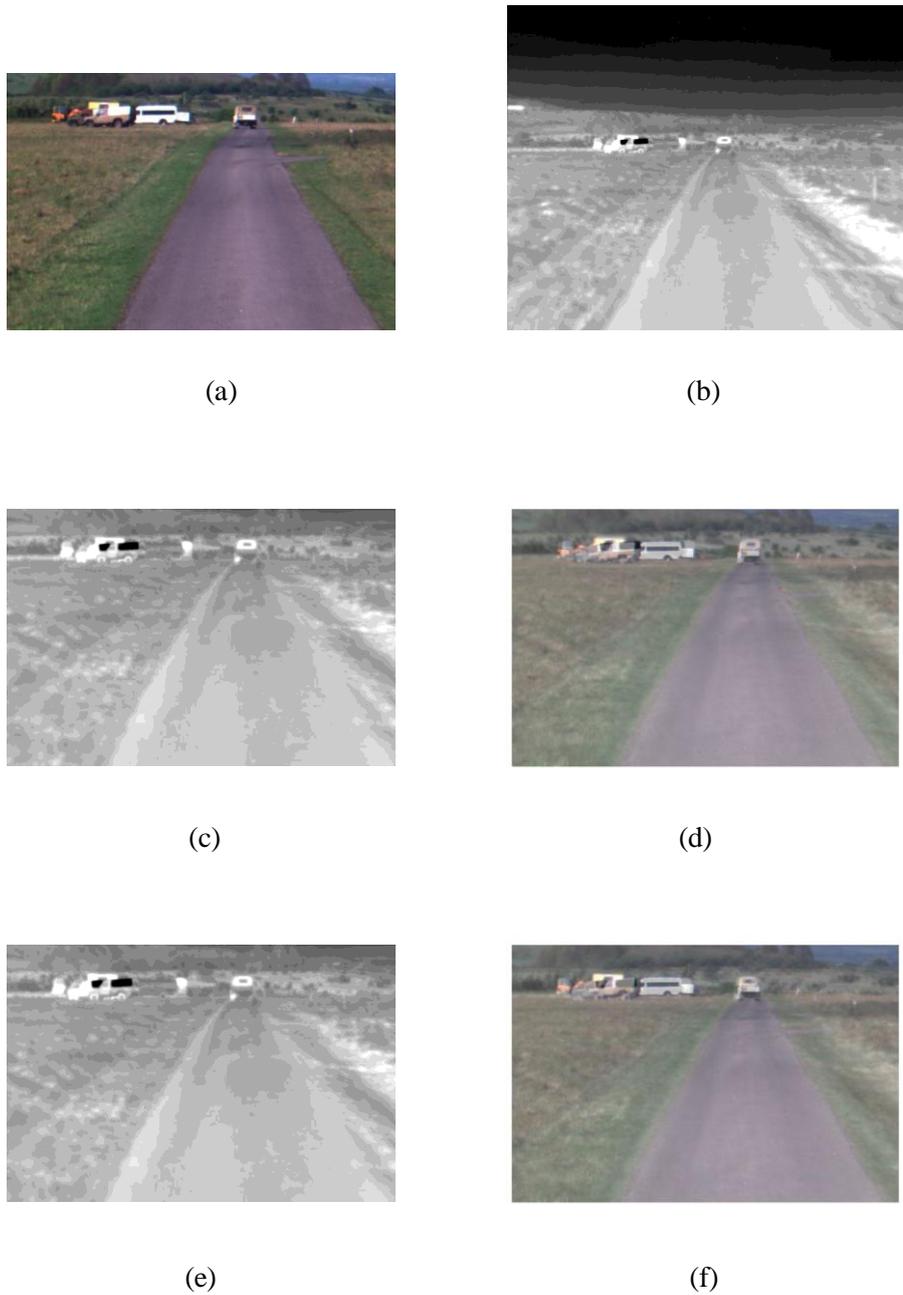


Figure 4.7: The second image registration of visible light/thermal image

(a): Visible light image, (b): Thermal image, (c): Registered thermal image using copula method, (d): Registered thermal image using copulas overlaid on the semi-transparent visible light image, (e): Registered thermal image using joint histogram method, (f): Registered thermal image using joint histogram method overlaid on semi-transparent visible light image.

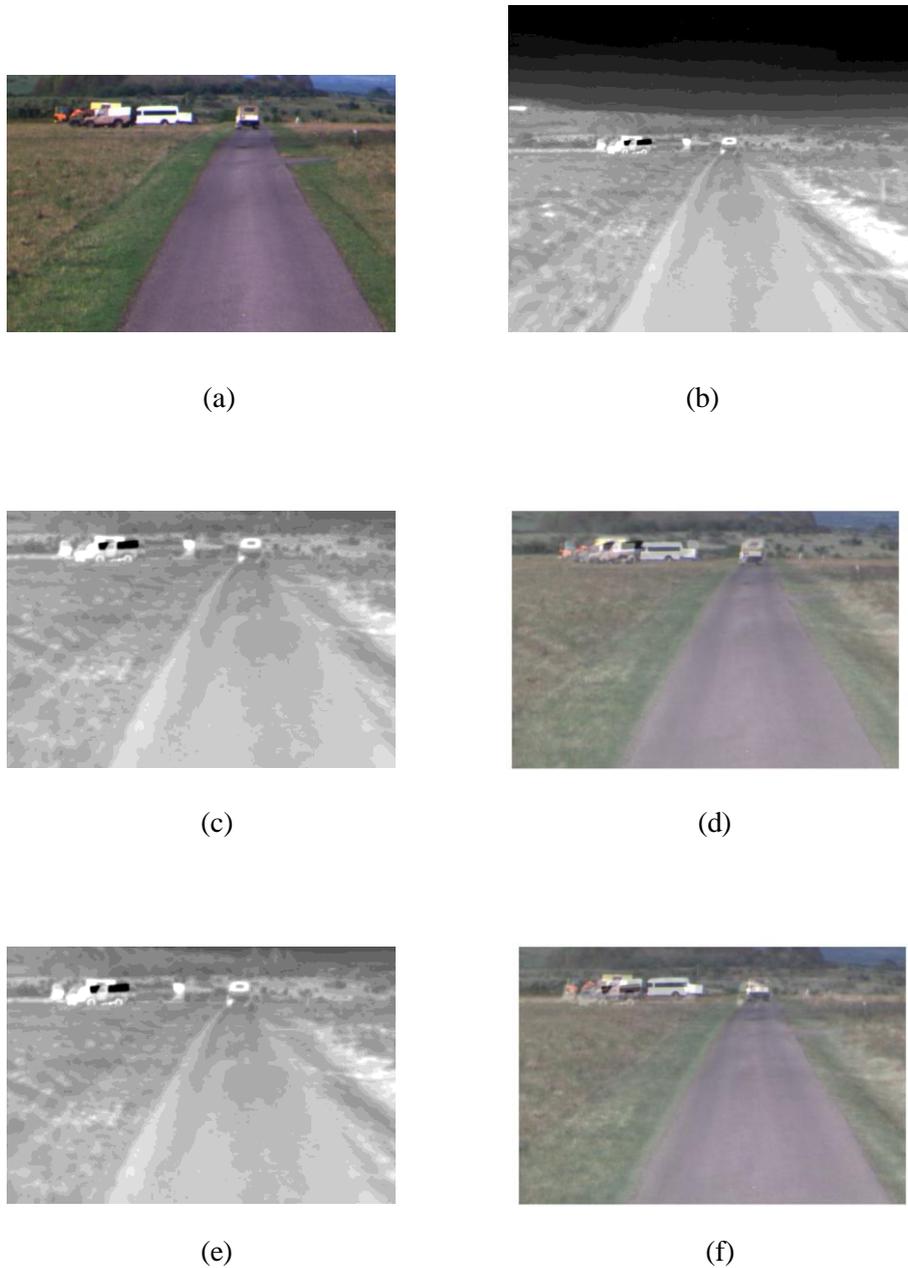


Figure 4.8: The third image registration of visible light/thermal image

(a): Visible light image, (b): Thermal image, (c): Registered thermal image using copula method, (d): Registered thermal image using copulas overlaid on the semi-transparent visible light image, (e): Registered thermal image using joint histogram method, (f): Registered thermal image using joint histogram method overlaid on semi-transparent visible light image.

4.4.2 Fusion Symmetry (*FS*)

The Fusion Symmetry has been defined as: [Stathaki, 2008]

$$FS = abs\left(\frac{I_{FX}(F, X)}{I_{FX}(F, X) + I_{FY}(F, Y)} - 0.5\right) \quad (4.5)$$

where I_{FX} represents the information between fused image and image X ; and the information I can be mutual information or Tsallis and Renyi divergence based information.

FF considers the factor whether the input images are fused symmetrically and it considers that the image fused symmetrically performs better than the fused image simply with high FF . The smaller the FS , better the performance of image fusion.

4.4.3 Divergence based information using copulas for Gaussian distributed data

In the following experiment, dataset are randomly generated with standard bivariate Gaussian distribution by using different Pearson correlations from 0 to 1. The result of mutual information using copulas and Gaussian assumption, Tsallis and Renyi divergence based information are compared. Tsallis and Renyi divergence based information with parameters that equal to 0.5 and 1.5 respectively have been computed using a Gaussian copula, and the copula parameters estimated by using the Canonical Maximum Likelihood (CML) technique [Cherubini et al., 2004].

The results are given in Figure 4.9. It may be observed that, the result of copula based mutual information is very close to the Gaussian assumption based mutual information for the Gaussian distributed data. Moreover, the parameters of Tsallis and Renyi based divergence can be adjusted so that they may offer better ability to control the measurement sensitivity, and hence better image fusion accuracy than the conventional Kullback-Leibler divergence.

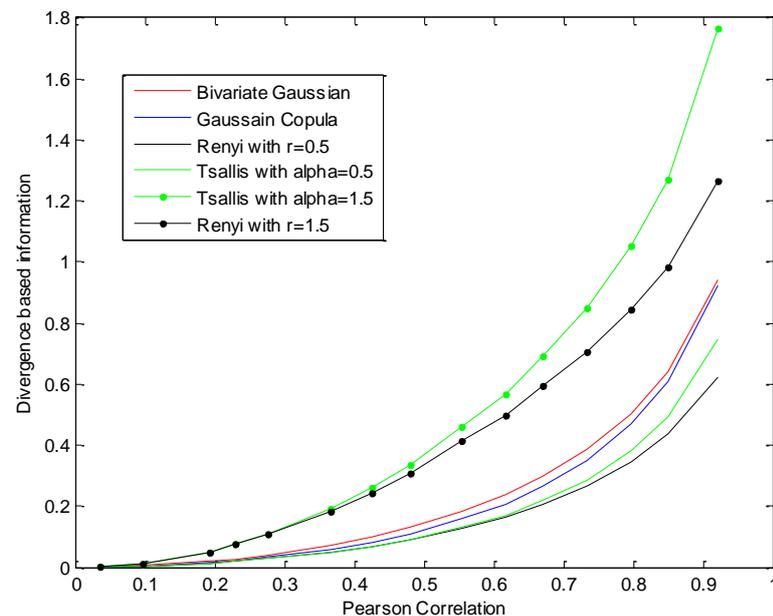


Figure 4.9: Mutual information and Tsallis, Renyi divergence based information using Gaussian copula

4.4.4 Performance evaluation of image fusion using copulas

Two images obtained at the same scene are shown in Figure 4.10 (a) and (b) where (a) is an infrared image and (b) is visible light image, they are available online at: (www.imagefusion.org). Several algorithms including the Simple Average (AVER), PCA, GP, LP, RP and the DWT have been applied for the fusion of these two images, and the results are given in Figure 4.10. To evaluate these methods, firstly, the mutual information based performance measure of FF shown in Eq. (4.4) was computed for all these algorithms. The results indicate the PCA method performs the best, since it has the highest MI_{FXY} value, while the other methods obtained approximately similar MI_{FXY} values. However there is a dichotomy in these observations, as the PCA method

Chapter 4: Copulas for Image Registration and Evaluation of Image Fusion Algorithms

is the worst performing since fused image by PCA method looks very similar to visible image but it is very ‘distant’ from infrared image. Note that there is a human in the visible image, but is not in the fused image at all.

Sine PCA fused image is very close to the visible image, so that very high mutual information is found between PCA fused image and the visible image. Although PCA fused image is ‘very distant’ from infrared image so that the mutual information between PCA fused image and the infrared is very low, mutual information is always equal or greater than 0. So the PCA method is mistakenly considered as the best algorithm. It shows that FF measure cannot indicate whether the images are fused symmetrically. To avoid this type of error, the Fusion Symmetry is used to deal with this problem.

The FF based mutual information and Renyi divergence based information and FS based mutual information, Renyi divergence and Tsallis divergence based information for all the mentioned fusion algorithm are summarized in Table 4.9. The algorithm with higher FF value or lower FS value is considered as better algorithm of image fusion. Gaussian copula was applied and CML method used to estimate the copula parameter. It may be found that FS measure is much better than the FF measure which simple sums the information between fused image and infrared image, visible image respectively.

The next step is to compare the methods based on mutual information, Tsallis and Renyi divergence based information. Since the fusion symmetric measure is obviously better than fusion factor, hence only the fusion symmetry measure is considered. Based on the criterion: the smaller FS , better the performance of image fusion. All of image fusion algorithms mentioned in this chapter can be ranked by using FS measure as:

Mutual information:

LP>AVER>DWT>GP>RP>PCA

Tsallis divergence based information with parameter $\alpha=3$.

LP>AVER>DWT>GP>RP>PCA

Renyi divergence based information with parameter $r=3$.

LP>AVER>DWT>GP>RP>PCA

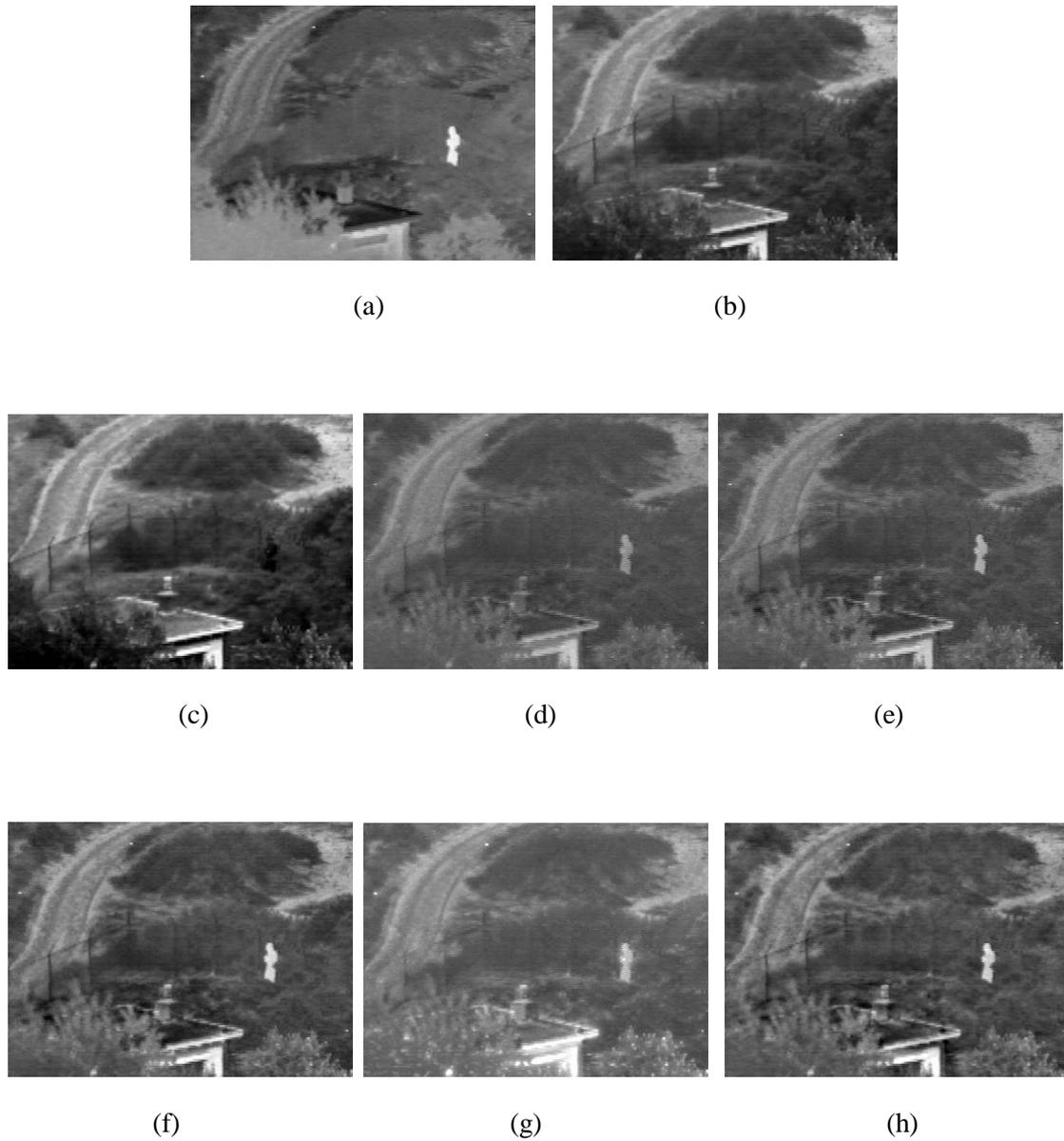


Figure 4.10: Fused images

(a): Infrared image. (b): Visible image. (c): Fused image by PCA. (d): Fused image by averaging. (e): Fused image by gradient pyramid. (f): Fused image by Laplace pyramid. (g): Fused image by Ratio pyramid. (h): Fused image by discrete wavelet transform.

	PCA	Aver	GP	LP	RP	DWT
MI(FF)	1.9751	0.374	0.3557	0.3377	0.2925	0.3062
Renyi(FF)	2.6282	0.8591	0.8250	0.7940	0.6914	0.7291
MI(FS)	0.4755	0.0890	0.1179	0.0683	0.2124	0.1146
Tsallis(FS)	0.4979	0.1078	0.1409	0.0816	0.2412	0.1338
Renyi(FS)	0.4497	0.0731	0.0978	0.0567	0.1834	0.0965

Table 4.9: Performance evaluation of image fusion

It may be observed that the performance of these three information based measures are exactly the same, and is also consistent with the rankings observed. 3 other experiments were carried out for performance evaluations of image fusion, and all of these experiments yielded the same results for mutual information, Tsallis and Renyi divergence based information with *FS* method, and the results were also consistent with the rankings observed. It should be noted that the significant advantage of Tsallis and Renyi method is that they can adjust the associated parameters to obtain better discrimination. For example, in the method of mutual information by *FS* measure, $DWT=0.1146$ is very close to $GP=0.1179$. If Tsallis method is used, and the parameter is adjusted to $\alpha=3$, the results obtained are: $DWT=0.1338$ and $GP=0.1409$. Here the difference between DWT and GP measures become clearer. This characteristic is useful for the situations when the very similar fused results are obtained by different algorithms, and the performance of image fusion is to be evaluated.

4.5 Conclusion

In this chapter, the divergence-based information by using copulas is applied for image registration. To reduce the computation complexity and keep the robustness, some techniques such as centre of gravity, circle neighbours, Bicubic interpolation are applied to improve image registration. The results are compared with the registration

Chapter 4: Copulas for Image Registration and Evaluation of Image Fusion Algorithms

methods that utilize the Gaussian assumption and joint histogram based mutual information. It is shown that the copulas-based method offer the better results especially for noisy and real images between visible light images and thermal images. Furthermore, the performance evaluation of image fusion using copulas has been presented. Gaussian copula density functions have been studied to estimate the mutual information, Tsallis and Renyi divergence based information, and their performance for image fusion is assessed, based on the fusion factor and fusion symmetry measures. Experiment shows that *FS* measure is much better than *FF* measure and the Tsallis divergence offers improved ability to discriminate by adjusting its parameter. An approach to choose the optimal values of the parameter will be researched in the future. The results of experiment also show that the copula density as an alternative and robust way can deal with any marginal distributions, to calculate the mutual information, and the Tsallis and Renyi divergence based information for the performance evaluation of image fusion.

Chapter 5

Band Selection for Hyperspectral Images

Summary

The previous chapter was concerned with the application of image registration which is achieved by maximizing the divergence based information between the overlapping area of reference image and transformed float image using copulas. In this chapter, copulas based information is applied to select the appropriate band images for hyperspectral images by choosing the band images that have higher copulas based mutual information with reference image. The result of experiment on Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) data shows copula based method offer more reasonable band selection results than conventional methods such as Gaussian assumption and joint histogram based mutual information.

5.1 Introduction

Hyperspectral images are taken by the use of specialist equipment that includes hundreds of different detectors, each within a wavelength sensitive range called spectral bandwidth, for example, the Airborne Visible/Infrared Imaging Spectrometer (AVIRIS) instrument, contains 224 different detectors which cover the entire wavelength range from 380nm to 2500nm, with approximately 10 nm spectral bandwidth. The detail can be found online at: (<http://aviris.jpl.nasa.gov>). Hyperspectral image contains much richer information and provide better discrimination ability than visible light images and are able to distinguish between objects which may have the same external appearance but with different interiors. As an illustration of hyperspectral imaging, mantis shrimp can see not only visible light images but also images from the ultraviolet to infrared light. The hyperspectral capabilities enable mantis shrimp to recognize different types of coral or prey which

Chapter 5: Band Selection for Hyperspectral Images

may appear as the same colour to the human eyes. Thus hyperspectral imaging offers a richer seam of image processing for the analysis and detection of image features.

Hundreds of bands imply high-dimensional data which requires huge storage space and transmission bandwidth. Moreover, the performance of hyperspectral image processing for functions such as image classification is strongly affected by the dimensionality, and high-dimensional data it is difficult to process [Guo et al., 2006]. Furthermore, it is likely that redundant information exists and only parts of the data can offer high discrimination, so that band selection between bands from the available high-dimensional data is usually necessary.

The objective of band selection is to remove the bands which contain little or no discriminatory information. In the past, many criteria have been applied for the band selection of hyperspectral images. The common methods include distance measures such as Bhattacharyya distance and Matusita distance, information-theoretic approaches such as mutual information and all kinds of divergence based information and eigen analysis measure such as Principal Components analysis (PCA) [Chang et al., 1999]. Bhattacharyya and Matusita distance [Ifarraguerri & Prairie 2004] are the special case of Renyi-like divergence and Csiszar divergence respectively which measures the similarity of two probability distributions and have been defined in Table 3.4 and Table 3.2 of Chapter 3 respectively. PCA [Pu & Gong' 2000] is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance of any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate. PCA has the distinction of being the optimal linear transformation that provides a subspace that has the largest variance. The components with low eigenvalues are neglected to reduce the dimensions. Specifically, suppose we have a matrix $X = [X_1, X_2, \dots, X_n]^T$, PCA looks for a linear transform of X into Y which satisfies $Y = A(X - \mu_x)$ where μ_x is the mean value of X. The algorithm using the covariance method can be achieved by the following steps:

Step 1: Input matrix data.

Step 2: Subtract the mean to generate a new data.

Step 3: Calculate the covariance matrix of the new data.

Chapter 5: Band Selection for Hyperspectral Images

Step 4: Calculate the eigenvectors and eigenvalues of the covariance matrix.

Step 5: Choose components and forming a feature vector. Once the eigenvector are found, the next step is to rank them by eigenvalue from high to low and then ignore the components which have lower eigenvalues. This does not lose much information.

Step 6: Using eigenvectors chosen in step 5 multiply the new data generated in Step 2.

In this chapter, we focus on the mutual information method which offers a measure of the dependence between random variables. This can be achieved by selecting specific bands for the analysis of hyperspectral images by comparing mutual information values between each of the band images and the reference map. The well known definition for mutual information between two random variables X and Y is given as:

$$MI(X, Y) = \iint f_{XY}(x, y) \log \frac{f_{XY}(x, y)}{f_X(x)f_Y(y)} dx dy \quad (5.1)$$

where $f_{XY}(x, y)$ is the joint probability density function of the variables (X, Y) , $f_X(x)$ and $f_Y(y)$ are the marginal densities of variables X and Y respectively.

5.2 Current techniques and problem description

Comparing to the PCA method, mutual information based method requires a reference map (Ground Truth) which may not always be available. A compensatory approach was proposed in [Guo et al., 2006] for the estimation of the reference map that employs a priori knowledge of spectral signature which significantly reduces the reliance of reference map. Spectral signatures are the specific combination of reflected and absorbed electromagnetic radiation at varying wavelengths, the approximate wavelength range for certain material, even for very similar materials can be identified since different material has the different reflectance in the same wavelength range so that it may uniquely identify as an object. The spectral signature has been collected in the spectral signature library such as USGS Digital Spectral Libraries which can be found online at: (<http://speclab.cr.usgs.gov>). The reference

Chapter 5: Band Selection for Hyperspectral Images

map can be synthesized by averaging the bands which have the significant different reflectance and these bands are called key bands.

The shortcomings of Gaussian assumption and joint histogram have been pointed out in Section 3.2. Copulas offer an alternative parametric technique for mutual information within the observed data and copula based mutual information can be written as:

$$MI(x, y) = \iint_{[0,1]^2} c(u, v) \log c(u, v) du dv \quad (5.2)$$

Furthermore, the rejection bandwidth and setting complementary threshold method is applied to remove the redundancy bands between neighbouring bands. This method is based on whether significant changes of mutual information are detected in the neighbouring bands. If there is no significant change of mutual information, it means that the redundancy information between neighbours, and then these band images can be neglected, even if they have higher mutual information with the reference image.

5.3 Application to hyperspectral images

Experiments were conducted on the AVIRIS data: 92AV3C hyperspectral image dataset collected over a test site called Indian Pine in north-western Indiana, and is available online at: (<ftp://ft.enc.purdue.edu/biehl/MultiSpec>). After removing four null data, 220 bands are available in the 92AV3C dataset. The size of each image is 145x145 pixels. The prior knowledge based on the spectral signature library such as USGS Digital Spectral Libraries can be used to estimate the reference image to reduce the reliance on a ground truth map. From the libraries, the approximate wavelength range for certain material is identified even for very similar materials. To choose the appropriate group of bands which have the greatest reflectance diversity between the interested material and these bands are used to estimate the reference map by averaging the corresponding band images as [Guo et al., 2006].

$$R = \frac{1}{N} \sum_{i=1}^N I_i \quad (5.3)$$

Chapter 5: Band Selection for Hyperspectral Images

where R is the estimated reference image, N is the number of selected band images, i is the index of images and $1 \leq i \leq N$.

Some band images samples are shown in Figures 5.1(a – e). The estimated reference map is shown in Figure 5.1 (f) which is computed by averaging images from the bands 170 to 210 using equation (5.3) [Guo et al., 2006].

After establishing the estimated reference image, the mutual information, between each band image and estimated reference image, is computed using equation (5.2). The Clayton and Gaussian Copula density functions are used to estimate mutual information and results are shown in Figure 5.2.

It can be found from Figure 5.2 that joint histogram method has the worst discrimination ability since the difference of mutual information between reference map and each band image is small. Similarly, the Gaussian assumption based method does not have a good ability to discriminate in the region of bands 10 to 30. The Clayton copula method does not have sufficient discrimination ability for low correlation values, since all the mutual information are approximately equal to 0 between bands 40 and 80. The Gaussian copula method seems to have the best discrimination ability.

Referring to the Figure 5.2, the bands with higher mutual information are patently obvious, and should be retained. However, some redundant information may exist in the neighbouring bands. To avoid inclusion of the redundant information between neighbours, two ancillary selection parameters are adopted here – (i): rejection bandwidth B which stands for a bandwidth measured from the centre of a selected band; (ii): complementary threshold t . These are used to reduce the redundant information between neighbouring bands. The measure is based on whether significant mutual information changes are detected in the neighbouring bands.

The algorithm is described in Appendix 3 [Guo et al., 2006], to select bands, the values for the rejection bandwidth B and complementary threshold t are chosen as 5 and 0.5 respectively and the results are given in Figure 5.3. Here the expected number (expnum) is selected as 15. As can be seen from Figure 5.3, some bands have been removed even if these bands have higher mutual information with respect to the

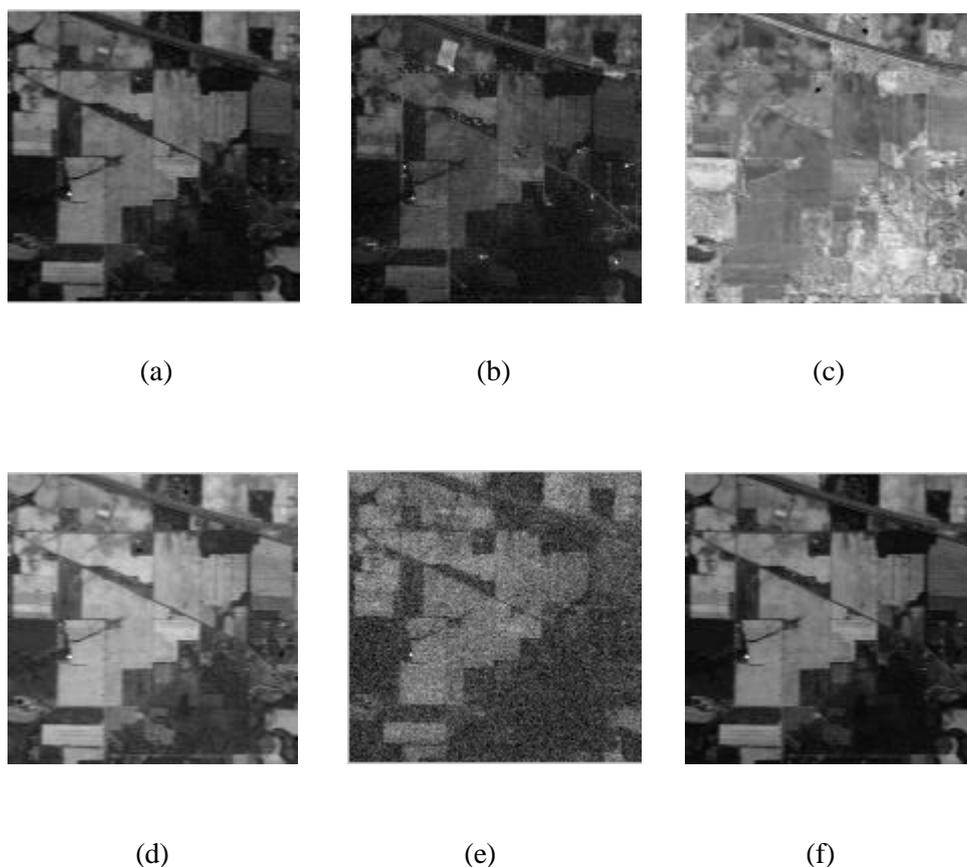


Figure 5.1: Hyperspectral images (a): Band 5 (b): Band 50 (c): Band 125 (d): Band 200 (e): Band 218 (f): Estimated Reference image

estimated reference image. This measure is more reasonable than a setup that uses a simple threshold, since it removes redundant information between neighbouring bands. The final bands selected using the Clayton, Gaussian copula, Gaussian assumption and joint histogram are shown in Table 5.1. It may be seen that the band selection should concentrate between 115 to 145 and 175 to 210. Bands from 5 to 35 may be also considered, when the retention of more bands are required. From the Figure 5.3 and Table 4.1, it is found Clayton and Gaussian copula provide more reasonable results for band selection than Gaussian assumption and joint histogram based mutual information.

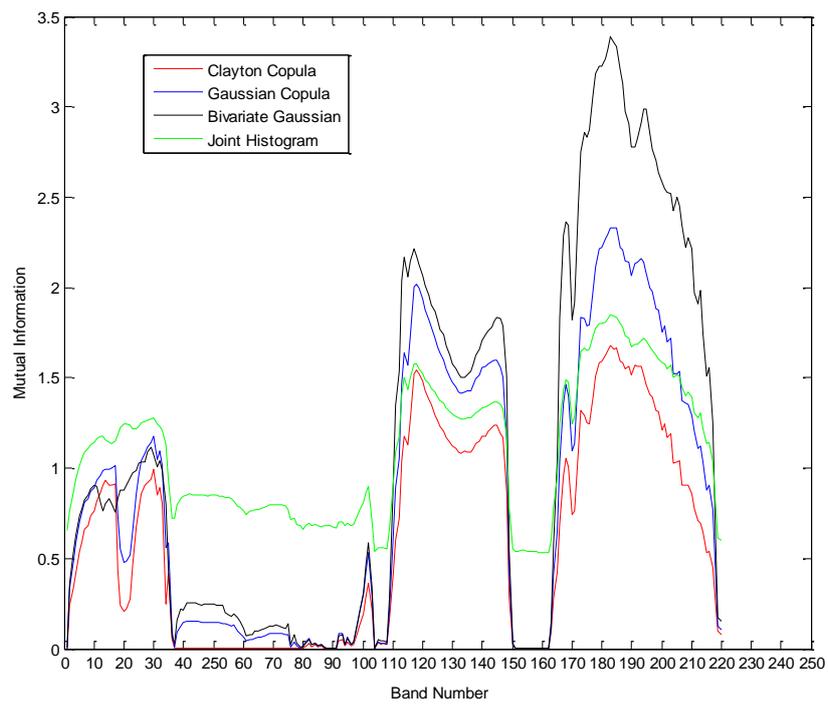


Figure 5.2: Mutual Information between each band image and reference image

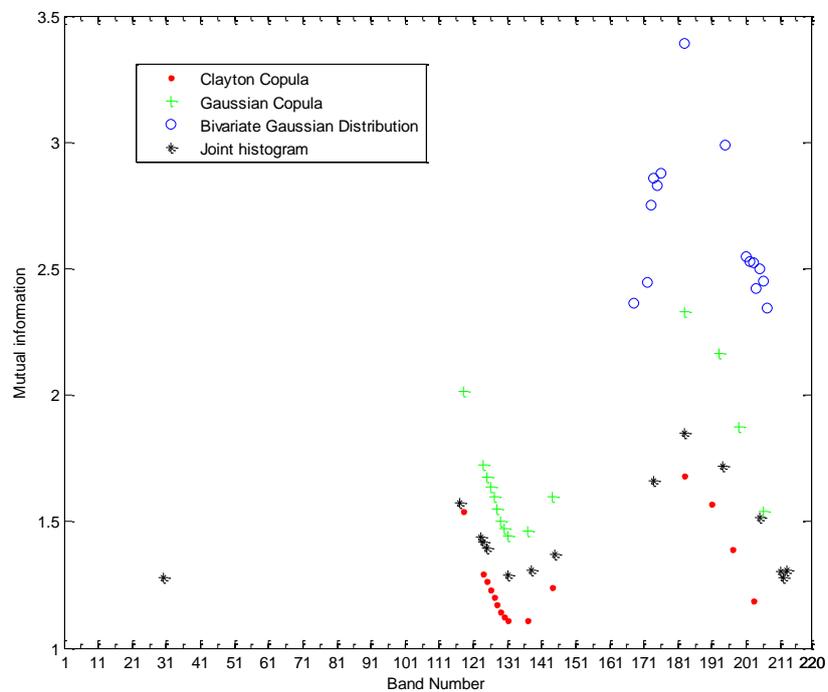


Figure 5.3: Selected band number using rejection bandwidth and complementary measure

Measure	Selected 15 bands														
Clayton Copula	118	124	125	126	127	128	129	130	131	137	144	183	191	197	203
Gaussian Copula	118	124	125	126	127	128	129	130	131	137	144	183	193	199	206
Bivariate Gaussian	168	172	173	174	175	176	183	195	201	202	203	204	205	206	207
Joint Histogram	30	117	123	124	125	131	138	145	174	183	194	205	211	212	213

Table 5.1: 15 Selected Bands

5.4 Conclusion

In this chapter, a band selection approach has been presented for hyperspectral images that exploit copula functions. The Clayton and Gaussian copula have been used to calculate the mutual information between each band image and the estimated reference image based on the prior knowledge of spectral signatures. Experiments on the AVIRIS dataset show that copulas offer an alternative and robust way to calculate the mutual information for band selection of hyperspectral images. The results are compared for methods using the Clayton, Gaussian copula, Gaussian assumption and joint histogram based mutual information, and found Gaussian copula method offers the better results than, Clayton copula, Gaussian assumption and joint histogram based methods.

Chapter 6

Conditional Copula for Image Change Detection

Summary

Conventional change detection techniques such as statistical similarity based measure cannot accurately deal with complicated scenes, e.g. where the objects in two images are very similar or even the same but their pixel intensities and the statistical distribution of pixel intensities can vary due to the external factors such as the use of different sensors or climate changing. In this chapter, conditional copulas are applied to deal with the problem of variability by training the ‘no change’ areas between two images. The results of copula based measure are better than ‘statistical similarity’ and ‘difference’ measure by observing and are validated by Receiver Operating Characteristic (ROC) curve.

6.1 Introduction

The objective of image change detection is to produce a binary map which represents two classes: change and no change for two images usually taken at the same scene but different time. Automated change detection is required in several fields such as remote monitoring [Bruzzone & Prieto, 2002], defence surveillance systems [Stauffer & Grimson, 2000], non-invasive assessment of medical conditions [Wakuya et al., 2007] and so on. In some cases, the sensors with different modalities are employed, for example, sensors may use hyperspectral bands to collect data in remote earth observations or as in medical imaging, different modalities such as MRI, CT and PET (Positron Emission Tomography) are used to monitor patients.

Chapter 6: Conditional Copula for Image Change Detection

In this chapter, the current techniques are first reviewed and the problems of complicated situation are considered where images are taken by the different sensors or where climate changes affect imaging condition. Next, the applications of image change detection are analysed for two cases: (i) Synthetic Aperture Radar (SAR) images: to estimate the scene change before and after earthquake. (ii) CT images: to monitor the calcification in human brain change before and after treatment and conditional copulas are developed for image change detection, early ideas were first proposed by [Mercier et al., 2008]. Finally, the results of copulas-based measure are compared with the conventional measure such as: Pixel-based and Statistical similarity based measures.

6.2 Image change detection techniques review

For two registered images taken at different time, the change detection can be accomplished in two stages: the first stage is to generate the change indicator and the next stage is to determine the optimal threshold for change indicator to achieve the final change detection. Both of these two stages are important for the performance of overall detection. In this chapter, we concentrate on the first stage: generation of change indicator. The algorithms of change detection indicator can be classified into two categories.

(i): Pixel-based measures [Inglada & Mercier, 2007]

(ii): Statistical similarity based measure [Inglada, 2003]

In the Pixel-based measure, the common change indicators are Mean Ratio Detector (MRD) and Difference Detector (DD). MRD is achieved by calculating the ratio of the local mean pixel intensity values for each pixel which is located at the same position within a fixed window of neighbourhood. MRD can be defined as:

$$D_{MRD}(X, Y) = 1 - \min \left\{ \frac{\mu_X}{\mu_Y}, \frac{\mu_Y}{\mu_X} \right\} \quad (6.1)$$

Chapter 6: Conditional Copula for Image Change Detection

where μ_x and μ_y are the mean values of each pixel computed in relation to its neighbourhood for the two images respectively.

The difference detector is obtained by calculating to pixel intensity difference between each pixel, which is at the same position within the two images.

It is known that the probability distribution of pixel intensity usually offers more stable information than pixel intensity themselves [Inglada, 2003]. The measure based on the local pixel probability distribution investigates the distance between two image pixel intensity probability distributions instead of the pixel intensities. The change indicator is achieved by the calculation of the symmetric Kullback-Leibler distance (SKLD). The Kullback-Leibler distance has been defined as [Inglada & Mercier, 2007]:

$$K(Y | X) = \int \log \frac{f_X(x)}{f_Y(x)} f_X(x) dx \quad (6.2)$$

where $f_X(x)$ and $f_Y(y)$ are the probability density functions (pdf) of the random variables X and Y respectively. The Kullback-Leibler distance is asymmetric and the symmetric version of Kullback -Leibler distance (SKLD) can be defined as [Inglada & Mercier, 2007]:

$$D(X, Y) = D(Y, X) = K(Y | X) + K(X | Y) \quad (6.3)$$

The key to Statistical similarity based measure is to estimate the probability density function for the local pixel intensity within a fixed window. Some well-known parametric distributions such as the Gaussian, Rayleigh distribution [White, 1991] have been proposed. Moreover, the Gamma distribution has been applied for Synthetic Aperture Radar (SAR) images since in this case; it is well known that the pixel intensity is distributed according to the Gamma distribution [Chatelain et al., 2007]. A more generalised distribution: Pearson System has been proposed for one dimensional probability density distribution estimation [Inglada, 2003]. The detail of Pearson System is shown in Appendix 4. Some non-parametric methods such as Gram-Charlier, Edgeworth series expansion also have been proposed [Inglada & Mercier, 2007], and the detail is given in Appendix 5. Furthermore, the kernel smoothing density estimation [Bowman & Azzalini, 1997] is an efficient and robust

tool for the estimation of probability density function, and is introduced in Appendix 6.

6.3: Comparison of Pearson system, Gram-Charlie series, Edgeworth series and kernel smoothing technique

The Pearson system has been applied for change detection in [Inglada, 2003], it includes eight probability densities but it is impossible to cover all the probability densities. Gram-Charlie and Edgeworth series have been applied for change detection in [Inglada & Mercier, 2007], but it has been found that the approach works for the probability distribution of image pixel intensity that is not far from the Gaussian distribution. In the following experiments, the Kernel smoothing technique seems a robust measure for the estimation of one dimensional probability density function.

A dataset with size 1500x2 was randomly generated by Gaussian distribution with parameters: mean value = 2.0220 and standard deviation = 4.9524. The results of simulations by using Gram-Charlier and Edgeworth series to estimate the probability density function for this dataset is shown in Figure 6.1 (a). With the size of dataset increases, the results of simulation become better and even as good as Gaussian probability density function itself.

Next, student t distribution is used to test the simulation ability of Gram-Charlie and Edgeworth series. A dataset with size 1500x2 is generated by Student t distribution with freedom degree = 8.0074, that is to say this student t is close to the Gaussian distribution since it has a high freedom degree. The simulation result is shown in Figure 6.1 (b). It may be found that the simulation results are not satisfactory when the freedom degree is low (lower than 7). The simulation result of freedom degree = 6.2413 is shown in the Figure 6.1 (c). That is to say, with the decreasing of freedom, the student t is more and more distant from Gaussian distribution, and the simulation results get worse.

The simulation of the probability density function for the same Student t distributed data with freedom degree = 6.2413 by using kernel smoothing technique is shown in Figure 6.1 (d). It can be seen that the result is much better than Gram-Charlier and Edgeworth series based probability density function estimation.

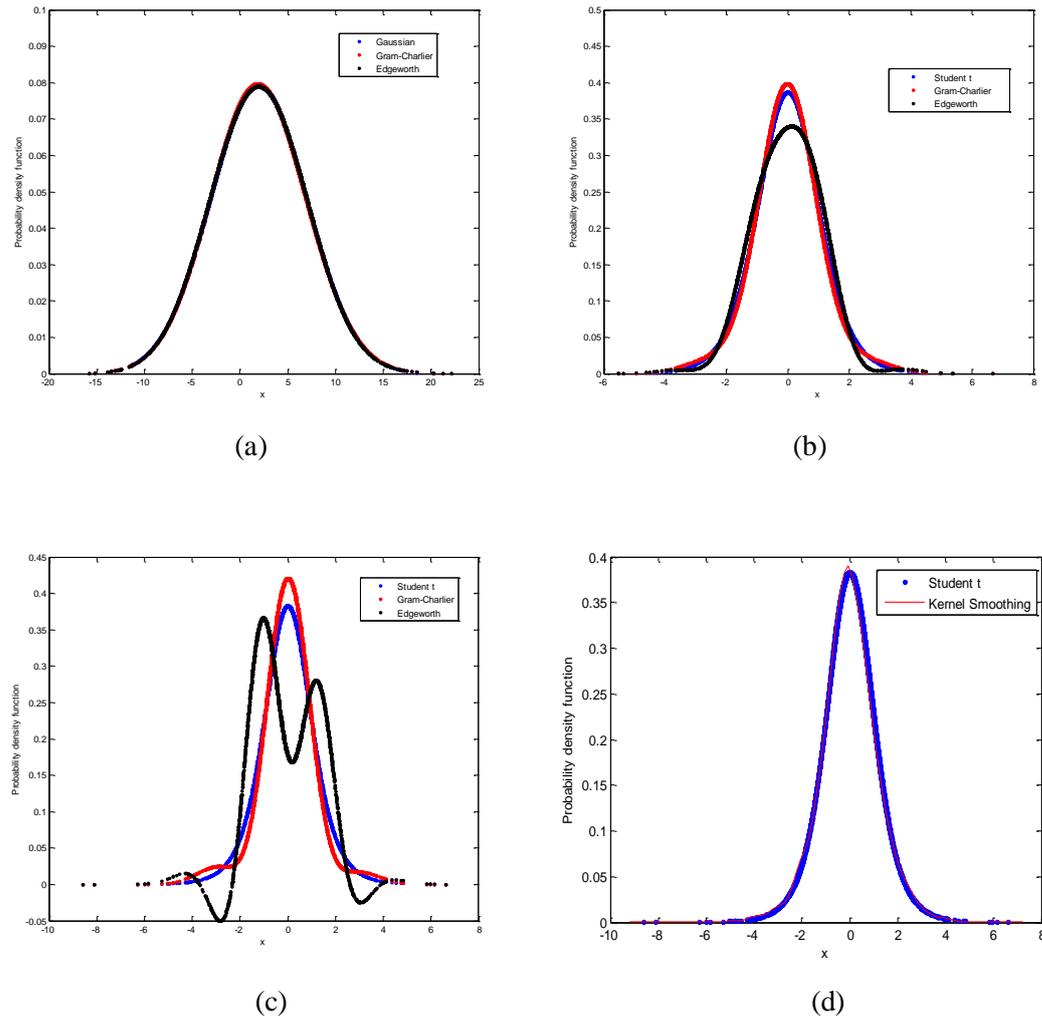


Figure 6.1: Probability density function estimation

6.4 Problems description

The Statistical similarity based measures are usually more reliable than Pixel based measures since the pixel intensity distribution offer more stable information than pixel intensities themselves. However, it has been found that the results are disappointing when the observations are very similar or even same, but the local pixel intensities and their statistical distribution varies remarkably due to the external factors such as the affect of climate changes, or use of different sensors. It leads to wrong change indicators so that the results of change detection are disappointing.

Chapter 6: Conditional Copula for Image Change Detection

An approach to compensate for this variability is to generate new pixel intensities to replace the pixel intensities of the first image by using the dependence between the ‘no change’ areas of the first and second images. This dependence may be non-linear and thanks to the advantage of copulas, this dependence can be modelled by copulas. The generated pixel intensities are that the pixel intensities of the first image should be if it was taken by the same imaging conditions with the second image. Finally, calculate the symmetric Kullback-Leibler distance between the generated pixel intensities and the pixel intensities of the second image to produce the change indicator.

6.5 Applications

The approach to image change detection using conditional copula is illustrated by the flow chart in Figure 6.2. Considering two registered I_1 (the first image) and I_2 (the second image), firstly choose the training area where ‘no change’ occurs in I_1 and I_2 . The ‘no change’ areas means that the objects in this area should be very similar or same, but actually they have different pixel intensities and statistical distribution of pixel intensity due to external factors such as affects of climate changes, or using different sensors.

Secondly, use the copula to model the dependence between the two ‘no change’ areas in I_1 and I_2 . Let $F_X(x)$ represents the pixel intensity cumulative distribution of I_1 . Transform the pixel intensity to uniform variant u by using empirical distribution which has been defined in Eq. (2.33). The corresponding conditional variant v can be obtained by using u and copula parameter by using the concept of conditional copula introduced in Eq. (2.37) as:

$$C_u(v) = \Pr[V \leq v | U = u] = \lim_{\Delta u \rightarrow 0} \frac{C(u + \Delta u, v) - C(u, v)}{\Delta u} = \frac{\partial C(u, v)}{\partial(u)}$$

For example, for Clayton copula, v can be estimated by using u and copula parameters θ as:

$$v = (u^{-\theta} (t^{-\frac{\theta}{\theta+1}} - 1) + 1)^{-\frac{1}{\theta}}$$

Chapter 6: Conditional Copula for Image Change Detection

where t is generated randomly to have the uniform distribution on $[0, 1]$.

The next step is to simulate new pixel intensities by using the cumulative distribution of the ‘no change’ area of the second image $F_2(x_2)$ which is usually different from the distribution of the second image and v using the following mapping as:

$$X' = F_2^{-1}(v) \quad (6.4)$$

Finally, the symmetric version of the Kullback-Leibler distance is calculated between the new simulated pixel intensities and the pixel intensities of the second image, as the final change indicator. The kernel smoothing technique can be used to estimate the probability density function. The flow chart of change detection using copulas is shown in Figure 6.2.

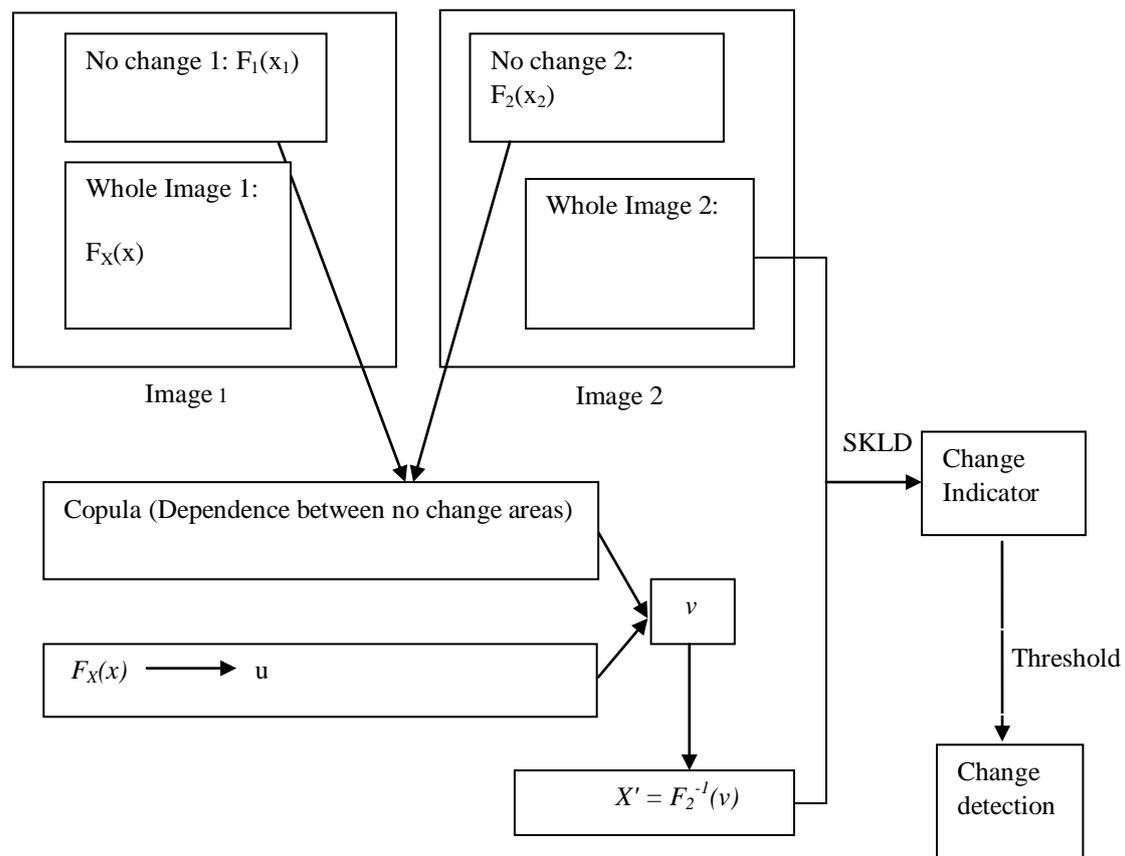


Figure 6.2: Flow chart of change detection using conditional copula

6.5.1 Change detection for remotely sensed images

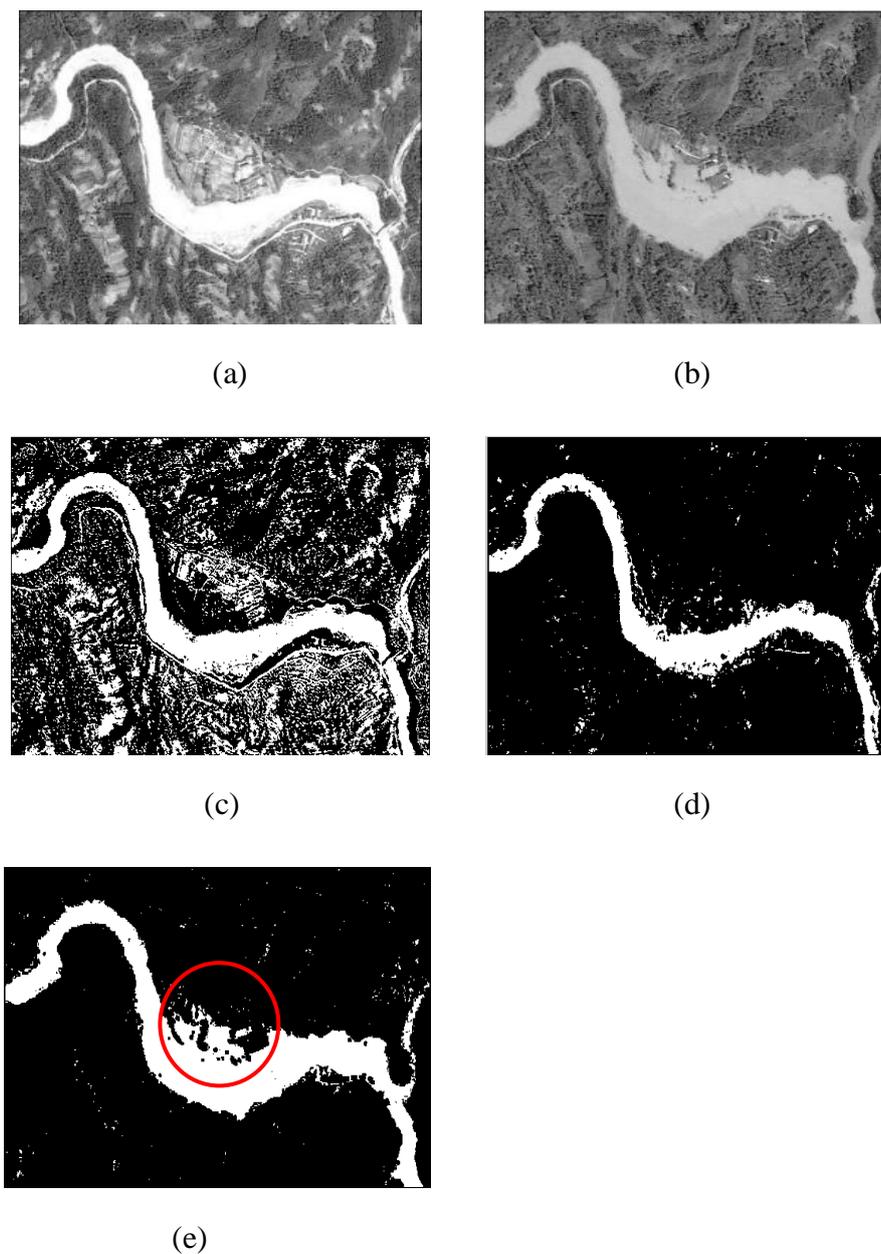


Figure 6.3: Change detection of SAR images of Earthquake using conditional copula

(a): SAR Image taken on 14 May 2006 (around two years before earthquake) (b): SAR image after flooding on 14 May 2008 (two days after earthquake) (c): Change detection results by pixel intensity ‘difference’ based change detector (d) Change detection result by Statistical Similarity based change detector (e) Change detection results using Clayton copula based change detector.

Chapter 6: Conditional Copula for Image Change Detection

Figure 6.3 represents the SAR images of Earthquake at Sichuan, China in May 2008 which can be found online at (<http://earthobservatory.nasa.gov>). After earthquake, the objects that were affected greatly were the river and the residence areas near the river which were flooded, but the objects that were far away from the river were affected slightly or even not affected at all, so we can select the same areas that far away from the river in Figure 6.3 (a) and (b) respectively as ‘no change’ areas.

The next step is to estimate the copula parameter for these two ‘no change’ areas to model the dependence between these two images. These parameters are calculated for Clayton, Frank and Gaussian copula as: 0.1238; 0.8759 and 0.1537 respectively.

In this application, it has been found that the Clayton copula has the minimal Euclidean distance from the empirical copula among Clayton, Frank and Gaussian copula, so that Clayton copula is chosen to applied to model the dependence between the ‘no change’ areas of these two images.

The new pixel intensities can be generated by using conditional copula and the cumulative distribution of the ‘no change’ area of image Figure 6.3 (b) using the Eq. (6.4). Finally, calculate the symmetric version of Kullback-Leibler distance between new generated pixel intensities and the pixel intensities of Figure 6.3(b).

The final result of change detection using Clayton copula is shown in Figure 6.3 (e). Figure 6.3 (c) is the change detection result using the Pixel-based measure of the differences between images Figure 6.3 (a) and (b). Figure 6.3 (d) is the change detection result obtained by the Statistical similarity base measure, obtained by using the symmetric Kullback-Leibler distance as change indicator. The same neighbourhood size 3x3 for each pixel and kernel smoothing technique are applied to estimate the probability density function for both Statistical similarity and copula based measures.

It can be seen that the key flooding areas (shown in the red circle in Figure 6.3 (e)) is missed out by the first and the second detector in Figure 6.3 (c) and (d) respectively but picked by using the copula approach.

To compare the performance of change detection among the ‘difference’, ‘Statistical similarity’ and copula measures, the Receiver Operating Characteristic (ROC) [Fawcett, 2006] is computed. ROC is defined as a plot of the true positive rate (as the

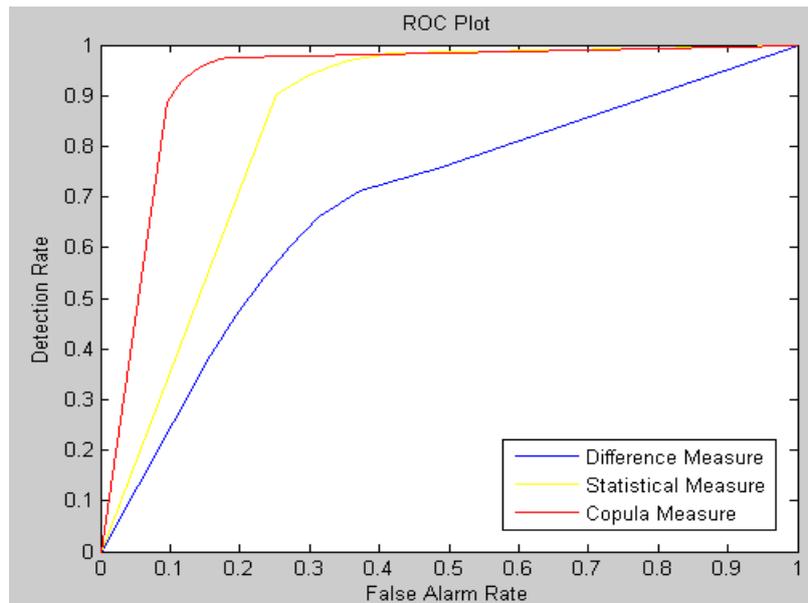


Figure 6.4: ROC plot of performance of image change detection

y coordinate) versus false positive rate (as the x coordinate). ROC offers a measure of the true detection performance against false alarm. The closer the curve is to the left-hand border and the top border of the ROC space, the more accurate the test. Since there is no ground truth available for this change detection, the ‘ground truth’ is manually obtained by the rule: the objects affected are only the whole river and parts of residence area flooded and the objects far away from river was considered as non-affected. The result of ROC test is shown in the Figure 6.4. It can be found the copula based measure provides better result than ‘statistical similarity’ and ‘difference’ measures.

6.5.2 Change detection for medical images

Figure 6.5 represents the CT image of human brain, the areas interested is the calcification area which changed before and after treatment, and the rest area should be very similar or even same, however actually they have the very different looking and statistical distribution of pixel intensity. It is a problem to obtain the satisfactory results by the conventional measures such as ‘difference’ and ‘statistical similarity’

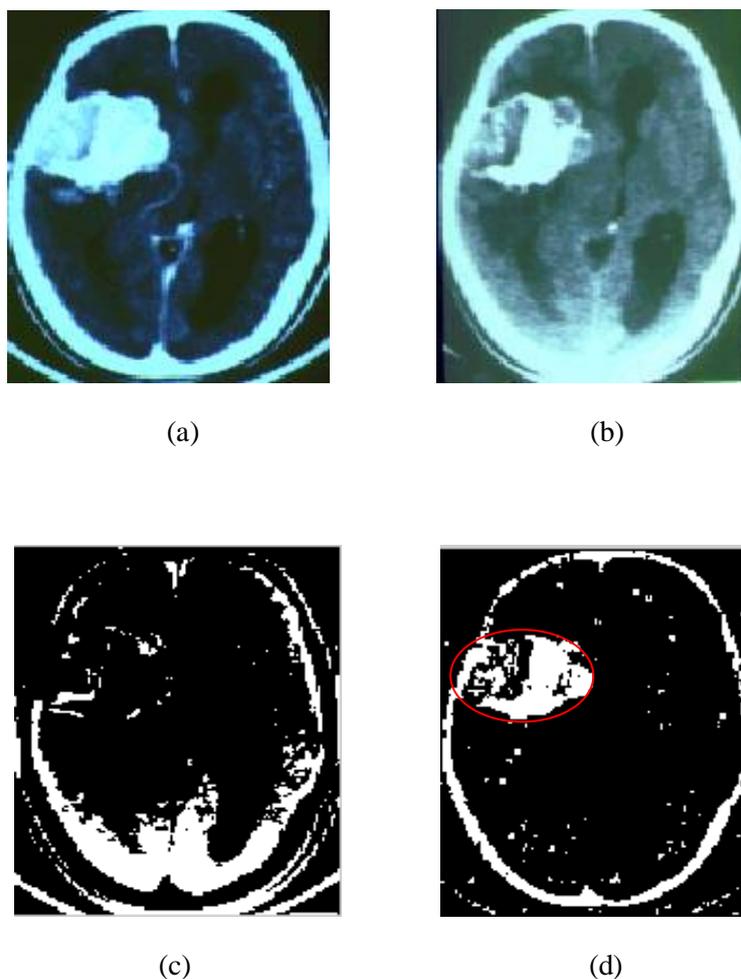


Figure 6.5: Change detection of CT images

(a): Before treatment (b): After treatment; (c): Change detection result of using difference measure. (d): Change detection using Frank copula measure.

measures. Copulas can be used to deal with this problem, firstly, the same areas located in the non-calcification areas of Figure 6.5 (a) and (b) are selected as the ‘no change’ areas, and then use these ‘no change’ areas to estimate the copula parameter. Then use the conditional copula and the cumulative distribution of the ‘no change’ area of Figure 6.5 (b) to simulate new pixel intensities. Finally calculate the symmetric Kullback-Leibler distance between the simulated pixel intensities and the

Chapter 6: Conditional Copula for Image Change Detection

pixel intensities of Figure 6.5 (b) as the change indicator, and the kernel smoothing technique is used to estimate the probability density function.

The result of change detection by using conditional copula is shown in the Figure 6.5 (d). The red circled area in Figure 6.5 (d) indicates the key change picked by the copula approach but missed by the difference measure.

In addition to the above two experiments, three other experiments were carried out. In all cases it was observed that the conditional copula based method always offered better results than the ‘difference’ and ‘statistical similarity’ methods when two images were obtained under different imaging conditions such as for different weather or when different sensors are used.

6.6 Conclusion

In this chapter, image change detection techniques based on conditional copula has been presented for the complicated situation where the objects in two images respectively are very similar or even same, but the pixel intensities their statistical distribution of pixel intensity vary remarkably due to the external factors such as climate changing or the use of different sensors. Here, copulas have been used to model the dependence structure between the ‘no change’ areas of two images, and then use the conditional copula and the cumulative distribution of the ‘no change’ area of the second image to simulate new pixel intensities. The simulated pixel intensities are that the pixel intensities of the first image should be if it was taken by the same imaging conditions with the second image. The final change indicator is obtained by calculating the symmetric version of Kullback-Leibler divergence between the simulated pixel intensities and the pixel intensities of the second image. It is shown that the results of copulas based measures are better than the ‘difference’ and ‘statistical similarity’ measures. Furthermore, the work is validated by plotting the Receiver Operating Characteristic (ROC) curve for the different methods.

Chapter 7

Conclusion and Further Research

7.1 Introduction

The purpose of the work presented in this thesis was to develop novel and original approaches for image processing using the copula functions. The consequential contributions are:

- Estimation of joint cumulative distributions and joint probability density functions from arbitrary marginal distributions using copula and copula density functions.
- The definitions of four categories of generalized divergences which include Csiszar, Renyi-like divergence, modified Bregman and Burbea-Rao divergence based information developed in terms of copula density functions only. These provided the ability to control measurement sensitivity. The modified Bregman divergence is established in terms of the smallest enclosing curve and K-means classification. Computing techniques for these divergence based information have been proposed, and the performance of these divergence based information were compared using Bernoulli distributions.
- An algorithm developed for maximizing the generalized divergence based information using copula density functions only for image registration. Two synthetic images with added noise, CT-MRI images and visible light image-

Chapter 7: Conclusion and Further Research

thermal images were employed to analyse the copula based techniques for image registration.

- Performance evaluation for different image fusion algorithms by generalized divergence such as Tsallis and Renyi divergence based information using copulas for multi-sensor images were proposed.
- A robust band selection method for hyperspectral images using copula based mutual information was developed and spectral signatures were introduced to reduce the reliance on reference image.
- The implementations of conditional copulas for complicated image change detection. The images taken by multiple sensors or under different climate change conditions were analysed. In such cases, the pixel intensities and the statistical distribution of pixel intensity are much different although the observations are very similar or even same. This makes change detection difficult using conventional techniques.

Within this thesis, an introduction to the thesis was proposed in Chapter 1; the chapter provides a brief introduction of image processing, the current problems of image processing and the reasons of using copulas for image processing. The subjects covered in this thesis include image registration, performance evaluation of image fusion, band selection for hyperspectral images and image change detection.

Copula function theory was introduced in Chapter 2 to provide the necessary background to the subject. In Chapter 2, the copula function was firstly introduced as a joint distribution function with uniform marginal distributions according to the Sklar's theorem, and then copula density function was deduced as the quotient of joint probability density function and product of marginal probability density functions.

Next, similar to the definition of joint conditional distributions, the conditional copula was developed, based on exploiting a set of partial differential operations. It was

Chapter 7: Conclusion and Further Research

shown that a marginal distribution can be estimated if another marginal distribution and the associated conditional copula is known, since the copula offers a measure of the dependence between these two marginal distributions.

Further, Kendall's tau, Spearman correlation and Maximum likelihood were introduced to estimate the parameters of copulas.

It was shown that the measures based on the Kendall's tau and Spearman correlation are only suitable for the copulas that have simple expressions linking the copula parameters and Kendall's tau or Spearman correlation. Some expressions for the two measures can be very complicated and some cannot be found. A further disadvantage of Kendall's tau and Spearman correlation is the computing effort involved. In such cases, the maximum likelihood measure should be applied.

In addition, the limitation of conventional dependency tools: the Pearson correlation was analyzed. It is known that the Pearson correlation, as a linear correlation, is only effective for the elliptical distributions. However in the real world, most of joint distributions of images are not necessarily elliptical. Moreover, a further significant limitation of all the known multivariate distribution models, such as multivariate Gaussian, Gamma, Exponential and Rayleigh distributions, is that they require consistent marginal distributions. By contrast, copulas are able to model the dependence which may not be linear and deal with arbitrary marginal distributions.

Experiment results in Chapter 2 showed that the joint density function estimated by copula density function has very similar results to the measure based on the classic joint Gaussian function for Gaussian distributed dataset. Further, using copula based measures, the joint cumulative distribution and probability density functions were also estimated for the datasets which have Gaussian and Student t marginal distributions respectively, thus proving the efficacy of copula techniques, in determining marginals that do not belong to the same family of distributions.

Finally, the dependency ranges of copulas were used for preliminary selection of optimal copulas, since some copulas such as Farlie-Gumbel-Morgenstern (FGM) copula only can model distributions that exist for the narrow bounds of Kendall's tau correlation, in the range $[-2/9, 2/9]$. Note that some copulas have very complicated expressions and are not convenient for applications. Our studies have shown that the

Chapter 7: Conclusion and Further Research

Clayton, Frank and Gaussian copulas are reasonable choices among the known families of copulas, as they are relatively simple expressions and are relatively easy to compute. Finally, empirical copula was introduced to determine the optimal copula from a finite set of copulas, and the candidate copula with the minimal Euclidean distance with the empirical copula was chosen as the optimal copula.

The divergence-based information using copula density function is introduced in Chapter 3. Firstly, Csiszar and Renyi-like divergence based information were proven that they can be expressed in terms of copula density function only.

Next, the limitation of the Bregman divergence was found. For example, analysing the Bregman divergence between two variables x and y , the Square loss and Itakura-Saito divergences are only sensitive to their difference $(x-y)$ and quotient (x/y) respectively. It means that the square loss and Itakura-Saito divergences do not have good discrimination ability in the case that have the same value of $(x-y)$ and (x/y) respectively.

To improve the ability of discrimination, a modified definition was proposed, by considering an extended Taylor expansion than that used for the conventional Bregman divergence. Similarly, the same modification was proposed for Burbea-Rao divergence since the Burbea-Rao divergence is the special case of Bregman divergence. Further, the Itakura-Saito divergence was modified, and then utilized as the distance for K-means classification. The classification results for Gaussian distributed data showed that the modified Itakura-Saito was more reliable and accurate than the conventional Itakura-Saito divergence. The smallest enclosing curve was applied to validate the effectiveness of modified Itakura-Saito divergence.

Experiments were conducted using two symmetrical Bernoulli distributions to analysis and compare all the divergences mentioned in this thesis, and it was shown that the general performance trend is similar for all the divergences, (though there are subtle nuances that differentiate them): each divergence has the minimal value of 0 when two marginal distributions have the same value, and the divergence values increase when two marginal distributions move apart. Moreover, Chi-square, Itakura-Saito, modified square loss and square t divergence have been found to offer better ability to control the measurement sensitivity, since these divergences change more obviously than other divergences for the same change of marginal distributions, and

Chapter 7: Conclusion and Further Research

other divergences have the approximately value of 0 when two marginal distributions are similar.

Finally, the estimation of mutual information using Gaussian copula was proposed based on the inverse standard Gaussian cumulative distribution function. The results showed that Gaussian copula based mutual information is very close to the bivariate Gaussian distribution based mutual information for Gaussian distributed data.

In chapter 4, the current techniques of image registration and performance evaluation of image fusion were firstly reviewed. This thesis then focused on the Area based image registration techniques, since some images do not contain enough distinctive objects that render them unsuitable for Feature based method. Image registration was achieved using copula density function techniques, by maximizing the divergence-based information between the overlapping parts of the first image (reference image) and the transformed second (float) image (that included rotation, translation and rescaling). Experiments on synthetic medical images, real medical (CT and MRI) images and real images (visible light and thermal images) showed that the copulas-based method offer more accurate results than the 'Gaussian assumption' and joint histogram based methods, especially for visible light and thermal images. This shows that copula based methods are more reliable than the Gaussian assumption based and joint histogram based methods for the estimation of divergence based information, since Gaussian assumptions are not accurate for these real images, and their pixel intensity distributions may be non-Gaussian. Note that the pixel intensity distributions usually offer more stable information than pixel intensities themselves, while the joint histogram method counts the number of occurrences of pixel intensity pairs.

As for the performance evaluation of image fusion, this thesis focuses on the information based measure since this measure does not require ground truth which, in many cases, may be not available. The algorithms of image fusion such as the Average method, Principal Component Analysis (PCA), Gradient Pyramid (GP), Laplacian Pyramid (LP), Ratio Pyramid (RP) and Discrete Wavelet Transform (DWT) methods are evaluated using the Tsallis and Renyi divergence based information for the fusion factor and fusion symmetry measures, and experiments were conducted on multi-sensor images. The results have shown that the Fusion Symmetry measure is much more accurate than Fusion Factor measure. Both Tsallis and Renyi divergence

Chapter 7: Conclusion and Further Research

based measure offered effective evaluation results based on the Fusion Symmetry measure. Moreover, Tsallis divergence based information offered improved ability of discrimination than conventional mutual information by adjusting its only parameter α to 3.

In chapter 5, the current techniques of band selection of hyperspectral images were firstly reviewed, and Spectral signatures were introduced to estimate a reference image in the absence of a reference image. Next, Clayton and Gaussian copula were applied to band selection for hyperspectral images, by calculating the mutual information between each band image and the reference image; and the band images which have higher mutual information with reference image were retained.

Experiments on AVIRIS dataset (220 band images) showed that Gaussian copulas based measures offer the most reasonable results for the band selection of hyperspectral images. The Clayton copula method does not offer sufficient discrimination ability when the dependency is low between reference image and band images from band 40 – 80, that make the Clayton copula based mutual information values very close to 0. The Gaussian assumption based method does not have a good ability to discriminate in the region of bands 10 to 30. The joint histogram based mutual information offered the worst results of band selection since the differences of mutual information between reference map and each band image are not evident. This means that Gaussian copulas offer a more robust and reliable measure to estimate mutual information between reference image and hyperspectral images than Clayton copula, bivariate Gaussian distribution and joint histogram based measures.

Finally, the rejection bandwidth and complementary measure were introduced to reduce the redundant information between neighbouring band images.

In chapter 6, the current image change detection techniques were reviewed firstly. The thesis focus on the Statistical similarity based measure rather than Pixel based measure since the statistical distribution of pixel intensity usually offers more stable information than the pixel intensities themselves. However, the results are disappointing for both the conventional Statistical similarity and Pixel based measure, for the complicated situation where the objects in two images respectively are very similar or same, even though, in this case, the statistical distribution of pixel

Chapter 7: Conclusion and Further Research

intensities may vary remarkably due to the external factors such as climate changing or the use of different sensors.

Copulas were used to model the dependence structure between two images by training the ‘no change’ areas of two images, and then using the conditional copula to simulate new pixel intensities to replace the pixel intensities of the first image. The final change indicator is obtained by calculating of the symmetric version of Kullback-Leibler divergence between the simulated pixel intensities and the second image. The Clayton copula and Frank copula were chosen, from amongst the Clayton, Frank and Gaussian copulas, to model the dependence for SAR and CT images respectively, since they had the minimal Euclidean distance from the empirical copula. The experimental results showed that copula based measures provided most accurate change detection than the ‘difference’ and ‘statistical similarity’ measures for SAR and CT images. The key changed areas (that indicated flooding) in SAR images, were only picked by the conditional copula measure, and missed out by methods based on the ‘difference’ and ‘Statistical similarity’ measures. The ‘difference’ measure offered the worst results since some areas that were not changed (affected) by flood but they were considered as changed areas since the images were affected by a great change in the weather.

Finally, the work was validated by plotting the Receiver Operating Characteristic (ROC) curve to verify the performance of the algorithms.

7.2 Future Research

Firstly, from the perspective of copula theory, there are two issues that need improvement. The first is to improve the estimation of copula parameters. Although Kendall’s tau correlation based method is quite reasonable, it is time-consuming, and the computing effort increases with the increase in data size. Moreover, the method is only applicable for a few copula functions. The maximum likelihood method is an alternative fast approach for estimating the copula parameters, and is feasible for any copula function; however the accuracy of estimation seems not good as Kendall’s tau correlation based method. So, the algorithm to calculate the Kendall’s tau correlation

Chapter 7: Conclusion and Further Research

should be further developed, and better and faster estimation methods should be explored.

The second is to design a better approach for choosing the best copula which can model the dependency structure optimally between marginal distributions since several copula families are available. Although an empirical copula has been proposed for the selection of the best copula, it is time-consuming with the increasing of data size and is only as effective as the finite set of copula families that are used to choose the best copula.

Moreover, from the point of view of applications, robust algorithms for optimization which are especially suitable for image registration are required. The algorithm of estimation of initial parameters for optimization should be studied for fast and automatic image registration using copulas.

In addition, the selection of appropriate parameters for generalized divergence such as Tsallis and Renyi divergence based information need to be researched to provide the ability to control the measurement sensitivity and hence achieve better accuracy and efficiency than the classic Kullback-Leibler divergence based information (Mutual Information), and their applications for image registration and performance evaluation for image fusion.

Another aspect of future work is to explore the optimal divergence for image processing since four categories of generalized divergences, eleven divergences using copulas have been researched.

Finally, the areas in signal processing can be further explored where the joint probability distributions, divergence based information and dependence between marginal distributions are required to be estimated. For examples, copulas may be explored to model signal correlation in Multi-Input Multi-Output (MIMO) for wireless communications. Moreover, copulas may be researched for Blind Source Separation (BSS) by copula based Independent Component Analysis (ICA) [Abayomi et al., 2007]. In addition, copulas may be applied for Bayesian classification using copula based joint probability density function.

Appendix

Appendix 1:

Computation and Simulation of Exponential and Rayleigh Copula

According the works of [Luke, 1962], [Brusset & Temme, 2007], the following double integrals can be calculated as:

$$\begin{aligned} & \int_0^{a_1} \int_0^{a_2} e^{-u-v} I_0(2\sqrt{\rho xy}) dx dy \\ &= (1 - e^{\rho a_2 - a_2}) + e^{\rho a_2 - a_2} K(\rho a_2, a_1) - e^{\rho a_1 - a_1} K(a_2, \rho a_1) \end{aligned}$$

where $K(x, y) = e^{-y} \int_0^x e^{-u} I_0(2\sqrt{uy}) du$

I_0 is called the first kind Modified Bessel function with order 0 and it can be expressed as:

$$I_0(z) = \sum_{k=0}^{\infty} \frac{2}{(k!)^2} \left(\frac{z}{2}\right)^{2k}$$

In the following computation, the Exponential and Rayleigh copula will be transformed to obtain the above expression fro computing the double integrals conveniently.

Computation of Exponential Copula:

The Exponential copula has been defined as [Durrani & Zeng, 2007]:

$$\int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{F^{-1}(v)} \frac{\lambda \mu}{1-\rho} \exp\left(-\frac{\lambda x}{1-\rho} - \frac{\mu y}{1-\rho}\right) I_0\left(\frac{2}{1-\rho} \sqrt{\rho \lambda \mu xy}\right) dx dy$$

Appendix

where λ, μ are the variances of random variables X and Y respectively.

Marginal probability density functions:

$$P_X(x) = \lambda e^{-\lambda x}; P_Y(y) = \mu e^{-\mu y}$$

Cumulative marginal distribution functions:

$$F_X(x) = 1 - e^{-\lambda x} \quad ; \quad F_Y(y) = 1 - e^{-\mu y}$$

Inverse marginal cumulative distribution function:

$$F_X^{-1}(u) = -\frac{\log(1-u)}{\lambda}; F_Y^{-1}(v) = -\frac{\log(1-v)}{\mu}$$

$$\text{Let } \frac{\lambda x}{1-\rho} = Z_1 \text{ and } \frac{\mu y}{1-\rho} = Z_2$$

$$\text{Since } 0 \leq x \leq -\frac{\log(1-u)}{\lambda} \text{ and } 0 \leq y \leq -\frac{\log(1-v)}{\mu},$$

$$\text{so } 0 \leq Z_1 \leq -\frac{\log(1-u)}{1-\rho} \text{ and } 0 \leq Z_2 \leq -\frac{\log(1-v)}{1-\rho}.$$

$$\text{The Jacobi determinant } J = \begin{pmatrix} \frac{\partial x}{\partial Z_1} & \frac{\partial x}{\partial Z_2} \\ \frac{\partial y}{\partial Z_1} & \frac{\partial y}{\partial Z_2} \end{pmatrix} = \begin{pmatrix} \frac{1-\rho}{\lambda} & 0 \\ 0 & \frac{1-\rho}{\mu} \end{pmatrix} = \frac{(1-\rho)^2}{\lambda\mu}.$$

$$\begin{aligned} & \int_0^{-\frac{\log(1-u)}{\lambda}} \int_0^{-\frac{\log(1-v)}{\mu}} \frac{\lambda\mu}{1-\rho} \exp\left(-\frac{\lambda x}{1-\rho} - \frac{\mu y}{1-\rho}\right) I_0\left(\frac{2}{1-\rho} \sqrt{\rho\lambda\mu xy}\right) dx dy \\ &= (1-\rho) \int_0^{b_1} \int_0^{b_2} e^{-Z_1 - Z_2} I_0(2\sqrt{\rho Z_1 Z_2}) dZ_1 dZ_2 \end{aligned}$$

$$\text{where } b_1 = -\frac{\log(1-u)}{1-\rho} \text{ and } b_2 = -\frac{\log(1-v)}{1-\rho}.$$

Computation of Rayleigh Copula:

Appendix

The Rayleigh copula can be written as [Durrani & Zeng, 2007]:

$$\int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{F^{-1}(v)} \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho^2)} \exp\left(-\frac{1}{(1-\rho^2)} \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2}\right)\right) I_0\left(\frac{xy\rho}{(1-\rho^2)\sigma_X\sigma_Y}\right) dx dy$$

Marginal propabitliy density functions:

$$P_X(x) = \frac{\pi}{\sigma_X \sqrt{2\pi x}} \exp\left(-\frac{x}{\sigma_X}\right); P_Y(y) = \frac{\pi}{\sigma_Y \sqrt{2\pi y}} \exp\left(-\frac{y}{\sigma_Y}\right)$$

Cumulative marginal distribution functions:

$$F_X(x) = 1 - \exp\left(-\frac{x^2}{2\sigma_X^2}\right); F_Y(y) = 1 - \exp\left(-\frac{y^2}{2\sigma_Y^2}\right)$$

Inverse marginal cumulative distribution function:

$$F_X^{-1}(u) = \sqrt{-2\sigma_X^2 \log(1-u)}; F_Y^{-1}(v) = \sqrt{-2\sigma_Y^2 \log(1-v)}$$

Let $F_X^{-1}(u) = a_1$ and $F_Y^{-1}(v) = a_2$, then

$$\begin{aligned} & \int_{-\infty}^{F^{-1}(u)} \int_{-\infty}^{F^{-1}(v)} \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho^2)} \exp\left(-\frac{1}{(1-\rho^2)} \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2}\right)\right) I_0\left(\frac{xy\rho}{(1-\rho^2)\sigma_X\sigma_Y}\right) dx dy \\ &= \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho^2)} \exp\left(-\frac{1}{(1-\rho^2)} \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2}\right)\right) I_0\left(\frac{xy\rho}{(1-\rho^2)\sigma_X\sigma_Y}\right) dx dy \end{aligned}$$

Let $\rho^2 = \rho'$, $\frac{x^2}{2(1-\rho')\sigma_X^2} = Z_1$ and $\frac{y^2}{2(1-\rho')\sigma_Y^2} = Z_2$, then

$$\begin{aligned} & \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho^2)} \exp\left(-\frac{1}{(1-\rho^2)} \left(\frac{x^2}{2\sigma_X^2} + \frac{y^2}{2\sigma_Y^2}\right)\right) I_0\left(\frac{xy\rho}{(1-\rho^2)\sigma_X\sigma_Y}\right) dx dy \\ &= \int_{-\infty}^{a_1} \int_{-\infty}^{a_2} \frac{xy}{\sigma_X^2 \sigma_Y^2 (1-\rho')} \exp(-Z_1 - Z_2) I_0(2\sqrt{\rho' Z_1 Z_2}) dx dy \end{aligned}$$

Since $0 \leq x \leq a_1$; $0 \leq y \leq a_2$ and $Z_1 = \frac{x^2}{2(1-\rho')\sigma_X^2}$; $Z_2 = \frac{y^2}{2(1-\rho')\sigma_Y^2}$,

Appendix

So, $0 \leq Z_1 \leq \frac{a_1^2}{2\sigma_x^2(1-\rho')}$; $0 \leq Z_2 \leq \frac{a_2^2}{2\sigma_y^2(1-\rho')}$.

Let $\frac{a_1^2}{2\sigma_x^2(1-\rho')} = b_1$; $\frac{a_2^2}{2\sigma_y^2(1-\rho')} = b_2$. The Jacobi determinant

$$J = \begin{pmatrix} \frac{\partial x}{\partial Z_1} & \frac{\partial x}{\partial Z_2} \\ \frac{\partial y}{\partial Z_1} & \frac{\partial y}{\partial Z_2} \end{pmatrix} = \begin{pmatrix} \frac{(1-\rho')\sigma_x^2}{x} & 0 \\ 0 & \frac{(1-\rho')\sigma_y^2}{y} \end{pmatrix} = \frac{(1-\rho')^2 \sigma_x^2 \sigma_y^2}{xy}$$

Consequently, the bivariate Rayleigh copula can be expressed as:

$$(1-\rho') \int_{-\infty}^{b_1} \int_{-\infty}^{b_2} \exp(-Z_1 - Z_2) I_0(2\sqrt{\rho' Z_1 Z_2}) dZ_1 dZ_2$$

where $b_1 = \frac{a_1^2}{2\sigma_x^2(1-\rho')}$; $b_2 = \frac{a_2^2}{2\sigma_y^2(1-\rho')}$

$$a_1 = \sqrt{-2\sigma_x^2 \log(1-u)}; \quad a_2 = \sqrt{-2\sigma_y^2 \log(1-v)}$$

$\rho' = \rho^2$ and ρ is the Pearson correlation between two variables x and y .

Simulation of Exponential copula:

Step 1:

Generate n by 2 uniform random variables.

$x = [x1; x2]$;

Step 2:

Let $Y = [-\frac{1}{\lambda} \log(x1); -\frac{1}{\mu} \log(x2)]$

Step 3:

$$\sigma = \begin{pmatrix} \text{var}(x) & \text{cov}(x, y) \\ \text{cov}(x, y) & \text{var}(y) \end{pmatrix} = \begin{pmatrix} \frac{1}{\lambda^2} & \frac{\rho}{\lambda\mu} \\ \frac{\rho}{\lambda\mu} & \frac{1}{\mu^2} \end{pmatrix}$$

Appendix

Let A is the Cholesky decomposition [David & Lloyd, 1997] of sigma.

Step 4:

Let $Z=Y*A$, where $Z=[Z_1; Z_2]$;

Let $u_1 = 1 - e^{-\lambda Z_1}$ and $u_2 = 1 - e^{-\lambda Z_2}$.

Simulation of Rayleigh Copula:

Step 1:

Generate n by 2 uniform random variables

Let $x = [x_1; x_2]$;

Step 2:

Let $s_1 = \sigma_X^2$ and $s_2 = \sigma_Y^2$.

Let $Y = [\sqrt{-2s_1 \log(1-x_1)}; \sqrt{-2s_2 \log(1-x_2)}]$

where Y is a n-by-2 matrix. Let $Y=[Y_1 Y_2]$;

Step 3:

$$\text{sigma} = \begin{pmatrix} \frac{(4-\pi)\sigma_X^2}{2} & \rho \frac{(4-\pi)\sigma_X\sigma_Y}{2} \\ \rho \frac{(4-\pi)\sigma_X\sigma_Y}{2} & \frac{(4-\pi)\sigma_Y^2}{2} \end{pmatrix}$$

Let A is the Cholesky decomposition of sigma.

Step 4:

Let $Z=Y*A$, where $Z=[Z_1; Z_2]$;

Let $u_1 = 1 - e^{-\frac{Z_1^2}{2s_1}}$ and $u_2 = 1 - e^{-\frac{Z_2^2}{2s_2}}$.

Appendix 2:

Simulation of Bivariate Gaussian distribution [Scheuer & Stoller, 1962];

Suppose z_1 and z_2 are two independent random variables with the standard normal

Appendix

distribution. Let

$$\begin{cases} X = \mu_1 + \sigma_1 z_1 \\ Y = \mu_2 + \sigma_2 \rho z_1 + \sigma_2 \sqrt{1 - \rho^2} z_2 \end{cases}$$

Then, the joint distribution $[X, Y]$ is the bivariate normal distribution with the parameters $[\mu_1, \mu_2, \sigma_1, \sigma_2]$ where μ_1, μ_2 are the mean values and σ_1, σ_2 are the standard deviations of two random variables X and Y respectively. Considering the special case for joint standard Gaussian distribution, let the mean values = 0 and standard deviations = 1, then,

$$\begin{cases} X = z_1 \\ Y = \rho z_1 + \sqrt{1 - \rho^2} z_2 \end{cases}$$

Appendix 3:

Algorithm of rejection bandwidth and setting the complementary threshold for band selection: [Guo, et al., 2006]

Let the expected selected band number be 'expnum' and selected band is set as 'sb' which has an initial value as empty.

Remaining bands are set as 'rb' and with initial values={1,2...220}. Mutual information change between neighbours is designated as: $d(n)=MI(n)-MI(n-1)$.

Next, the algorithm for rejection bandwidth and setting the complementary threshold for band selection can be executed as:

Appendix

```

while(length(sb)<expnum)
  sbi=argmax(MI(between s and estimated reference))
  where s ∈ rb and sbi is the selected band index.
  neighbour set N={n=sbi-(B+1),...sbi,...sbi+B};
  if max(d(n))<t then
    sb=sb∪ sbi; rb = {rb remove sbi and N from rb};
  else
    sb=sb∪ sbi; rb = {rb remove sbi from rb};
  end if
end while

```

Appendix 4:**Pearson System distribution**

Pearson system [Stuart & Ord, 1987] shows many probability density functions $f(x)$ satisfy a differential equation as follows:

$$\frac{f'(x)}{f(x)} = \frac{x-a}{b_0+b_1x+b_2x^2} \text{ where } \begin{cases} b_0 = -\frac{\mu_2(4\beta_2-3\beta_1)}{10\beta_2-12\beta_1-18} \\ b_1 = a = -\frac{\sqrt{\mu_2}\sqrt{\beta_1}(\beta_2+3)}{10\beta_2-12\beta_1-18} \\ b_2 = -\frac{2\beta_2-3\beta_1-6}{10\beta_2-12\beta_1-18} \end{cases}$$

parameters β_1 and β_2 are:

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \text{skewness}^2 ; \beta_2 = \frac{\mu_4}{\mu_2^2} = \text{kurtosis}$$

Appendix

where μ_k is called central moment of order k of random variables X and has been defined as:

$$\mu_k = E\{[X - E(X)]^k\}$$

The different probability distributions of Pearson system can be distinguished by β_1 and β_2 . Hence the probability distribution of the Pearson system can be evaluated by calculating the first four central moments. For examples, the Gaussian distribution is located at $\beta_1 = 0$ and $\beta_2 = 3$, and the Gamma distribution can be determined by

$\beta_2 = \frac{3}{2}\beta_1 + 3$, more distributions can be found in [Stuart & Ord, 1987].

Appendix 5:

Gram-Charlie and Edgeworth Series

Considering a probability density which is not too far from and has same mean value and variance with Gaussian probability density may be expressed in terms of the Gram-Charlie series as [Stuart & Ord, 1987]:

$$f_X(x) = (1 + \frac{1}{6}\mu_3 H_3(x) + \frac{1}{24}(\mu_4 - 3)H_4(x) + \dots)G_X(x)$$

where $H(x)$ is the Hermite polynomials. Considering the standard Gaussian probability density function:

$$G_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

Its derivatives can be calculate as:

Appendix

$$\begin{cases} G'_X(x) = -xG_X(x) \\ G''_X(x) = (x^2 - 1)G_X(x) \\ G'''_X(x) = (3x - x^3)G_X(x) \\ \dots \end{cases}$$

The Hermite polynomial $H(x)$ for standard Gaussian distribution is defined in the following formula as:

$$G_X^{[r]}(x) = H_r(x)G_X(x)$$

The first six items has been calculated as:

$$\begin{cases} H_0(x) = 1 \\ H_1(x) = x \\ H_2(x) = x^2 - 1 \\ H_3(x) = x^3 - 3x \\ H_4(x) = x^4 - 6x^2 + 3 \\ H_5(x) = x^5 - 10x^3 + 15x \\ H_6(x) = x^5 - 15x^4 + 45x^2 - 15 \end{cases}$$

The Edgeworth series provides the approximation of the probability density function in terms of cumulant K_X and Hermite polynomials $H(x)$ of order r which is truncated to 6 as [Stuart & Ord, 1987]:

$$f_X(x) = [1 + \frac{K_{X3}}{6} H_3(x) + \frac{K_{X4}}{24} H_4(x) + \frac{K_{X5}}{120} H_5(x) + \frac{K_{X6} + 10K_{X3}^2}{720} H_6(x)]G_X(x)$$

Formally, the cumulant k_r is defined as [Stuart & Ord, 1987]:

$$\exp\left(\sum_{r=1}^{\infty} \frac{k_r t^r}{r!}\right) = \sum_{r=0}^{\infty} \frac{\mu'_r t^r}{r!}$$

where μ'_r is called raw moment and has been defined as:

$$\mu'_r = E\{X^r\}$$

Appendix

The relationship between the raw moment and cumulant has been found as [Stuart & Ord, 1987]:

$$k_n = \mu'_n - \sum_{k=1}^{n-1} \binom{n-1}{k-1} k_k \mu'_{n-k}$$

The first cumulant is the expected value; the second and third cumulants are the second and third central moments respectively (the second central moment is the variance); but the higher cumulants are neither raw moments nor central moments, but rather more complicated polynomial functions of the moments. The first 6 equations between cumulant and raw moment are:

$$\left\{ \begin{array}{l} K_{X1} = \mu'_{X1} \\ K_{X2} = \mu'_{X2} - K_{X1} \\ K_{X3} = \mu'_{X3} - 3K_{X2}K_{X1} - K_{X1}^3 \\ K_{X4} = \mu'_{X4} - 4K_{X3}K_{X1} - 3K_{X2}^2 + 6K_{X2}K_{X1}^2 - K_{X1}^4 \\ K_{X5} = \mu'_{X5} - 5K_{X4}K_{X1} - 10K_{X3}K_{X2} - 10K_{X3}K_{X1}^2 - 15K_{X2}^2K_{X1} - 10K_{X2}K_{X1}^3 - K_{X1}^5 \\ K_{X6} = \mu'_{X6} - 6K_{X5}K_{X1} - 15K_{X4}K_{X2} - 15K_{X4}K_{X1}^2 - 10K_{X3}^2 - 60K_{X3}K_{X2}K_{X1} - 20K_{X3}K_{X1}^3 \\ \quad - 15K_{X2}^3 - 45K_{X2}^2K_{X1}^2 - 15K_{X2}K_{X1}^4 - K_{X1}^6 \end{array} \right.$$

Appendix 6:

Kernel smoothing technique

Kernel smoothing which has been defined as [Bowman & Azzalini, 1997]:

$$f(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right)$$

where n is the number of sample, h is called smoothing parameter or bandwidth, K is called kernel function and satisfy the condition:

Appendix

$$\int_{-\infty}^{\infty} K(x)dx = 1$$

The common kernel function is standard Gaussian probability density function which is also known as Parzen window technique.

$$K\left(\frac{x-x_i}{h}\right) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-x_i)^2}{2h^2}}$$

It should note that the selection of bandwidth is much important than the selection of kernel function. The optimal bandwidth of Gaussian kernel smoothing is suggested as [Bowman & Azzalini, 1997]:

$$h = \left(\frac{4}{3n}\right)^{\frac{1}{5}} \sigma$$

where σ is the standard deviation.

Bibliography

Abayomi, K.A., Lall, U. & Dela Pena,V.H. (2007). “Copula Based Independent Component Analysis”,[Online]: <http://www2.isye.gatech.edu/~kabayomi3/papers/papers/CICA.pdf>.

Abramowitz, J., & Stegun, I. A. (1965). “Handbook of Mathematical Functions”, National Bureau of Standards, Applied Math. Dover Publications.

Abu-Dayya, A. A., & Beaulieu, N. C. (1994). “Analysis of switched diversity systems on generalized-fading channels”, IEEE Transaction on Communication, Volume 42 , pp. 2959–2966.

Adams, R. (1993). “Radial Decomposition of Discs and Spheres”, Graphical models and Image Processing, Volume. 55, Issue 5, pp. 325–332.

Ali, S and Silvey, S. (1966). “A general class of coefficients of divergence of one distribution from another”, Journal of the Royal Statistical Society, Series B, 28(1), pp. 201-217.

Assen, H, Egmon-Petersen, M and Reiber, J H.C. (2002). “Accurate object localization in gray level images using the center”, IEEE transactions on image processing, Volume 11, No.12, pp. 1379-1384.

Banerjee, A, Merugu, S, Dhillon, I.S, Ghosh, J. (2005). “Clustering with bregman divergences”, Journal of Machine Learning Research(6), pp. 1705–1749.

Barbara, Z and Flusser, J. (2003). “Image registration methods: a survey”, Image and Vision Computing (21), pp. 977–1000.

Bowman, A. W., & Azzalini, A. (1997). “Applied Smoothing Techniques for Data Analysis” Vouolume 18, Oxford Statistical Science Series.

Bregman, L M. (1967). “The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming”, USSR Comp Math and Math. Phys, 7, pp. 200-217.

Bibliography

Brusset, X & Temme, N.M. (2007). "Optimizing an objective function under a bivariate probability model", *European Journal of Operational Research*, Volume 179, pp. 444-458.

Bruzzone, L. & Prieto, D.F. (2002). "A partially unsupervised cascade classifier for the analysis of multitemporal remote-sensing images". *Pattern Recognition Letters*, Volume 23, Issue 9, pp. 1063-1071.

Burt, P.J. & Adelson, E.H. (1983). "The Laplacian pyramid as a compact image code", *IEEE Transaction on Communication*, Volume 31, Issue 4, pp. 532-540.

Burt, P.J. & Kolczynski, R.J. (1993). "Enhanced image capture through fusion", *Proceeding of 4th International Conference of Computer Vision*, pp. 173-182.

Carrier, F., Krook, M., & Pearson, C. E. (1983). "Functions of a complex variable: theory and technique", Hod Books.

Chatelain, F., Tourneret, J. Y., Inglada, J., & Ferrari, A. (2007). "Bivariate Gamma Distributions for Image Registration and Change Detection". *IEEE Transactions on Image Processing*, Volume 16, Issue. 7 , pp. 1796-1806.

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). "Copula Methods in Finance", John Willey & Sons, Ltd.

Cho, Y.D, Al-Naimi, K. & Kondo, A. (2001). "Mixed decision-based noise adaptation for speech enhancement", *IET Electronics Letters*, Volume 37, Issue 8, pp. 540-542.

Christensen, D. (2005). "Fast algorithms for the calculation of Kendall's tau", *Physica-Verlag Computational Statistics* 20, pp.51-62,

Clayton, D. G. (1978). "A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence", *Biometrika* 65 , pp. 141-151.

Csiszar, I. (1967). "Information-type measures of difference of probability distribution and indirect observations", *Studia Scientiarum Mathematicarum Hungarica*, 2, pp. 299-318.

Cvejic, N., Canagarajah, C.N and Bull, D.R. (2006) "Image Fusion Metric based on Mutual Information and Tsallis Entropy", *IET Electronics Letters*, Volume. 42, Issue. 11, pp. 626-627.

Bibliography

David, B. I., & Lloyd N, T. (1997). "Numerical linear algebra", Philadelphia: Society for Industrial and Applied Mathematics.

Downton, F. (1970). "Bivariate exponential distributions in reliability theory", Journal of Royal Statistical Society. B 32 , pp. 408–417.

Durrani, T. S, & Zeng, X. (2007). "Copulas for bivariate probability distributions", IET Electronic Letters , Volume 43 , Issue 4, pp. 248-249.

Durrleman, V., Nikeghbali, A., & Roncalli, T. (2000). "Which copula is the right one? roupe de Recherche Operationnelle", Credit Lyonnais, Working paper.

Dwork, C., Kumar, R., Naor, M. & Sivakumar, D. (2001), "Rank Aggregation Revisited", 10th International World Wide Web Conference, pp. 613-622.

Fang, K.T., Kotzk, S, & Ng, K.W. (1987). "Symmetric multivariate and related distributions", London: Chapman & Hall.

Fawcett, T. (2006). "An introduction to ROC analysis", Pattern Recognition Letters, 27, pp.861–874.

Frank, M. J. (1979). "On the simultaneous associativity of $F(x,y)$ and $x+y-F(x,y)$ ", Aequationes Math. 19 , pp. 194-226.

Gander, W. & Gautschi, W. (2000). "Adaptive Quadrature-Revisited", Bit Numerical Mathematics, Volume 40, No. 1, pp. 84-101.

Genz, A. (2004). "Numerical computation of rectangular bivariate and trivariate normal and t probabilities", Journal of statistics and computing, Springer Netherlands. Volume 14, No. 3 , pp. 251-260.

Gibbons, J. D. (1985). "Nonparametric Statistical Inference", Second edition, Marcel Dekker, New York.

Gonzalez R.C & Woods R.E. (2008). "Digital Image Processing", Third Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey.

Gray, R, Buzo. A, Gray, A, Matusugama, Y. (1980). "Distortion measures for speech processing", IEEE Transactions on Acoustics, Speech and Signal Processing, Volume 28, Issue 4, pp. 367-376.

Guo, B., Gunn, S. R., Damper, R. I. and Nelson, J. D. B. (2006). "Band selection for hyperspectral image classification using mutual information". IEEE Geoscience and Remote Sensing Letters, Volume 3, Issue 4. pp. 522-526.

Bibliography

Hamedani, G. G., & Tata, M. N. (1975). "On the determination of the bivariate normal distribution from distributions of linear combinations of the variables", *The American Mathematical Monthly*, Volume 82, Issue 9, pp. 913–915.

He, Y, Hamza, B and Krim, H. (2003). "A generalized divergence measure for robust image registration", *IEEE Transactions on Signal Processing*, Volume 51, No.5, pp. 1211-1220.

Huard, D., Evin, G. & Favre A. (2006). "Bayesian Copula Selection", *Elsevier Computational Statistics & Data Analysis*, Volume. 51, Issue. 2, pp. 809-822.

Ifarraguerri, A. and Prairie, M. W. (2004). "Visual method for spectral band selection.", *IEEE Letters on Geoscience and Remote Sensing*, Volume 1, Issue. 2, pp. 101-106.

Inglada, J. (2003). "Change detection on SAR images by using a parametric estimation of the Kullback-Leibler divergence", *International Geoscience and Remote Sensing Symposium*, Volume 6, pp. 4104-4106.

Inglada, J., & Mercier, G. (2007). "A New Statistical Similarity Measure for Change Detection in Multitemporal SAR images and Its Extension to Multiscale Change Analysis", *IEEE Transactions on Geoscience and Remote Sensing*, Volume 45, No. 5, pp.1432-1445.

Jain, A. K. (2010). "Data clustering: 50 years beyond K-means", *Pattern Recognition Letters*, Volume 31, Issue 18, pp. 651-666.

Jia, Y. (1998). "Fusion of Landsat TM and SAR Images Based on Principal Component Analysis", *Remote Sensing Technology and Application*, Volume 13, Issue 1, pp. 4649-4654.

Keys, R. (1981). "Cubic convolution interpolation for digital image processing", *IEEE Transactions on Signal Processing, Acoustics, Speech, and Signal Processing*, Volume 29, pp.1153 -1160.

Kruskal, W. H. (1958). "Ordinal Measures of Association". *Journal of the American Statistical Association*, Volume 53, No. 284, pp. 814–861.

Kumar, P. and Shoukri, M.M. (2007). "Copula based prediction models: an application to an aortic regurgitation study", *BMC Medical Research Methodology*, Volume 7.

Bibliography

Li, Y., Shi, B. & Zhang, J. (2008). "Copula joint function and its application in probability seismic hazard analysis", Volume 21, No 3, pp 296-305.

Lagarias, J C, Reeds, A.J, Wright, H.M and Wright, E.P. (1998), "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions", Society for Industrial and Applied Mathematics Journal of Optimization, Volume 9, No.1, pp. 112-147.

Landsman, Z., Valdez, A. E. (1999). "Tail Conditional Expectations for Elliptical Distributions", Technical Report N 02-04.

Luke, Y.L. (1962). "Integrals of Bessel Function", New York, McGraw-Hill.

Maes, F, Collignona, A, Vandermenlen, A & Marchal, G. (1997). "Multimodality image registration by maximization of mutual information", IEEE Transactions on Medical Imaging, Volume 16, Issue 2, pp. 187-198.

Mallt, S. (1989). "A theory for multiresolution signal decomposition: the wavelet representation", IEEE Pattern Analysis and Machine Intelligence, Volume 11, No.7, pp. 674-693.

Maritz, J. S. (1981). "Distribution-Free Statistical Methods", Chapman & Hall.

Martin, S. (2006). "On Information Measures for Medical Image Registration", PhD thesis, University of Strathclyde.

Martin, S and Durrani, T. S. (2007). "A new divergence measure for medical image registration", IEEE Transacation on Image Processing, Volume 16, No. 4, pp. 957-966.

Martin, S, Morison, G, Nailon, W and Durrani, T.S. (2004). "Fast and accurate image registration using Tsallis entropy and simultaneous perturbation stochastic approximation", IET Electronic Letters, Volume. 40, Issue 10, pp. 595-597.

Mercier, G., Moser, G., & Serpico, S. B. (2008). "Conditional copula for change detection on heterogeneous SAR data", IEEE Transaction on Geoscience and Remote Sensing, Volume 46, Issue 5, pp.1428-1441.

Murphy, F and Murphy, B. (2006). "Copulas as a risk management tool", Finance Magazine, Ireland.

Nelsen, R. B. (1999). "An Introduction to Copulas", Springer Verlag, Volume 139, New York.

Bibliography

Nock, R and Nielsen, F. (2005). "Fitting the smallest enclosing bregman ball", Machine Learning: ECML , Volume 3720, pp. 649-656.

Osterreicher, F. (2002). "Csiszar's f-divergences-Basic properties", Report Collection, [Online]: <http://rgmia.vu.edu.au/monographs/csiszar.htm>.

Otsu, N. (1979). "A Threshold Selection Method from Gray Level Histograms", IEEE Transactions on Systems, Man and Cybernetics, Volume 9, Issue 1, pp. 62-66.

Ozturk, Y. and Abut, H. (1990). "Applications of Itakura-Saito Type Spectral Distortion Measures to Image Analysis and Classification", Twenty-fourth Asilomar Conference on Signals, Systems & Computers, November 5-7, Pacific Grove, California.

Pardo, M. C. and Vajda, I. (1997). "About distances of discrete distributions satisfying the data processing theorem of information theory", IEEE Transactions on Information Theory, Volume 43, Issue 4, pp. 1288-1293.

Pardo, M. C and Vajda, I. (2003). "On asymptotic properties of information-theoretic divergences", IEEE Transactions on Information Theory, Volume 49, Issue 7, pp. 1860-1867.

Parker, J.A., Kenyon, R.V. and Troxel D.E. (1983). "Comparison of interpolating methods for image resampling", IEEE Transaction on Medical Imaging, Volume 2, Issue.1, pp. 31-39.

Perrone, L.F., Wieland, F.P., Liu, J., Lawson, B.G., Nicol, D.M and Fujimoto, M. (2006). "Using Copulas in Risk Analysis", 38th Winter Simulation Conference, Monterey, California.

Philip, J R. (1960). "The function $\text{inverfc } \theta$." Australian Journal of Physics, Volume 13, pp. 13-20.

Pluim, J.P., Maintz, J.B.A & Viergever, M.A. (2004). "f-information measures in medical image registration", IEEE Transactions on Medical Imaging, Volume 23, Issue 12, pp. 1508-1516.

Press, W. H, Flannery, P.B, Tenkolsky, A.S and Vetterling, T.W. (1992). "Numerical Recipes in C", Cambridge University Press.

Pu, R, and Gong, P. (2000). "Band Selection from Hyperspectral Data for Conifer Species Idenification", Geogrphic Information Sciences, Volume. 6, No. 2, pp.137-142.

Bibliography

Sawhney, H. S and Kumar, R. (1999). "True multi-image alignment and its applications to mosaicing and lens distortion correction", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 21, Issue 3, pp. 235–243.

Scheuer, E.M. & Stoller, D. (1962). "On the Generation of Normal random Vectors", Technometrics, Volume 4, No.2, pp. 278-281.

Schweizer, B. & Wolff, E.F. (1981). "On nonparametric Measures of Dependence for Random Variables", The Annals of Statistics, Volume 9, No. 4, pp. 879-885.

Seber, G. (1984). "Multivariate Observations", Wiley, New York.

Sharma, R K and Pavel, M. (1997). "Multisensor image registration", Journal of Society for Information Display, pp. 951-954.

Shorack, G.R. & Wellner, J.A. (1986). "Empirical Processes with applications to Statistical", John Wiley & Sons Inc.

Sijbers, J, Dekker, A.J, Scheunders, P and Van Dyck, D. (1998). "Maximum-Likelihood Estimation of Rician Distribution Parameters", IEEE Transactions on Medical Imaging, Volume. 17, No.3, pp. 357-361.

Sklar, A. (1959). "Fonctions de repartition a n dimensions et leurs marges", Publ. Ins. Statist. Univ. Pairs 8 , pp. 229-231.

Spath, H. (1985). "Cluster Dissection and Analysis: Theory, FORTRAN Programs", Halsted Press, New York.

Stathaki, T. (2008). "Image fusion: algorithms and applications", Academic Press, 2nd Edition.

Stauffer, C., Eric, W. & Grimson, L. (2000). "Learning Patterns of Activity Using Real-Time Tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, Volume 22, pp. 747-757.

Stuart, A. & Ord, J.K. (1987). "Kendall's Advanced Theory of Statistics", 5th Edition, Volume 1, Charles Griffin & Company Limited, London.

Thomas, C. M. & Joy, T.A. (1991). "Elements of Information Theory", John Wiley & Sons, Inc.

Toet, A. (1989). "Image Fusion by a Ratio of low-pass pyramid", Pattern Recognition Letters, Volume 9, pp. 245-253.

Bibliography

Trivedi, P. K., & Zimmer, D. M. (2007). “Copula Modeling: An Introduction for Practitioners”, World Scientific Publishing.

Viola, P and Wells, W M. (1997). “Alignment by maximization of mutual information”, *International Journal of Computer Vision*, Volume 24 , No. 2, pp. 137–154.

Van den Boomgard, R., & van Balen, R. (1992). “Methods for Fast Morphological Image Transforms Using Bitmapped Images”, *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, Volume. 54, No. 3, pp. 252–254.

Wakuya, H., Yamashita, K., & Shida, K. (2007). “Signal processing for brain temperature change detection by noninvasive measurement: An independent component analysis approach”, *Systems and Computers in Japan*, Volume 38, Issue 6, pp. 80-90.

White, R.G. (1991). “Change detection in SAR imagery, *International Journal of Remote Sensing*”, Volume 12, Issue 2, pp. 339 – 360.

Winitzki, S. (2008). “A Handy Approximation for the error function and its inverse” [Online] <http://homepages.physik.uni-muenchen.de/~Winitzki/erf-approx.pdf>.